



**HAL**  
open science

# Organisation du complexe d'espèce et décryptage des structures des génomes en mosaïque interspécifiques chez les agrumes cultivés

Franck Curk

► **To cite this version:**

Franck Curk. Organisation du complexe d'espèce et décryptage des structures des génomes en mosaïque interspécifiques chez les agrumes cultivés. Génétique des plantes. Université Montpellier II - Sciences et Techniques du Languedoc, 2014. Français. NNT : 2014MON20223 . tel-02290617

**HAL Id: tel-02290617**

**<https://theses.hal.science/tel-02290617>**

Submitted on 17 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

Pour obtenir le grade de  
Docteur

Délivré par **UNIVERSITE MONTPELLIER 2**

Préparée au sein de l'école doctorale SIBAGHE  
Et des unités de recherche Amélioration Génétique et  
Adaptation des Plantes (Agap) et *Centro de Proteccion  
Vegetal y Biotecnologia*, IVIA, Valence, Espagne.

Spécialité : **Biologie moléculaire**

Présentée par **Franck CURK**

**Organisation du complexe d'espèce et  
décryptage des structures des génomes en  
mosaïque interspécifiques chez les  
agrumes cultivés**

Soutenue le 15 décembre 2014 devant le jury composé de

Mme Perla HAMON, Directeur de recherche, IRD	Président du jury
Mme Karine ALIX, Maître de Conférences, AgroParisTech,	Rapporteur
M. Thierry ROBERT, Maître de Conférences, Univ. Paris 6	Rapporteur
M. Patrick OLLITRAULT, Directeur de recherche, Cirad	Directeur de thèse
M. Luis NAVARRO, Professeur, IVIA	Co-directeur de thèse

---

# Dédicace

---

Afin d'être reconnaissant envers celles et ceux qui m'ont appuyé et encouragé à effectuer ce travail de recherche, je dédie ce mémoire :

À **mes très chers parents** pour tous les sentiments d'affection et d'amour qu'ils m'ont apporté.

À **Monsieur Thomas** pour son dévouement et la confiance qu'il m'a toujours accordé : *sans vous et votre sagesse, Monsieur Thomas, rien de tout cela n'aurait été possible.*

Et à tous les autres **membres de ma famille**, sans exception.

---

# Remerciements personnels

---

« Mesurer l'importance des choses à la toise de son carré de navets est odieuse méthode, qui oublie l'essentiel en faveur d'un détail de rien. L'homme à courtes vue troublerait l'ordre de l'univers pour la conservation d'une douzaine de pruneaux... »

Jean-Henri Fabre, *Souvenirs Entomologiques*.

C'est parce que, enfant, j'ai lu les *Souvenirs entomologiques* de **Jean-Henri Fabre** que la science en général et la biologie en particulier est réellement devenue le moteur de mes choix de vie.

Je tiens à remercier dans un premier temps les gens qui m'ont accordé leur confiance et ont accepté ce projet fou de commencer une thèse de doctorat après dix ans d'Inra : **Mme Françoise Dosba** qui lança l'idée, **Mme Hélène Lucas** qui l'accepta, **Mmes Anne-Françoise Adam-Blondon et Carole Caranta** pour avoir continué l'aventure, **M. Raphaël Morillon** qui me proposa immédiatement un premier sujet et enfin, et surtout, **M. Patrick Ollitrault**, mon directeur de thèse, qui assumait la responsabilité de me prendre dans son équipe.

Non seulement **M. Patrick Ollitrault** a pris le risque de croire en moi, mais il a sacrifié une grande partie de son temps, de son confort et de sa santé pour que cette thèse aboutisse. **Patrick, ton dévouement à la science et aux gens que tu accompagnes est un exemple pour moi et devrait l'être pour un grand nombre de chercheurs et de professeurs ! Merci.**

Bien évidemment, ce projet n'aurait de toute façon pas abouti sans les efforts bienveillants de **M. Luis Navarro** qui non seulement accepta de me recevoir dans son laboratoire mais a également mis tout en œuvre sur le plan humain, matériel et financier pour que mes recherches aboutissent.

Merci à **M. Manuel Talon, M. Javier Terol Alcayde, Mlle Victoria Ibañez et toute l'équipe du Centro de Genómica de l'ivia** en Espagne pour leur aide et leur participation à ces travaux.

Merci également à **tout le personnel del Instituto Valenciano de Investigaciones Agrarias** qui m'a reçu et accueilli les bras grands ouverts depuis les réceptionnistes et gardiens jusqu'aux chercheurs et directeurs en passant par les techniciens de terrain, le personnel administratif et les simples stagiaires de cet extraordinaire établissement. Merci à vous tous.

J'ai, pendant ces 4 années passées à Valence en Espagne, pu bénéficier de l'aide, du soutien et des efforts de toute une équipe soudée et efficace ; merci **Caroline** pour ta formation à la cuisine moléculaire ; Merci **Andres** pour ta patience et ta disponibilité, ce fut un grand atout d'avoir un ami au labo ; Merci **Frédérique** pour ton efficacité et ton soutien, qu'aurais-je fait sans toi ? Merci **Gema** pour ton appui et tes leçons de Castillan ; Merci **Juan** pour ta disponibilité et ta gentillesse ; Merci **Mari-Cruz, José, Maria, Montse, Vero, los Pablo(s), Ana, Berta, Marta, Regina, Nuria, Jesus, Giovanni, Fernando, Alida, Tony, Tania** et tous les autres pour votre soutien et votre humour, ¡Gracias!

Merci à **Mme Nuria Duran Villa** pour sa bonne humeur et son soutien. Merci à **M. Salvador Zaragoza** pour sa passion contagieuse des agrumes et de leur histoire.

Merci à **M. Francisco Tadeo** et toute sa famille et belle famille pour l'accueil extraordinaire qu'ils m'ont réservé, *Quico, Ya sabes que si vienes a Córcega comeremos una mona de pascua ¡como dios manda!*

Merci à **Cali et toute son équipe** pour ces inoubliables *Almuerzos, cafe solos, arroz negros, arroz a la cubana, arroz y verduras, arroz, arroz, arroz y paellas ¡he disfrutado mogollón!*

Un merci particulier à **Elsa** pour l'extraordinaire rencontre qu'elle m'a permis de faire...

---

En Espagne je n'ai pas oublié mes collègues de Corse qui ont participé à la faisabilité et à la réussite de ce projet. Merci **Jean-Marc, François, Yann, les Gilles, Isabelle, Thibaut, Marie, Hager, Olivier, Marion, Véro, Emmanuelle, Dominique, Emmanuel, Albert** et bien d'autres, pour votre aide. Merci à l'ensemble des membres de **la commission locale de la formation permanente** de m'avoir accompagné avant et pendant cette formation diplômante. Merci à **l'ensemble du personnel Inra-Cirad du Centre Inra de Corse** sur qui j'ai toujours pu compter.

Merci à **la commission nationale de la formation permanente de l'Inra** d'avoir accepté de participer à ce projet.

Merci aux agents de **l'UMR Agap** qui m'ont accueilli dans leur équipe en cours de thèse. Merci à **M. Jean-Christophe Glassman, Mme Marie-France Duval** ainsi que **M. Xavier Perrier** et **M. Jean-Pierre Jacquemoud-Collet** pour leur aide et leur soutien.

Un grand Merci à **Camille Jacquemond** pour ces années de formation et d'apprentissage. **Camille, tu m'as transmis l'essentiel : aimer les agrumes !**

Merci à **M. Serge Hamon** et **Mme Dominique Brunel** d'avoir accepté de m'accompagner tout au long de ces années en participant à mon comité de thèse. *Vos recommandations et vos conseils m'ont été précieux.*

Merci à **David Karp** pour son aide et pour l'intérêt qu'il a toujours porté à mes travaux.

Merci à **Nicolas Antomarchi** dont l'amour des langues modernes et anciennes m'a beaucoup apporté, et merci à son frère **Jacques** pour son soutien et son amitié. Merci à **Céline Riffard** et son **papa** pour leur aide discrète mais efficace.

Bien évidemment, merci à toute **ma famille** qui m'a beaucoup manqué pendant ces années d'expatriation et merci à toute la **famille Alcocer-Cordellat** qui m'a accueilli et accepté en son sein comme un membre à part entière et qui a fait de mon séjour en Espagne un moment de joie et de bonheur inoubliable.

Enfin, un grand **MERCI** à **Carmen** qui, pendant ces années à Valence puis en Corse pour les derniers mois de rédaction, m'a soutenu et surtout supporté...

*¡muchas gracias mi Carmencita!*

---

# Remerciements institutionnels

---

Je tiens tout particulièrement à remercier toute l'équipe *del Centro de Protección Vegetal y Biotecnología* dont les efforts et le professionnalisme nous ont permis d'avoir de quoi mener à bien nos recherches :

**Luis Navarro**, *Director*

**Pablo Aleza**, *Phylogeny and molecular evaluation*,

**Gema Ancillo**, *Germplasm*,

**Diego Conchilla**, *Greenhouses*,

**José Manuel Conchilla**, *Field evaluation*,

**José Cuenca**, *Molecular and field evaluation*,

**Andrés García-Lor**, *Phylogeny and molecular evaluation*,

**Maria Hernández Garrido**, *Phylogeny and molecular evaluation*

**José Juárez**, *Germplasm and Tissue culture*,

**Vicente Lloris**, *Field evaluation*,

**Rafael Montalt**, *Germplasm*,

**Antonio Navarro**, *Tissue culture*,

**Frédérique Ollitrault**, *Phylogeny and molecular evaluation*

**Carmen Ortega**, *Tissue culture*,

**José Antonio Pina**, *Germplasm and field evaluation*,

**Juan Romero**, *Phylogeny and molecular evaluation*

**Mari Carmen Vives**, *Germplasm*,

---

Un remerciement particulier également au **consortium CITRUSEQ/CITRUSGENN** et à ces membres fondateurs : *l'Instituto Valenciano de Investigaciones Agrarias (IVIA), l'Instituto de Recursos Naturales y Agrobiología de Sevilla (IRNASE), la Fundación de la Comunidad Valenciana Centro de Investigación Príncipe Felipe (CIPF), Fundación Ruralcaja Valencia, Eurosemillas, Source Citrus Geensis SL, GMC Variedades Vegetales A.I.E., Anecoop, et l'ICCSA* pour m'avoir donné accès à leurs données.

La liste suivante reprend les noms des personnes du Consortium Citrusseq/Citrusgenn ayant participé aux différentes phases du séquençage du génome des agrumes ou aux analyses postérieures ayant permis de générer la base de données du Consortium d'où ont été extraites les données utilisées dans le chapitre IV de cette thèse, qu'elles en soient remerciées :

#### **IVIA**

**Manuel Cercós**, *Doctor en Ciencias Biológicas, conceptual work, results interpretation and design;*  
**Francisco Tadeo**, *Doctor en Ciencias Biológicas, conceptual work, results interpretation and design;*  
**Javier Terol**, *Doctor en Ciencias Biológicas, conceptual work, results interpretation and design, matrices, data mining, bioinformatics;*  
**Manuel Talón**, *Doctor en Ciencias Biológicas, conceptual work, results interpretation and design, genome sequencing;*  
**Vicent Arbona**, *Doctor en Ciencias Biológicas, laboratory tasks;*  
**Victoria Ibañez**, *Doctor Ingeniero Agrónomo, conceptual work, results interpretation and design, matrices, data mining;*  
**Amparo Herrero**, *Licenciada Ciencias Biológicas, DNA extractions;*  
**Juan V Muñoz-Sanz**, *Doctor Ingeniero Agrónomo, DNA extractions;*  
**Leandro H. Estornell**, *Doctor en Ciencias Biológicas, DNA extractions;*  
**Isabel López**, *Técnico laboratorio, laboratory tasks;*  
**Ana Almenar**, *Técnico laboratorio, laboratory tasks;*  
**Angel Boix**, *Técnico de campo, field work;*  
**Antonio López-García Usach**, *Ingeniero Técnico Agrícola, conceptual work, results interpretation and design, field work;*  
**Elena Blázquez**, *Técnico laboratorio, laboratory tasks;*  
**Cristina Martínez**, *Técnico laboratorio, laboratory tasks;*  
**Mariano Montoro**, *Técnico de campo, field work;*  
**Juana Ramirez**, *Técnico laboratorio, DNA extractions and laboratory tasks;*  
**Antonio Prieto**, *Técnico laboratorio, DNA extractions and laboratory tasks;*  
**Isabel Sanchís**, *Técnico laboratorio, DNA extractions and laboratory tasks;*  
**Matilde Sancho**, *Técnico laboratorio, DNA extractions and laboratory tasks.*

#### **CIPF**

**Joaquín Dopazo**, *Doctor en Ciencias Biológicas, conceptual work, results interpretation and design, genome sequencing, genome assembling, mapping, variant calling;*  
**Ana Conesa**, *Doctor Ingeniero Agrónomo, conceptual work;*  
**Javier Santoyo**, *Doctor en Ciencias Biológicas, conceptual work;*  
**Ignacio Medina**, *Biólogo e Informático, conceptual work;*  
**David Montaner**, *Estadístico, data mining, bioinformatics;*  
**José Carbonell**, *Ingeniero Informático, genome sequencing, genome assembling, mapping, variant calling, matrices, data mining, bioinformatics;*  
**Roberto Alonso**, *Ingeniero Informático, genome sequencing, genome assembling, mapping, variant calling, matrices, data mining, bioinformatics.*

---

### **ANECOOP**

**Angel del Pino**, *Doctor Ing. Agrónomo, funding and design;*

**Nicola Juste Vidal**, *Técnico, field work;*

**Gerra Sáez Gromaz**, *Administrativa, management;*

**Miguel Martínez**, *Doctor Ing. Agrónomo, funding and design.*

### **ICCSA**

**D. Jorge Garcia**, *Ingeniero Agrónomo, funding;*

**Juan Carlos Claramonte**, *Ing. Agrónomo, conceptual work, results interpretation and design;*

### **EUROSEMILLAS**

**Jose Pellicer**, *Ing Agrónomo, funding, conceptual work, results interpretation and design;*

### **GCM, Variedades Vegetales, A.I. E**

**D. Adrián Agustí**, *Ingeniero Agrónomo, funding;*

**Rosario C. Marín Pastor**, *Ing. Técnico Agrícola, conceptual work, field work.*

### **SOURCE CITRUS GENESIS, SL**

**Josep Jove Estiarte**, *Técnico Superior Comercio Internacional, field work*

**María Paz Fernández Quijada**, *Doctora en Ciencias Biológicas, laboratory tasks;*

**David Alba**, *Ing Agrónomo, funding, conceptual work;*

**Francisco Mañez**, *Técnico Superior, field work;*

### **FUNDACION RURALCAJA**

**José Mariano Aguilar**, *Ing. Técnico Agrícola, field work;*

**Carlos Baixauli Soria**, *Ing. Agrónomo, funding, conceptual work, results interpretation and design;*

**Inma Nájera Juan**, *Ing. Agrónomo, management;*

**Imma Benlloch Tortosa**, *Administrativa, funding, management.*

### **INSTITUTO DE RECURSOS NATURALES Y AGROBIOLOGÍA DE SEVILLA (IRNASE).**

**José Manuel Colmenero-Flores**, *Doctor en Ciencias Biológicas, conceptual work, results interpretation and design.*



---

# Résumé

---

Les études préexistantes identifient quatre taxons de base (*C. reticulata* les mandariniers, *C. maxima* les pamplemoussiers, *C. medica* les cédratiers et *C. micrantha*) à l'origine de l'ensemble des formes cultivées suite à des événements de réticulations. Il en résulte des structures génotypiques complexes, généralement fixées par l'apomixie, fortement hétérozygotes et formées d'une mosaïque de grands fragments chromosomiques d'origines phylogénétiques différentes. La structuration de la variabilité phénotypique suggère que la différenciation initiale des taxons ancestraux est à l'origine d'une part importante de la variabilité utile des agrumes. La connaissance de l'origine des formes cultivées et de leurs structures phylogénomiques est donc indispensable à la bonne gestion des collections et à l'optimisation des programmes d'amélioration génétique. A cette fin, cette thèse explore différentes approches d'analyse de la diversité des génomes. Elle a bénéficié de l'évolution rapide des NGS et propose une utilisation raisonnée des outils disponibles en fonction des questions de recherches. Une analyse plus poussée a été conduite sur les limettiers et citronniers. Le pyroséquençage 454 (Roche) d'amplicons a été utilisé pour décrypter la structure en mosaïque interspécifique du chromosome 2 de 50 variétés à partir d'une information haplotypique multilocus et pour identifier des marqueurs SNP diagnostiques des taxons ancestraux. Ces marqueurs ont permis, en association avec des SSR et indels, d'apporter un nouvel éclairage sur l'origine des limettiers et citronniers, par un génotypage exhaustif des collections Inra/Cirad et Ivia. Enfin, les données de re-séquençage complet Illumina réalisé par le Consortium Citrusseq/Citrusgenn de sept variétés de limettiers et de citronniers comparées à celles de représentants des taxons ancestraux nous ont permis de reconstituer la structure interspécifique de leurs génomes et de schématiser leurs caryotypes phylogénomiques. Les différentes approches ont conduit à des conclusions convergentes. Nos résultats confirment les hypothèses concernant la séquence évolutive à l'origine des bigaradiers (*C. aurantium*), des orangers (*C. sinensis*) et des pomelos (*C. paradisi*) à partir des pools géniques de *C. maxima* et *C. reticulata*. Ils mettent en évidence de fréquentes introgressions de *C. maxima* dans le génome de mandariniers considérées comme représentatifs de *C. reticulata*. Les contributions relatives de ces deux taxons ancestraux aux génomes de nombreuses variétés de petits agrumes (mandariniers, tangors et tangelos) ont pu être estimées. Les limettiers et citronniers résultent de multiples événements de réticulation et *C. medica* est identifié comme parent mâle de la majorité des variétés diploïdes. Deux grands groupes de citronniers, sont différenciés, ceux issus d'hybridations directes *C. reticulata* × *C. medica* et ceux impliquant trois taxons ancestraux (*C. maxima*, *C. reticulata* et *C. medica*). Le bigaradier serait le parent femelle à l'origine des citronniers type Lisbonne (*C. limon*). Les limettiers de type Mexicain (*C. aurantifolia*) seraient issus d'une hybridation directe *C. micrantha* × *C. medica*. Enfin, les limes à gros fruits, triploïdes, ont deux origines. Les types Tahiti résulteraient probablement de la fécondation d'un ovule de citronnier type Lisbonne par un gamète diploïde de limettier type Mexicain. L'autre grand type serait issu d'un backcross entre *C. aurantifolia* (gamète diploïde) et *C. medica*. Ces connaissances sur la structure génomique des espèces secondaires permettent d'envisager une reconstruction d'idéotypes à partir du germplasm des taxons ancestraux. Elles ouvrent également la voie à des études de génétique d'association s'appuyant sur la phylogénomique des gènes impliqués dans l'élaboration des caractères de qualité, de résistance et d'adaptation. Enfin, les marqueurs diagnostiques d'espèces développés trouveront de nombreuses applications pour la caractérisation des collections et diverses études de génétiques.

**Mots clés :** agrumes, haplotype, structure génétique, genome, SNP, NGS, limes, citrons

# Abstract

---

## **Citrus gene pool organization and nuclear genomic interspecific admixture of cultivated citrus**

Citrus fruit, the most important fruit crop in the world, show a wide phenotypic diversity. Previous studies (molecular markers) identified four ancestral taxa (*Citrus reticulata* Blanco, mandarins; *C. maxima* (Burm.) Merr., pummelos; *C. medica* L., citrons; *C. micrantha* Wester, papedas) as the ancestors of all cultivated *Citrus* after reticulate evolutions. As a result, modern citrus varieties have complex and highly heterozygous genotypic structures, generally fixed by apomixis, and formed by a mosaic of large chromosomal fragments of different phylogenetic origins. Furthermore, the structuration of the phenotypic variability suggests that the initial differentiation of the basic taxa is the main source of most of the variability of the useful citrus phenotypic diversity. A thorough knowledge of the origin of cultivated citrus and their phylogenomic structure are essential for the management of biological resources and breeding program optimization.

This thesis explores different approaches for analyzing genome diversity in order to identify the phylogenetic origins of the various horticultural citrus groups and to decipher their phylogenomic genome's structures. We focused on limes and lemons, less studied than most other citrus. This thesis takes advantage of the rapid evolution of NGS and proposes a rational use of available tools, based on research questions. Roche 454 parallel sequencing of amplicons provides multi-loci haplotype information on 500 base fragments. It was used to decipher the interspecific mosaic structure of chromosome 2 for fifty varieties and to identify ancestral taxa diagnostic SNP markers. The genotyping of all limes and lemons of the Inra/Cirad and Ivia germplasm with these markers, in association with SSR and indel markers, allowed to propose new hypothesis on the origins of limes and lemons. Data from Illumina whole genome re-sequencing, done by the Citrusseq/Citrusgenn Consortium, of 7 varieties of limes and lemons, compared to those of representatives of the ancestral taxa, allowed to infer the interspecific structure of their genomes and to map out, for the first time, their phylogenomic karyotypes.

The different approaches led to similar conclusions. Our results confirm previous hypothesis about the evolutionary steps at the origin of sour orange (*C. aurantium*), sweet orange (*C. sinensis*) and grapefruit (*C. paradisi*) involving *C. maxima* and *C. reticulata* gene pools. They highlight frequent introgressions of *C. maxima* in the genome of mandarin varieties despite the fact they were considered as representative of *C. reticulata*. We were also able to quantify the relative proportions of these two ancestral taxa in the genome of many varieties of small citrus fruit (mandarin hybrids, tangors and tangelos). Our work on limes and lemons demonstrate that *C. medica* is the male parent of this varietal group at the diploid level. Two groups of lemons are clearly differentiated: one from direct hybridizations between *C. reticulata* and *C. medica*, and one from crosses between hybrids (*C. maxima* × *C. reticulata*) and *C. medica*. Sour orange seems to be the female parent of 'Eureka' type lemons (*C. limon*). The 'Mexican' type limes (*C. aurantifolia*) seems to come from a direct hybridization *C. micrantha* × *C. medica*. Finally, triploid big fruit limes have two major origins. The 'Tahiti' type probably results from an 'Eureka' type lemon (*C. limon*) ovule fecundated by a diploid gamete of a 'Mexican' type lime (*C. aurantifolia*), while the other type would come from a back-cross between *C. aurantifolia* (diploid gamete) and *C. medica*.

This new insights in genomic structure of secondary species makes to consider possible a reconstruction of these ideotypes from ancestral taxa germplasm. They also open new ways for association genetic studies based on phylogenomics of genes involved in the development of quality, resistance and adaptation traits. Finally, developed specific taxa diagnostic markers will find many applications for the characterization of collections and further genetic studies.

**Keywords:** Citrus, haplotype, evolution, SNP, NGS, genome admixture, lime, lemon

**Instituto Valenciano de Investigaciones Agrarias, Lab. Cirad-Ivia, Centro de Protección Vegetal y Biotecnología, Carretera CV-315, Km 10,7 - ES-46113 Moncada (Valence, Espagne)**

# Resumen

## Organización de las especies de cítricos y determinación de su estructura genómica nuclear inter-específica

Los cítricos, el cultivo frutal más importante a nivel mundial, muestran una diversidad fenotípica muy variada. Estudios previos mediante marcadores moleculares identificaron cuatro taxones ancestrales (*Citrus reticulata* Blanco, mandarinas; *C. maxima* (Brum.) Merr., zamboa; *C. medica* L., cidros; *C. micrantha* Wester, papedas) que han dado lugar a todos los cítricos cultivados a través de una evolución reticulada. Como resultado, las variedades modernas de cítricos tienen unas estructuras genotípicas altamente complejas y heterocigóticas, generalmente fijadas por la apomixis, y formadas por fragmentos cromosómicos de gran tamaño en forma de mosaico procedentes de diferentes orígenes filogenéticos. Es más, la estructuración de la variabilidad fenotípica sugiere que la diferenciación inicial de los taxones básicos es la fuente principal de la mayoría de la variabilidad de la diversidad fenotípica útil en los cítricos. Un mayor conocimiento del origen de los cítricos cultivados es fundamental para el manejo de los recursos biológicos y la optimización de los programas de mejora genética.

Esta tesis explora diferentes aproximaciones para el análisis de la diversidad del genoma para identificar los orígenes filogenéticos de los distintos grupos de cítricos cultivados y determinar sus estructuras genómicas. Nos hemos centrado en las limas y limones, menos estudiados que el resto de cítricos. Esta tesis se aprovecha de las ventajas de la rápida evolución en las técnicas de secuenciación para su uso en la respuesta de cuestiones científicas. La secuenciación de amplicones mediante la técnica de Roche 454 proporciona la información de haplotipos en fragmentos de 500 pares de bases. Esta técnica se utilizó para determinar la estructura inter-específica en mosaico del cromosoma 2 en 50 variedades y así identificar marcadores moleculares (SNP) diagnóstico de los taxones ancestrales. El genotipado de limas y limones de los bancos de germoplasma del Inra/Cirad y del Ivia mediante dichos marcadores, combinados con marcadores SSR e indel, permitió proponer nuevas hipótesis sobre el origen de las limas y limones. Datos procedentes de la re-secuenciación del genoma de siete variedades de limas y limones mediante Illumina, comparados con los procedentes de representantes de los taxones ancestrales, permitieron determinar la estructura inter-específica de sus genomas y mapear, por primera vez, sus cariotipos filogenómicos.

Las diferentes aproximaciones dieron lugar a conclusiones similares. Nuestros resultados confirman hipótesis previas sobre el origen del naranjo amargo (*C. aurantium*), naranjo dulce (*C. sinensis*) y pomelo (*C. paradisi*) que involucran a *C. maxima* y *C. reticulata*. Además, hemos observado que variedades de mandarino consideradas como representantes de *C. reticulata* presentan introgresión de *C. maxima*. También hemos sido capaces de cuantificar las proporciones relativas de estos dos taxones ancestrales en el genoma de distintas variedades de pequeños cítricos (híbridos de mandarino, tangors y tangelos). Nuestro trabajo en limas y limones demuestra que *C. medica* es el parental masculino de este grupo varietal a nivel diploide. Dos grupos de limones se diferencian claramente: uno procedente de hibridaciones directas entre *C. reticulata* y *C. medica*, y otro a partir de cruces entre híbridos (*C. maxima* x *C. reticulata*) y *C. medica*. El naranjo amargo parece ser el parental femenino de híbridos tipo 'Eureka' (*C. limon*). Las limas tipo 'Mejicana' (*C. aurantifolia*) parecen proceder de la hibridación directa entre *C. micrantha* y *C. medica*. Finalmente, frutos de limas triploides grandes tienen dos orígenes principales: las de tipo 'Tahiti' probablemente resultaron de un óvulo fecundado de un limón tipo 'Eureka' (*C. limon*) por un gameto diploide de lima tipo 'Mejicana' (*C. aurantifolia*), mientras que el otro tipo procedería de un retrocruce entre *C. aurantifolia* (gameto diploide) y *C. medica*.

Estos resultados novedosos sobre la estructura genómica de las especies secundarias permiten considerar como posible la reconstrucción de estos ideotipos a partir de los taxones ancestrales dentro del germoplasma. Estos también abren nuevas vías para estudios de genética de asociación basados en filogenómica de genes relacionados con la calidad, resistencia y caracteres de adaptación. Finalmente, los marcadores de diagnóstico específico de taxón darán lugar a numerosas aplicaciones para la caracterización de colecciones y futuros estudios genéticos.

**Palabras clave:** cítricos, haplotipo, evolución, SNP, NGS, genoma, lima, limón

# Resum

## Organització de les espècies de cítrics i determinació de la seua estructura genòmica interespecífica

Els cítrics, el cultiu fruiter més important del món, mostren una gran diversitat fenotípica. Estudis previs mitjançant marcadors moleculars identificaren quatre taxons ancestrals (*Citrus reticulata* Blanco, mandarines; *C. maxima* (Brum.) Merr., pummelos; *C. medica* L., poncilers; *C. micrantha* Wester, papedas) com a parentals de tots els cítrics cultivats després d'un període d'evolució reticulada. Com a resultat, les varietats de cítrics modernes tenen unes estructures genotípiques complexes i molt heterozigòtiques, generalment fixades per apomixis, i formades per un mosaic de llargs fragments cromosòmics de diferent origen filogenètic. A més a més, l'estructura de la variabilitat fenotípica suggereix que la diferenciació inicial dels taxons ancestrals és la principal font principal de la major part de la variabilitat fenotípica útil dels cítrics. Un coneixement més a fons de l'origen dels cítrics cultivats i de la seua estructura filogenòmica son essencials per a l'utilització dels recursos biològics i l'optimització dels programes de millora genètica.

Aquesta tesi explora diferents aproximacions per a l'anàlisi de la diversitat del genoma amb l'objectiu d'identificar els orígens filogenètics dels diversos grups de cítrics i determinar les seues estructures filogenòmiques. Nosaltres ens hem centrat en les llimeres i llimoners, menys estudiats que altres cítrics. Aquesta tesi aprofita la ràpida evolució de les tècniques de seqüenciació i proposa un ús raonable de les eines disponibles, basant-se en qüestions científiques. La seqüenciació paral·lela d'amplicons de Roche 454 proporciona una informació d'haplotipus en fragments de 500 parelles de bases. Es va utilitzar per a determinar l'estructura interespecífica en mosaic del cromosoma 2 en 50 varietats i per a identificar marcadors SNP diagnòstics. El genotipatge de les llimeres y llimoners dels bancs de germoplasma de l'Inra/Cirad i de l'Ivia mitjançant aquests marcadors, en combinació amb SSR e indel, ha permès proposar noves hipòtesis sobre els seus orígens. La informació generada amb la re-seqüenciació amb Illumina de genomes sencers de set varietats de llimeres i llimoners, comparada amb els representants dels taxons ancestrals, va permetre determinar l'estructura interespecífica dels seus genomes i mapejar, per primera volta, els seus cariotipus filogenòmics.

Les diferents aproximacions portaren a conclusions similars. Els nostres resultats confirmen hipòtesis prèvies de l'origen del taronger amarg (*C. aurantium*), taronger dolç (*C. sinensis*) i pomelo (*C. paradisi*) on estan involucrats *C. maxima* i *C. reticulata*. També destaquen les freqüents introgressions de *C. maxima* en el genoma de mandarines que eren considerades com a representatives de *C. reticulata*. Hem pogut quantificar la proporció relativa d'aquests dos taxons ancestrals en el genoma de moltes varietats dels cítrics xicotets (híbrids de mandarina, tangors i tangelos). El nostre treball en llimeres y llimoners demostra que *C. medica* és el parental masculí d'aquest grup varietal a nivell diploide. Dos grups de llimoners estan clarament diferenciats: un d'hibridacions directes entre *C. reticulata* i *C. medica*, i un de creuaments entre híbrids (*C. maxima* x *C. reticulata*) i *C. medica*. El taronger amarg pareix ser el parental femení dels llimoners tipus 'Eureka' (*C. limon*). Les llimeres 'Mexicanes' (*C. aurantifolia*) pareixen haver sigut originades per una hibridació directa *C. micrantha* i *C. medica*. Finalment, les llimeres grans triploides tenen dos orígens principals. El tipus 'Tahiti' probablement prové d'un òvul fecundat de tipus 'Eureka' (*C. limon*) per un gamet diploide de llima tipus 'Mexicana' (*C. aurantifolia*), mentre que l'altre tipus pareix vindre d'un retrocreuament *C. aurantifolia* (gamet diploide) i *C. medica*.

Estos nous resultats de l'estructura genòmica de les espècies secundaries porta a considerar la possibilitat de reconstruir els idiotipus a partir dels taxons ancestrals. Ells també obren nous camins per a estudis de genètica d'associació basats en gens involucrats en el desenvolupament de la qualitat, resistència i adaptació. Finalment, els marcadors desenvolupats en la tesi trobaran moltes aplicacions per a la caracterització de les col·leccions de germoplasma y futurs estudis genètics.

**Paraules claus:** cítrics, haplotip, evolució, SNP, NGS, genoma, llima, llimó



# Table des matières

<b>Dédicace</b>	
<b>Remerciements personnels</b>	
<b>Remerciements institutionnels</b>	
<b>Résumé en français</b>	
<b>Abstract (résumé en anglais)</b>	
<b>Resumen (résumé en espagnol)</b>	
<b>Resum (résumé en valencien)</b>	
<b>Table des abréviations</b>	16
<b>Introduction</b>	18
1. Les agrumes : origines et diffusion	19
1.1. Les noms	19
1.2. Les origines	21
1.3. La diffusion	23
2. Les agrumes : botanique et origines génétiques	26
2.1. Ordre des <i>Sapindales</i> , famille des <i>Rutaceae</i>	26
2.2. Les agrumes vrais : six genres botaniques	27
2.2.1. <i>Clymenia</i> Swingle	27
2.2.2. <i>Eremocitrus</i> Swingle	28
2.2.3. <i>Microcitrus</i> Swingle	28
2.2.4. <i>Poncirus</i> Raf.	28
2.2.5. <i>Fortunella</i> Swingle	29
2.2.6. <i>Citrus</i> L.	30
2.3. Le genre <i>Citrus</i>	30
2.3.1. Les taxons de base	30
2.3.1.1. Les cédratiers	32
2.3.1.2. Les mandariniers	33
2.3.1.3. Les pamplemoussiers	34
2.3.1.4. Les papedas	35
2.3.2. Origine génétiques des espèces cultivées	35
2.3.2.1. L'oranger ( <i>Citrus sinensis</i> (L.) Osb.)	36
2.3.2.2. Le bigaradier ou oranger amer ( <i>C. aurantium</i> L.)	37
2.3.2.3. Le pomelo ( <i>C. paradisi</i> Macf.)	37
2.3.2.4. Le clémentinier ( <i>C. clementina</i> Hort. ex Tan.)	37
2.3.3. Le cas particulier des limetiers et des citronniers	37
2.3.3.1. Les citronniers ( <i>C. limon</i> )	39
2.3.3.2. Les limettiers ( <i>C. aurantifolia</i> )	39
2.3.3.3. Les autres limettiers et citronniers	40
3. Les agrumes : ressources génétiques	42
3.1. Les centres de ressources phytogénétiques	42
3.2. La conservation des ressources génétiques agrumes	44
3.3. La gestion des ressources génétiques agrumes	45
4. Les agrumes : Importance socio-économique	46
4.1. Le marché des agrumes	46
4.1.1. La production	46
4.1.2. Les surfaces	48
4.1.3. Les exportations	49
4.1.4. Les importations	50
4.2. Cas particulier des citrons et des limes	50
4.2.1. Le citron	52
4.2.2. Les limes	54
5. Agrumiculture moderne : les contraintes	56
5.1. Les stress abiotiques	56

5.2. Les stress biotiques .....	56
6. Agrumiculture moderne : les apports de l'amélioration des plantes en réponse aux enjeux et contraintes .....	59
6.1. Les objectifs d'amélioration et de sélection des porte-greffe .....	59
6.2. Les objectifs de l'amélioration des scions .....	60
6.3. Les facteurs limitant l'amélioration génétique chez les agrumes .....	60
6.4. Les principales stratégies d'amélioration .....	62
6.4.1. La sélection de mutants spontanés et la mutagenèse induite .....	62
6.4.2. La recombinaison sexuée au niveau diploïde .....	62
6.4.3. La polyploïdie .....	65
6.4.4. Les cybrides .....	66
6.4.5. La transgénèse .....	66
6.4.6. La mutagenèse dirigée .....	66
6.4.7. La reconstruction des structures génomiques hybrides .....	67
7. Les Agrumes : évolution des outils moléculaires .....	68
7.1. Analyse des génomes nucléaires .....	68
7.1.1. AFLP ( <i>Amplification Fragment Length Polymorphism</i> ) .....	68
7.1.2. CAPS ( <i>Cleaved Amplified Polymorphic Sequences</i> ) <sup>57</sup> ou PCR-RFLP .....	68
7.1.3. Indel ( <i>Insertions - délétions</i> ) .....	69
7.1.4. IRAP ( <i>Inter-Retrotransposon Amplified Polymorphism</i> ) .....	69
7.1.5. ISSR ( <i>Inter Simple Sequence Repeat</i> ) .....	69
7.1.6. RAPD ( <i>Random Amplified Polymorphic DNA</i> ) .....	69
7.1.7. RFLP ( <i>Restriction Fragment Length Polymorphism</i> ) .....	70
7.1.8. SNP ( <i>Single Nucleotide Polymorphisms</i> ) .....	70
7.1.9. SSR ( <i>Simple Sequence Repeats</i> ) .....	71
7.2. Marqueurs cytoplasmiques .....	72
7.3. Données génétiques et génomiques disponibles .....	74
7.4. Les promesses des NGS .....	76
7.4.1. Séquençage à haut débit, de deuxième génération .....	76
7.4.1.1. Pyroséquençage (454 Roche) .....	76
7.4.1.2. Séquençage à l'aide de terminateurs réversibles (Illumina) .....	76
7.4.1.3. Séquençage par ligature ( <i>ligation</i> ) (Applied Biosystems) .....	77
7.4.1.4. séquençage par semi-conducteur ( <i>Ion Torrent</i> ) .....	77
7.4.2. Séquençage de troisième génération (TGS) .....	78
7.5. Exploiter et intégrer les données des séquençages .....	78
7.5.1. Biopython .....	79
7.5.2. South Green .....	79
7.5.3. France Génomique .....	79
8. Objectifs et activités de thèse .....	80
<b>Chapitre I</b> .....	<b>84</b>
<i>Next generation haplotyping to decipher the nuclear genomic interspecific admixture in Citrus species; analysis of the Chromosome 2</i>	
<b>Chapitre II</b> .....	<b>111</b>
<i>Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern Citrus varieties</i>	
<b>Chapitre III</b> .....	<b>143</b>
<i>Nuclear, mitochondrial and chloroplastic SSRs, InDels and SNP markers reveal the diversity and phylogenetic origin of limes and lemons</i>	
<b>Chapitre IV</b> .....	<b>173</b>
<i>Deciphering the interspecific genome structure of limes and lemons by whole-genome re-sequencing</i>	
<b>Discussion</b> .....	<b>205</b>
<b>Conclusion</b> .....	<b>226</b>
<b>Références bibliographiques</b> .....	<b>229</b>

<b>Annexes</b>	260
<b>Annexes I : Articles additionnels et contributions à congrès en relation avec la thèse</b>	261
Annexe I.1. <i>Multilocus Haplotyping by Parallel Sequencing to Decipher the Interspecific Mosaic Genome Structure of Cultivated Citrus (Curk et al.; in press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012)</i>	262
Annexe I.2. <i>Multilocus SNPs analysis allows phylogenetic assignation of DNA fragments to decipher the intraspecific mosaic genome structure of cultivated Citrus (Curk et al.; 2011; Plant Genome Evolution Congress 2011, 4 - 6 Sep 2011, P2.21)</i>	263
Annexe I.3. <i>Two-step PCR targeted amplicon library for next generation haplotyping; an example in Citrus (Curk et al.)</i>	264
Annexe I.4. <i>New Insights on Limes and Lemons Origin from Nuclear and Cytoplasmic Markers Genotyping and Targeted Nuclear Gene Sequencing (Curk et al.; in press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012 and poster)</i>	288
Annexe I.5. <i>Comparative values of SSRs, SNPs and InDels for citrus genetic diversity analysis (Ollitrault et al.; in press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012; S02P05)</i>	289
Annexe I.6. <i>A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species (Garcia-Lor et al. 2013b; Annals of Botany; 111 (1): 1-19.; doi: 10.1093/aob/mcs227; First published online: October 26, 2012)</i>	290
<b>Annexes II</b>	291
<i>Supplementary information – Chapitre I</i>	
<b>Annexes III</b>	313
<i>Supplementary information – Chapitre II</i>	
<b>Annexes IV</b>	341
<i>Supplementary information – Chapitre III</i>	
<b>Annexes V</b>	363
<i>Supplementary information – Chapitre IV</i>	



## Table des abréviations

ACP	Analyse en composantes principales
ADN	Acide désoxyribonucléique
AFC	Analyse factorielle des correspondances
AFLP	<i>Amplification Fragment Length Polymorphism</i> (polymorphisme de longueur des fragments amplifiés)
ARN	Acide ribonucléique
AGL	<i>Programa AGL : Recursos y tecnologías Agroalimentarias</i>
BC	<i>Backcross</i> (rétrocroisement ou croisement en retour)
CA	<i>Correspondence analysis</i>
CAPS	<i>Cleaved Amplified Polymorphic Sequences</i> ou PCR-RFLP
CCD	<i>Charge-Coupled Device</i> (dispositif à transfert de charge)
CEA	Commissariat à l'énergie atomique et aux énergies alternatives
CEVd	<i>Citrus exocortis viroid</i>
CIPF	<i>Centro de Investigación Príncipe Felipe</i>
Cirad	<i>Centre de coopération internationale en recherche agronomique pour le développement</i>
CNRS	Centre national de la recherche scientifique
CPsV	<i>Citrus psorosis virus</i>
CT	type cytoplasmique
CTV	<i>Citrus tristeza virus</i>
DNA	<i>Deoxyribonucleic acid</i>
EST	<i>Expressed sequence tag</i> (marqueur de séquence exprimée)
FAO	Organisation des Nations Unies pour l'alimentation et l'agriculture
HLB	Huanglongbing
FEDER	Fonds européen de développement régional
GATK	<i>Genome Analysis Toolkit</i>
ICCSA	<i>Investigación Citrícola Castellón SA</i>
ICGC	<i>International Citrus Genome Consortium</i>
ICVN	<i>International citrus variety numbering</i>
Indel	Insertions - délétions
Inra	Institut national de la recherche agronomique
Inserm	Institut national de la santé et de la recherche médicale
IRAP	<i>Inter-Retrotransposon Amplified Polymorphism</i>
IRD	Institut de recherche pour le développement
Irnase	<i>Instituto de Recursos Naturales y Agrobiología de Sevilla</i>
ISSR	<i>Inter Simple Sequence Repeat</i>
Ivia	<i>Instituto Valenciano de Investigaciones Agrarias</i>
LD	Déséquilibre de liaison
MID	<i>Multiplex identifier</i>
MLG	<i>Multi Locus Genotypes</i>
NGS	<i>New Generation Sequencing</i> (séquençage de nouvelle génération)
NJA	<i>Neighbour joining analysis</i>
OGM	Organisme génétiquement modifié
PCA	<i>Principal component analysis</i>
PCR	Réaction en chaîne par polymérase
RAPD	<i>Random Amplified Polymorphic DNA</i>

RFLP	<i>Restriction Fragment Length Polymorphism</i> (polymorphisme de longueur des fragments de restriction)
RNA	<i>Ribonucleic acid</i>
SNP	<i>Single Nucleotide Polymorphisms</i> (Polimorphime nucléotidique)
SRA	Station de Recherches Agronomiques
SSCP	<i>Single Strand Conformation Polymorphisms</i>
SSR	<i>Simple Sequence Repeats</i> (séquence microsatellite)
TGS	<i>Third-generation sequencing</i> (Séquençage de troisième génération)
UMR Agap	Unité mixte de recherche, Amélioration génétique et adaptation des plantes méditerranéennes et tropicales
UMS	Unité mixte de service

---

# Introduction

---

## 1. Les agrumes : origines et diffusion

*Kennst du das Land, wo die Zitronen blühen,  
Im dunkeln Laub die Goldorangen glühen,  
Ein sanfter Wind vom blauen Himmel weht,  
Die Myrte still und hoch der Lorbeer steht?  
Kennst du es wohl? Dahin!  
Dahin möcht' ich mit dir,  
O mein Geliebter, ziehen.*

Connais-tu le pays où les citronniers fleurissent,  
Les oranges d'or dans le sombre feuillage flamboient,  
Un doux zéphyr souffle dans l'azur du ciel,  
Où poussent le calme myrte et le grand laurier ?  
Le connais-tu bien ?  
Là-bas ! Là-bas  
Je voudrais aller avec toi, mon amour.

Johann Wolfgang von Goethe (1795)

« Connaissez-vous le pays où les citronniers fleurissent, (...) » nous chante Mignon en évoquant l'Italie dans *Les Années d'apprentissage de Wilhelm Meister* de Goethe (1795). Pour beaucoup d'Européens, les agrumes sont le symbole du soleil en général et de la Méditerranée en particulier. Pourtant, leur origine est bien plus exotique et, même si leurs extraordinaires couleurs et leur parfums particuliers ont toujours attiré l'attention des voyageurs depuis des temps immémoriaux, leur distribution à travers le monde a été relativement lente (Webber *et al.*, 1967).

### 1.1. Les noms

A travers l'étymologie des noms d'agrumes les plus communs nous pouvons retrouver une trace de leur histoire. Ainsi, le mot « agrumes » en français (*agrumi* en italien ou *agrum* en occitan), dérive du latin médiéval *acrumen* qui désignait les fruits acides (*acer* : âcre, aigre en latin classique) (Couplan, 2012). Il s'utilise au pluriel et désigne indifféremment les arbres et les fruits. Il n'est réellement adopté dans le langage courant qu'au XX<sup>e</sup> siècle ; il est absent des dictionnaires généralistes avant 1930 et entre à l'Académie Française en 1940. C'est le développement économique de la production d'agrumes qui généralise ce mot pourtant aujourd'hui connu de tous. Les Anglais et les Espagnols se distinguent des Français et Italiens en n'utilisant pas le terme « agrumes » mais celui de *citrus* et son dérivé *citricos* (Salette, 2000) qui est, semble-il, issu du grec ancien κέδρος (Kédros, cèdre) et qui a donné *citrus* en latin puis cédrat et citron en français et *citron* en anglais (cédrat). *Citrus* en anglais génère une certaine confusion notamment au moment de différencier les agrumes « citrus » (en romain et sans majuscule en anglais) comprenant plusieurs genres et espèces et le nom de genre utilisé en latin « *Citrus* » (en italique et avec majuscule) ne représentant qu'une partie des agrumes.

Le mot « orange » viendrait lui du mot *nar*, parfum, puis *naranga* en sanscrit, repris en langue perse puis en arabe, d'où *naranja* en espagnol ou *arancia* en italien (à la suite de la disparition du « n » initial du fait de l'utilisation de l'article « *un'arancia* »), et enfin « orange » en anglais et en français, certainement influencé par le nom de la ville d'Orange et de la couleur de l'or. C'est d'ailleurs le fruit qui donna son nom à la couleur et non l'inverse. Le mot orange désignait autrefois l'orange amère ou bigarade (*Citrus aurantium* L.) et ce n'est qu'après l'arrivée de l'orange douce (*Citrus sinensis* (L.) Osb.) en occident que le fruit amer adopta son nom actuel hérité du provençal *bigarrat* (CNRS - ATILF, 2014).

Le mot « limon » qui désignait autrefois les citrons acides est arrivé en France au XIV<sup>e</sup> siècle de l'italien *limone* après avoir voyagé de l'Asie à l'Europe : *nimbú* désigne un type

d'agrumes en sanscrit et a donné *limún* en persan et *laymún* en arabe qui a donné *lima* en espagnol. C'est de ce *lima* espagnol que sont issus les noms de « lime » et « limette », introduits au XVI<sup>e</sup> siècle dans la langue française, et c'est des limons et limes français que sont nés les *lemons and limes* de la langue anglaise (Salette, 2000).

Le mot « mandarine » vient de « mandarin », nom donné au XVI<sup>e</sup> siècle par les Portugais aux lettrés-fonctionnaires de l'Empire chinois (Larousse) ; les mandarines ayant la couleur orange d'or de la tunique des mandarins, elles en adoptèrent le nom.

Le mot français « pamplemousse » vient du néerlandais *pompelmoes* (*pompe* : enflé ou gros comme une citrouille *pompoen*, et *limoes* : agrumes ou citron venant du portugais ou du malais) (Chauvet, 2001). Comme son nom l'indique, il s'agit d'un agrume de taille respectable et dont l'épaisseur de peau lui confère une excellente capacité de conservation. Le naturaliste néerlandais George Everhard Rumphius (1627-1702), auteur du *Cabinet de curiosité d'Ambon* (Rumphius, 1999 (1661)) nous indique qu'il s'agit « d'un fruit excellent pour les voyages en mer, car on peut le conserver longtemps sans qu'il pourrisse » ; caractéristique qui permis aux navigateurs néerlandais d'utiliser les pamplemousses comme source de vitamine C (antiscorbutique) lors de leurs longs voyages et ainsi de diffuser à partir d'Asie le fruit et son nom. Les « vrais » pamplemousses (*Citrus maxima* (burm.) Merr.) sont souvent confondus avec les pomelos (*Citrus paradisi* Macf.). Aux États-Unis le mot *pomelo* fut d'abord proposé pour désigner les *Citrus paradisi*, nouveaux hybrides découverts dans les Caraïbes au XVIII<sup>e</sup> siècle, mais ce nom fut vite remplacé dans le commerce par *grapefruit* (Chauvet, 2001) probablement du fait que, contrairement aux pamplemousses, les pomelos poussent en grappes (figure 1). En français, les agronomes d'Algérie, qui connaissaient le pamplemousse exotique, déjà présent en Méditerranée, adoptèrent le mot « pomelo » ou « pomélo », emprunté de l'anglais, pour distinguer les deux types de fruits (Chauvet, 2001). Pourtant les pomelos sont encore aujourd'hui très souvent qualifiés de pamplemousses au grand dam des spécialistes.



Figure 1 : Grappe de pomelo Star Ruby © F. Curk

L'origine du nom de la bergamote est plus controversée, il semble qu'il y ait deux origines possibles. Soit la ville d'Asie mineure de *Bergama*, forme arabo turque de *Pergamo*, aurait donné son nom à ce fruit au parfum tant recherché, soit sa forme de poire aurait évoqué « la poire du prince », *beg-armudi* ou *bey armudu* en turque (Barnhart, 1988), qu'évoquait Rabelais (1546) dans le *Tiers Livre* : « Vous mangerez bonnes poyres Crustemenies, & Berguamottes » (Rabelais, 1546; CNRS - ATILF, 2014).

Enfin la clémentine, née au début du XX<sup>e</sup> siècle, doit son nom à la Société Algéroise d'agriculture qui baptisa ce nouveau fruit « Clémentine » en l'honneur de son sélectionneur le Père Clément (de son vrai nom Vital Rodier, 1839-1904) (Trabut, 1902).

## 1.2. Les origines

Le Sud-Est asiatique, allant des versants sud de l'Himalaya jusqu'à l'Indonésie en passant par la Chine, le Vietnam, la Thaïlande et l'archipel malaisien, est aujourd'hui considéré comme le véritable berceau originel des agrumes. Dans cette zone, la culture de certains agrumes remonterait à plusieurs millénaires. En Chine, les agrumes seraient cultivés depuis plus de 4 000 ans. La plus ancienne référence manuscrite connue serait citée dans un texte de l'un des 5 *classiques* attribués à Confucius, le *Shu Jing*, appelé aussi *Livre des Histoires* ou *Classique des documents*, qui compile des écrits remontant jusqu'au III<sup>e</sup> millénaire av. J.-C. Il y serait expliqué que deux types d'agrumes, un grand et un petit, devaient être remis comme offrandes à l'empereur (XXIII<sup>e</sup> siècle av. J.-C.). Il est bien évidemment impossible d'identifier ces deux fruits mais d'après la description qui en est faite, il pourrait s'agir de pamplemousses (*C. maxima*) et de mandarines (*C. reticulata*) (Cooper, 1982; Deng, 2008). Ce sont également les Chinois qui revendiquent la première monographie attestée sur les agrumes, le *Ju Lu*, qui daterait de 1178 et écrit par Han Yanzhi. Dans cet ouvrage 27 agrumes y seraient décrits ainsi que les différentes étapes de leur culture, de la propagation à la récolte des fruits (Deng, 2008).

Le cédrat (*C. medica* L.) (figure 2) était sacré en Inde et associé à la divinité à tête d'éléphant, Ganesh, dieu de la sagesse et de l'intelligence. Chez les bouddhistes le dieu des richesses, Buberu, est parfois représenté avec un cédrat dans une main (Scora, 1975). Le cédrat serait le premier agrume importé en occident au III<sup>e</sup> siècle av. J.-C. Certains auteurs expliquent que son introduction en Grèce aurait coïncidé avec l'épopée d'Alexandre le Grand jusqu'à la vallée de l'Indus où le cédratier était cultivé. Appelé à l'époque « Pomme de Perse » ou « Pomme de Médie », le cédrat fut décrit par Théophraste (372 av. J.-C., 288 av. J.-C.) dans son *Histoire des plantes*.

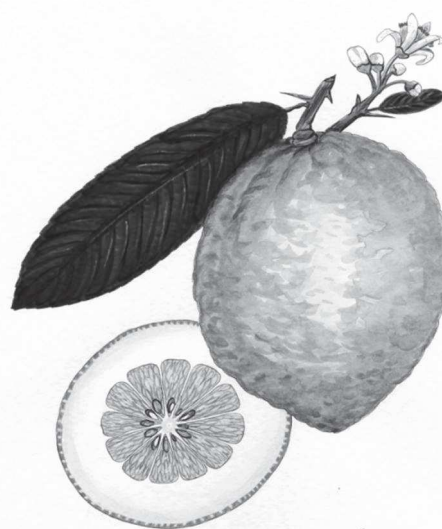


Figure 2 : Cédrat de Corse (aquarelle) ;  
© F. Curk

Ce fruit a rapidement conquis le reste de la Méditerranée. Il n'était pas consommé, mais ses utilisations cosmétiques ou pharmaceutiques étaient réputées. Malgré la présence attestée du cédrat en Méditerranée dans les écrits de Théophraste, il semble qu'il n'y ait pas d'agrumes parmi les plus de 200 plantes citées dans la bible (Moldenke et Moldenke, 1952). Pourtant, certains auteurs (Isaac, 1959) pensent reconnaître le cédrat dans le Lévitique 23:40 « Vous prendrez, le premier jour, du **fruit des beaux (Hadar) arbres**, des branches de palmiers, des rameaux d'arbres touffus et des saules de rivière ; et vous vous réjouirez devant l'Éternel, votre Dieu, pendant sept jours ». Le plus bel arbre (*Hadar*, traduisible en français par beaux, majesté, splendeur, magnificence) est donc reconnu comme étant le cédratier qui donne le fruit sacré des Juifs qui s'offre encore aujourd'hui (figure 3), et depuis 136 av. J.-C., pour la fête des Tabernacles (Loret, 1891; Nicolosi *et al.*, 2005). À cette époque (II<sup>e</sup> siècle av. J.-C.) la culture du cédrat semble répandue de la Perse à la Palestine et sa diffusion suit celle de la diaspora juive

dans tout le bassin méditerranéen. Au premier siècle après J.-C. le cédrat serait présent au sud de l'Italie et sur ses îles ainsi qu'en Corse (Praloran, 1971).

À l'exception du cédrat, les origines et les phénomènes d'expansion des agrumes restent mal connus. La difficulté est certainement liée au fait que, comme tout événement historique, les traces écrites sont généralement associées à des événements guerriers souvent sujets à des approximations ou des mensonges

historiques, ou propagandes, visant à dénigrer ou rabaisser les avancées techniques ou scientifiques des peuples conquis. De plus, un grand nombre de traces non écrites (fresques, peintures, dessins, sculptures...) sont peu ou difficilement exploitables comme document historique fiable. Ainsi la présence antique d'agrumes en Égypte et en Italie semble-t-elle probable même si non encore prouvée et serait en contradiction avec l'affirmation que le tout premier agrume introduit en Méditerranée fut le cédrat, en Grèce, au III<sup>e</sup> siècle av. J.-C. Victor Loret (1891; 1892) affirme avoir identifié des agrumes sur des peintures du temple de Karnac alors que d'autres auteurs les considèrent comme difficilement identifiables (Praloran, 1971). Il cite également un fruit momifié décrit comme étant un agrume et conservé au musée du Louvre (Loret, 1891; 1892).

Un article récent (Pagnoux *et al.*, 2013) évoque la présence du cédratier mais également du citronnier en Italie au III<sup>e</sup> et II<sup>e</sup> siècle av. J.-C. en se basant sur l'étude de semences et de grains de pollen retrouvés dans des fouilles archéologiques réalisées à Rome et Pompéi. Pompéi, où l'on peut observer, sur une mosaïque, des fruits oranges qui pourraient être des agrumes (oranges ou bigarades) et qui donc, y auraient été cultivés avant la destruction de cette ville en 79 ap. J.-C. (Luro *et al.*, 2013). De même, il est intéressant d'évoquer une statue d'Hercule en bronze doré du musée Capitolini de Rome, représentant Hercule tenant à la main les pommes d'or qu'il vient de voler du jardin des Hespérides, après avoir tué Ladon, le dragon qui en gardait l'entrée. Les pommes d'or du jardin des Hespérides, souvent considérées comme des pommes ou des coings, ont souvent été comparées à des oranges à tel point que l'on considère les agrumes comme constituant le groupe des hespéridées (*hesperideae*) et que l'on nomme les fruits d'agrumes « hespéridies » (Lecoq et Juillet, 1831; Praloran, 1971). Même si rien dans la littérature ne nous permet d'identifier la nature de ces fruits gardés précieusement par les muses Hespérides, certains artistes ont donc fait le choix de les représenter comme des oranges et parfois, à des époques inattendues. C'est le cas pour les fruits que l'on découvre dans la main de la statue d'Hercule du musée Capitolini (figure 4a) évoquée par Ferrari en 1646 : *Equidem non unius duntaxat in Herculis capitolini, sed in aliorum quoque complurium simulacris id genus poma spectari facili dederim: Quod arbitrariam secuti sententiam*



Figure 3 : Sélection et conditionnement de cédrats (variété Diamante) fraîchement récoltés en vue de leur expédition à New York où ils seront vendus pour la fête des Tabernacles (Santa Maria del Cedro, Calabre, Italie, 2008) ; © F. Curk-Inra.

*veteres artifices mala cotonea, citreis quippe notiora & imitationi opportuniora, praetulerint*<sup>1</sup>. Ces fruits ne ressemblent en rien à des pommes ni à des coings mais ont bien une forme évoquant les quartiers d'un agrume ; un cédrat, toujours selon Ferrari (1646), puisque l'orange, qui pourtant se prête mieux à la comparaison (figure 4b), était, et est encore, considérée comme absente du bassin méditerranéen au II<sup>e</sup> siècle apr. J.-C., époque à laquelle aurait été réalisée cette statue.



Figure 4 : Statue d'Hercule en bronze doré, 241 cm, sculpture du II<sup>e</sup> siècle apr. J.-C., Musée *Capitolini*, Rome (Italie) ; a. vue d'ensemble de la statue ; b. Détail de la main d'Hercule avec les pommes d'or, fruits du jardin des Hespérides ;  
© Nuria Duran Vila.

### 1.3. La diffusion

Malgré ces débats, il est généralement admis que plusieurs siècles se seraient écoulés avant que ne soient introduites en Occident d'autres variétés d'agrumes que le cédrat (figure 5). Ce serait lors des échanges commerciaux avec l'Asie, à partir du XI<sup>e</sup> siècle, que les Arabes auraient introduit en Méditerranée la bigarade (signalée en Perse en 1030 puis en Sicile en 1094), le citronnier (XII<sup>e</sup>) et les limes (XIII<sup>e</sup>). L'aire d'origine du citron n'est pas tout à fait définie ; elle serait comprise entre le nord de l'Inde et le sud-est de la Chine ou le nord du Myanmar. Il serait cité clairement pour la première fois à partir du XII<sup>e</sup> siècle et n'apparaît pas dans les écrits chinois ou indiens anciens (Bonavia, 1973; Zaragoza, 1991, 2007). Aucun citronnier sauvage n'ayant été rencontré dans la zone supposée d'origine, il semble que le citronnier soit né relativement récemment d'une hybridation d'agrumes plus anciens, dont le cédrat, déjà cité comme parent possible par Malik (1974).

<sup>1</sup> « Assurément, je donnerais facilement à voir ce genre de fruits ; et pas seulement sur la statue de l'Hercule Capitolini, mais aussi sur bien d'autres ; en effet, selon un avis douteux, les artistes anciens auraient préféré les coings, sous prétexte qu'ils étaient plus connus que les cédrats et qu'ils se prêtaient donc mieux à l'imitation. »



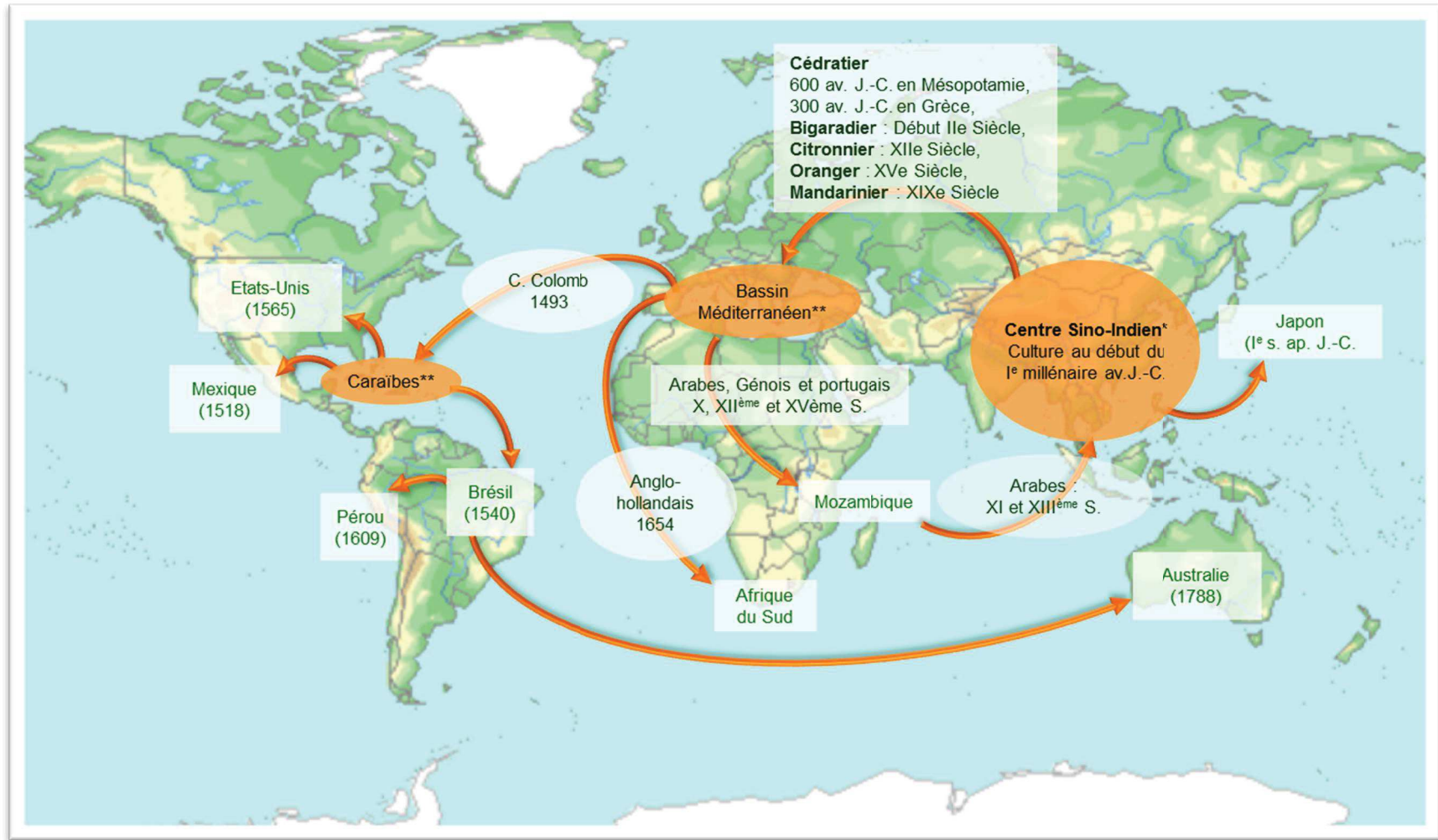


Figure 5 : Diffusion des agrumes dans le monde. \* Origine et aire de répartition primaire. \*\* Aires de répartition et de diversification secondaire (d'après Praloran, 1971).

L'introduction de l'oranger dans le bassin méditerranéen est attribuée, selon les auteurs, soit aux Génois (vers 1400) soit aux Portugais (en 1548). La thèse selon laquelle les Génois furent les premiers à les introduire en Méditerranée est défendue par Praloran (1971), affirmant que des documents italiens datant du XV<sup>e</sup> siècle attestent de la présence d'agrumes dans le bassin méditerranéen. On retrouve la trace de vergers d'orangers dès 1523 en Sicile, Calabre et Ligurie, dès 1515 en Espagne et avant 1600 en Corse (dans la région d'Aregno, en Balagne). Quoi qu'il en soit les Portugais sont certainement à l'origine de la diffusion de l'oranger dans tout le bassin méditerranéen puisque, comme le concède Praloran (1971), de nombreuses variétés d'oranges portent le nom de « Portugaise » et qu'en de nombreuses langues, l'orange porte le nom de « Portugal » : *Bourtougan* en arabe, *Pôrthogal* en persan, *Portakal* en turc, *Pirteqal* en kurde, *портюкал* (portokal) en bulgare ou *aranciu portugallu* (« orange Portugal » pour la différencier de la mandarine : *aranciu mandarinu*, « orange mandarine ») en corse (Association pour le Développement des Etudes Archéologiques, Historiques, linguistiques et Naturalistes du Centre-Est de la Corse, ).

Officiellement, la Méditerranée n'est donc pas le bassin d'origine mais bel et bien le bassin à partir duquel se sont diffusés les agrumes dans le reste du monde (figure 5). D'abord par les Maures qui implantèrent la culture des orangers dans tout le Maghreb, l'ouest de la Méditerranée ainsi que sur une grande partie du continent africain. Les vertus antiscorbutiques des agrumes leur permirent de contribuer au développement de la navigation maritime et d'être introduits dans le Nouveau Monde (Caraïbes), lors du second voyage de Christophe Colomb en 1493. Des Caraïbes, les agrumes furent introduits dans de nombreuses régions du continent américain au début du XVI<sup>e</sup> siècle (USA, Brésil, Argentine, Mexique...). Puis en 1654 les Anglo-Hollandais les introduisent en Afrique du Sud. Enfin l'Australie, pourtant terre d'origine des *Microcitrus* ne connaîtra officiellement ses premiers *Citrus* qu'en 1788 (Praloran, 1971).

Alors que la culture des petits agrumes de type mandarine est une des plus importantes du bassin méditerranéen, le mandarinier (*C. reticulata*) n'y a été introduit qu'à partir du XIX<sup>e</sup> siècle, en Italie. Ce n'est qu'après que la Méditerranée va devenir un centre de diversification secondaire (figure 5) notamment à partir de la naissance de la clémentine en Algérie à la fin du XIX<sup>e</sup> siècle et diffusée d'abord dans les pays du pourtour méditerranéen puis dans le monde entier. La Méditerranée est également considérée comme le centre d'origine des limettes (*C. limetta*).

Enfin, les Caraïbes, du fait de la découverte du pomelo (*C. paradisi*) au XIX<sup>e</sup> siècle, sont également considérées comme un centre de diversification secondaire à partir duquel le pomelo fut introduit en Floride (en 1823) où sa culture s'est intensifiée (Webber, 1943; Scora *et al.*, 1982; Kumamoto *et al.*, 1987; Herrero *et al.*, 1996; de Moraes *et al.*, 2007; Garcia-Lor *et al.*, 2012).

## 2. Les agrumes : botanique et origines génétiques

### 2.1. Ordre des *Sapindales*, famille des *Rutaceae*...

Alors que l'étymologie du nom « agrumes » ne fait pas débat<sup>2</sup>, sa taxinomie elle, soulève, encore aujourd'hui, d'interminables discussions entre spécialistes. L'extraordinaire facilité qu'ont les agrumes à s'hybrider avec de nombreuses espèces et genres proches ainsi qu'à muter de manière spontanée, n'a pas facilité la tâche des premiers taxinomistes qui de plus, pour la plupart, n'ont pas eu la possibilité d'observer des plants dans leur milieu naturel. Nous héritons donc aujourd'hui d'une histoire naturelle des agrumes riche en classifications et points de vue différents et avec un grand nombre de variétés, mutants et hybrides décrits comme des espèces à part entière. Les taxinomistes, après avoir classé les agrumes dans l'ordre des Geraniales (Swingle et Reece, 1967), semblent aujourd'hui d'accord sur le fait que les agrumes appartiennent à l'ordre des *Sapindales* Berchtold & J. Presl, famille des *Rutaceae* Jussieu (Stevens, 2013; NCBI, 2014). La famille des Rutacées, dont le nom vient du genre *Ruta* (Couplan, 2012), rue (plantes odorantes), regroupe des végétaux (plantes herbacées, arbustes et arbres) à glandes à huiles essentielles. C'est la division en tribu, sous tribu, genres et espèces continue à faire polémique (Praloran, 1971). Selon Swingle et Reece (1967) les agrumes font partie des *Aurantioideae* (ou *Aurantieae* Reichenbach ; *Amyridaceae* Kunth, *Aurantiaceae* Jussieu, *Citraceae* Roussel (Stevens, 2013), une sous famille de l'ordre des *Rutaceae* qui se divise en deux tribus : les *Clauseneae* (avec 5 genres) et les *Citreae* (avec 28 genres). La tribu des *Clauseneae* serait plus primitive que celle des *Citreae*. Les *Citreae* se divisent en trois sous tribus : les *Triphasiinae*, les *Balsamocitrinae* et les *Citrinae*. Les *Citrinae* se divisant également en trois groupes dont le plus important comprend les agrumes dits vrais représentés par 6 genres botaniques : *Fortunella*, *Eremocitrus*, *Poncirus*, *Clymenia*, *Microcitrus* et *Citrus* (tableau 1).

---

<sup>2</sup> cf. *Les noms*, point 1.1.

Tableau 1 : Position taxinomique des agrumes selon Swingle &amp; Reece (1967)

Tribu	Sous tribu	Genre	
<i>Clauseneae</i>	<i>Clauseninae</i>	<i>Clausena</i>	
		<i>Glycosmis</i>	
	<i>Merrilliinae</i>	<i>Murraya</i>	
		<i>Merrillia</i>	
	<i>Micromelinae</i>	<i>Micromelum</i>	
<i>Citreae</i>	<i>Balsamocitrinae</i>	<i>Aegle</i>	
		<i>Aeglopsis</i>	
		<i>Afraegle</i>	
		<i>Balsamocitru</i>	
		<i>Feroniella</i>	
		<i>Swinglea</i>	
			<i>Feronia</i>
		Agrumes primitifs	<i>Burkillanthus</i>
			<i>Limnocitrus</i>
		Fruits proches des agrumes	<i>Pleiospermium</i>
			<i>Severinia</i>
	<b>Agrumes vrais</b>	<i>Hesperethusa</i>	
		<i>Atalantia</i>	
		<i>Citropsis</i>	
		<b><i>Citrus</i></b>	
		<b><i>Clymenia</i></b>	
		<b><i>Fortunella</i></b>	
		<b><i>Poncirus</i></b>	
		<b><i>Eremocitrus</i></b>	
		<b><i>Microcitrus</i></b>	
	<i>Triphasiinae</i>	<i>Luvunga</i>	
		<i>Merope</i>	
		<i>Monanthocitrus</i>	
		<i>Oxanthera</i>	
		<i>Pamburus</i>	
		<i>Paramignya</i>	
		<i>Triphasia</i>	
	<i>Wenzelia</i>		

La sous tribu des *Citrinae* se caractérise par la structure de ses fruits en quartiers remplis de sacs à jus ou poils succulents. Cette structure que l'on ne rencontre chez aucune autre plante, même au sein des Rutacées, se nomme hesperidie (Praloran, 1971).

## 2.2. Les agrumes vrais : six genres botaniques

### 2.2.1. *Clymenia* Swingle

Ce genre a des feuilles pérennes, coriaces, avec un pétiole non articulé et court (de 1/10 à 1/20 de la longueur du limbe). Elles sont lancéolées avec un apex acuminé. Les fleurs, souvent simples, apparaissent à la base des feuilles et ont de nombreuses étamines (entre 10 et 20 fois plus que de pétales). Les fruits ovoïdes de la taille d'une petite orange, ont entre 14 et 16 quartiers. Les vésicules à jus sont généralement adhérentes à l'endocarpe. Ce genre ne comprendrait qu'une seule espèce, *Clymenia polyandra* (Tan.) Swing. ou *Citrus polyandra* Tan. Elle fut d'abord classée dans les *Citrus*. Toutefois, Tanaka (1954), puis Swingle et Reece (1967), considérant ce genre comme l'un des plus primitifs des agrumes,

l'ont séparé. Alors que Berhow (2000) évoque la possibilité qu'il s'agisse en fait d'un hybride entre *Fortunella* et *Citrus*, une étude moléculaire récente (Garcia-Lor *et al.*, 2013a) infirme cette hypothèse en mettant en évidence le statut particulier de *Clymenia polyandra*.

### 2.2.2. *Eremocitrus* Swingle

*Eremocitrus* est un genre mono spécifique (*E. glauca* (lindl.) Swing.). Les feuilles sont pérennes, petites, très allongées, gris-vert ou vert glauque, coriaces et velues sur les deux faces. Les fleurs sont le plus souvent isolées mais peuvent également se présenter en petits bouquets, avec des ovaires de 3 à 5 loges. Les fruits sont petits, ovoïdes à piriformes. *Eremocitrus* est originaire des zones désertiques australiennes et est connu pour être relativement résistant au froid et à la sécheresse (Swingle et Reece, 1967).

### 2.2.3. *Microcitrus* Swingle

Les *Microcitrus* présentent un dimorphisme foliaire ; les jeunes plants présentent des feuilles plus petites que les plants plus âgés. Ces feuilles sont pérennes et plus ou moins coriaces. Ils ont de petites fleurs aux étamines libres, ovaire de 4 à 8 loges possédant 4 à 8 ovules par loge. Les vésicules à jus sont plus ou moins allongées, mais souvent globuleuses d'où le surnom de « citron caviar » donné à certaines variétés (figure 6). Les fruits, ronds, ovoïdes ou cylindrique en fonction des espèces, présentent des glandes à huiles essentielles sur le zeste et des sortes de gouttelettes d'huile à l'intérieur du fruit, le long de l'axe central. L'arbre est souvent arbustif et buissonnant. D'après Swingle (1967) ce genre est constitué de six espèces dont cinq (*M. australasica* (F. Muell.) Swing. (figure 6) ; *M. australis* (Planch.) Swing. ; *M. garrowayi* (F. M. Bail.) Swing. ; *M. inodora* (F. M. Bail.) Swing. et *M. maideniana* (Domin) Swing.) sont originaires de l'est de l'Australie (Queensland et nord de la Nouvelle Galles du Sud) et une (*M. warburgiana* (F. M. Bail.) Tan.) du sud-est de la Nouvelle Guinée.



Figure 6 : Vésicules à jus d'un fruit de *Microcitrus australasica* (F. Muell.) Swing., ou citron caviar, coupé en deux ; © F. Curk-Inra.

### 2.2.4. *Poncirus* Raf.

Il s'agit de l'unique genre à feuilles trifoliées (figure 7) et caduques, avec pétiole ailé plus ou moins développé. Les arbres fleurissent au tout début du printemps avant l'apparition des premières feuilles. Les ovaires ont de 6 à 8 loges. Les fruits sont pubescents (figure 7) et le zeste est riche en huiles essentielles (Swingle et Reece, 1967). D'après Swingle et Reece (1967), il s'agirait d'un ancêtre possible des agrumes vrais, adapté aux conditions de froid

extrême du nord de la Chine. Il existe, en fonction des auteurs, une (*Poncirus trifoliata* (L.) Raf. ; (Swingle et Reece, 1967)) ou deux (*P. trifoliata* et *P. polyandra*, (Ding *et al.*, 1984; Wu *et al.*, 1994)) espèces. Le *Poncirus trifoliata* est très utilisé comme porte-greffe adapté aux zones humides à sols acides. Cette espèce a également été et est encore très utilisée comme géniteur dans les programmes de création de porte-greffe comme les Citranges (*Citrus sinensis* (L.) Osb. x *P. trifoliata* (L.) Raf.) et les Citrumelos (*Citrus paradisi* Macf. x *P. trifoliata* (L.) Raf.) par exemple.



Figure 7 : Fruit pubescent et feuille trifoliée de *Poncirus trifoliata* (L.) Raf. ; © F. Curk-Inra.

### 2.2.5. *Fortunella* Swingle

Les kumquats, dont le nom de genre *Fortunella* est dédié à Robert Fortune (Swingle et Reece, 1967) qui les introduisit en Europe en 1846 (à la *London Horticultural Society*), se caractérisent par des fruits plutôt petits (figure 8) dont on peut consommer la peau de certaines espèces. Ils possèdent des ovaires de 3 à 7 loges. Les feuilles sont coriaces et riches en huiles essentielles. Les arbres, de taille variable selon les espèces, sont à feuilles pérennes. Les *Fortunella* sont considérés comme des arbres plus résistants au froid que la plupart des agrumes mais moins que les *Poncirus*. Il s'agit d'un genre originaire du sud-est de la Chine qui, selon Swingle et Reece (1967), n'existerait plus à l'état sauvage. Les *Fortunella* sont surtout cultivés comme arbres d'ornement même s'il existe une production de fruits frais. Selon Krueger et Navarro (2007) il existe quatre espèces de *Fortunella* : *F. margarita* (Lour.) Swing., *F. japonica* (Thunb.) Swing., *F. polyandra* (Ridl.) Tan. et *F. hindsii* (Champ.) Swing. (figure 8).



Figure 8 : Kumquat Hong Kong, (*Fortunella hindsii* (Champ.) Swing.) ; © F. Curk-Inra.

### 2.2.6. *Citrus* L.

Le genre *Citrus* regroupe l'essentiel des espèces d'agrumes cultivés et est le plus diversifié des genres d'agrumes (oranges, citrons, cédrats, mandarines, pamplemousses, bergamotes, bigarades, limes, pomelos... figure 9). La morphologie des arbres varie d'une espèce à l'autre en forme, taille et port. La taille, la forme et la couleur des feuilles sont variables ainsi que la taille des fruits. D'une espèce à l'autre, le diamètre des fruits peut aller d'environ 3 cm, pour la mandarine Nasnaran par exemple, à plus de 25 cm pour certains pamplemousses ou cédrats. La forme des fruits est également très variable : sphériques comme une orange, piriformes comme la poire du commandeur ou digitées comme le cédrat Main de Bouddha, avec ou sans mamelon, avec ou sans col. Les principales espèces qui composent ce genre étant sexuellement compatibles entre elles et avec les genres précédents, il n'y a pas vraiment de consensus quant au nombre d'espèces de *Citrus*.

## 2.3. Le genre *Citrus*

### 2.3.1. Les taxons de base

En fonction des taxinomistes, le genre *Citrus* comprend seize (Swingle et Reece, 1967) ou cent cinquante-sept (Tanaka, 1954, 1961) espèces. La taxinomie de Swingle et Reece (1967) semble plus en adéquation avec les grands axes de structuration de la diversité phénotypique que celle de Tanaka (1954, 1961). Cette dernière, tout en ayant l'avantage de donner une dénomination latine à des variétés cultivées et donc d'être plus précise au moment de comparer des variétés ayant différentes dénominations locales, s'éloigne certainement des principes fondamentaux de la taxinomie qui a pour objectif de regrouper des formes ayant des traits communs. Swingle et Reece (1967) ont divisé le genre *Citrus* en deux sous genres : les *Citrus* avec dix espèces et les *Papedas* avec 6 espèces. Cette subdivision est basée sur les différences morphologiques et les composés chimiques des fruits et des fleurs. En 1954, Tanaka dans son *Species Problem in Citrus*, divise le genre *Citrus* en deux sous genres (les *Archicitrus* et les *Metacitrus*), huit sections, 13 sous sections, 8 groupes, deux microgroupes et 145 espèces. En 1961, il y ajoute deux nouvelles sous sections, un nouveau groupe et 12 nouvelles espèces pour arriver finalement à un total de 157 espèces. On considère que Tanaka a réalisé une description des agrumes plus exhaustive que celle de Swingle et Reece en divisant par exemple les mandarines en 36 espèces alors que Swingle et Reece les regroupent presque toutes sous un seul nom d'espèce : *Citrus reticulata* Blanco. Plus récemment une nouvelle taxinomie des agrumes a été proposée par Mabberley (1997) qui, par rapport aux deux autres classifications précédentes, respecte plus les notions d'inter fertilité et d'espèce en regroupant les six genres des agrumes vrais, *Poncirus*, *Fortunella*, *Citrus*, *Eremocitrus*, *Microcitrus* et *Clymenia*, sous une seule dénomination de genre : *Citrus*. Ce genre est ensuite décliné en 3 espèces et 4 groupes hybrides. Bien avant déjà, des études basées sur des caractères biochimiques (Scora, 1975) et morphologiques (Barrett et Rhodes, 1976) suggéraient que la majorité des espèces du genre *Citrus* étaient probablement issus d'hybridations directes ou successives de trois espèces ancestrales que sont les cédrats (*Citrus medica* L.), les mandarines (*Citrus reticulata* Blanco) et les pamplemousses (*Citrus maxima* (burm.) Merr.). Cette hypothèse est aujourd'hui confirmée par de nombreuses études sur l'organisation de la diversité phénotypique et génétique (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003; Fanciullino *et al.*, 2006). En plus de ces trois taxons de base, considérés comme ancestraux, *Citrus*

*micrantha* Wester (*Papeda*) est aujourd'hui considéré par de nombreux auteurs (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b) comme le quatrième taxon de base à l'origine en particulier des citrons vert tels que la lime Mexicaine (*Citrus aurantifolia* (Christm.) Swing.). Enfin, une liste de trente trois noms d'espèces de *Citrus* acceptés est aujourd'hui disponible sur la page internet *The Plant List* (Royal Botanic Gardens, Kew and Missouri Botanical Garden, 2013). Il s'agit d'une liste dite de travail des espèces connues de plantes, générée à partir d'une collaboration entre le *Royal Botanic Gardens, Kew Garden* et *Missouri Botanical Garden*. Les noms d'espèces présents dans cette liste sont en constante évolution.



Figure 9 : Exemple de diversité phénotypique du genre *Citrus*. Photomontage à partir de photos de variétés d'agrumes de la collection Inra-Cirad de San Giuliano en Corse. © F. Curk-Inra



### 2.3.1.1. Les cédratiers

Probablement originaires d'une zone comprise entre le sud de la Chine et le nord de l'Inde, les cédratiers sont généralement des arbustes épineux à port irrégulier souvent étalé. Les feuilles sont plutôt grandes, coriaces, obovales, dentelées, et antocyanées (pourpres à roses) chez les variétés à pulpe acide quand elles sont jeunes (figure 10a, b et c) et complètement vertes une fois entièrement développées. Le pétiole est court, non articulé et sans ailette. Les fleurs sont généralement de grande taille et antocyanée (roses à roses pâles) chez les variétés à pulpe acide (figure 10a, b et c) et blanches chez les variétés à pulpe douce (figure 10d).

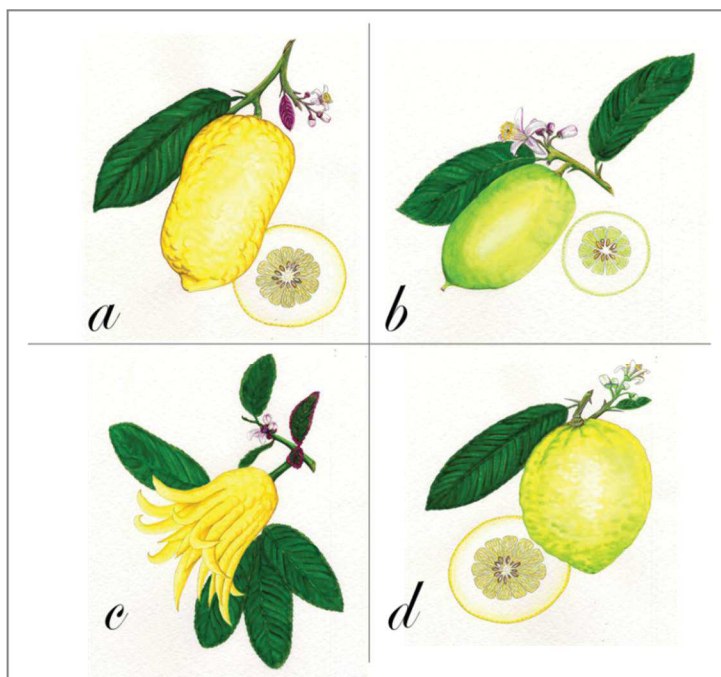


Figure 10 : Exemples de *Citrus medica* : a. Cédrot Diamante ; b. Cédrot Etrog ; c. Cédrot Main de Bouddha ; d. Cédrot de Corse ; Aquarelles, gauche ; © F. Curk

Il n'est pas rare de rencontrer des fleurs stériles femelles, sans pistil. De plus, en général, les organes reproducteurs mâles et femelles des fleurs non stériles arrivent à maturité en même temps juste avant l'ouverture de la fleur. Les fruits sont généralement de grande taille (pouvant dépasser les 2 kg chez certaines variétés) ils peuvent être sphéroïdes à ellipsoïdes, souvent irréguliers. Les fruits, de couleur verte à jaune orangé, ont un albédo charnu très épais, souvent très ferme, adhérant fortement aux quartiers. Les pépins sont monoembryonnés<sup>3</sup> (Swingle et Reece, 1967). Les cédratiers sont sensibles au froid et au vent et sont aujourd'hui cultivés pour la confiserie, les liqueurs et l'ornement dans le bassin méditerranéen, et en Amérique du Sud ; pour l'ornement et la médecine traditionnelle en Asie. Les cédratiers sont généralement divisés en deux catégories, ceux à pulpe acide (cédrot Diamante [figure 10a], cédrot Etrog ou Ethrog [figure 10b]...) et à pulpe douce (cédrot de Corse ou Corsican [figure 10d]) (Swingle et Reece, 1967; Praloran, 1971). Une étude récente (Luro *et al.*, 2012) montre que, malgré le faible taux d'hétérozygotie déjà observé chez les cédrats (Barrett et Rhodes, 1976; Barkley *et al.*, 2006), certainement du à sa facilité d'autofécondation (les fleurs s'autopollinisant avant ouverture), il existe une hétérozygotie suffisante pour structurer la diversité des variétés de cédrats cultivés dans le bassin méditerranéen. Citons enfin le cédrot Main de Bouddha, ou *sarcodactylis* (figure 10c), qui a une fructification particulière puisque le fruit se divise en segments ressemblant à des doigts et est généralement dépourvu de pulpe. Cette variété est essentiellement cultivée comme plante d'ornement ou pour la pharmacopée traditionnelle en Chine (figure 11).

<sup>3</sup> cf. *Les facteurs limitants l'amélioration génétique chez les agrumes*, point 6.3.



Figure 11 : Sachet de jeunes fruits séchés de cédrats Main de Bouddha destinés à la pharmacopée traditionnelle Chinoise, achetés sur un marché Chinois (Collection privée, Michel Baches) ; © F. Curk-Inra

### 2.3.1.2. Les mandariniers

Comme nous l'avons déjà évoqué la classification des mandarines est très complexe et controversée. Un grand nombre de groupes dit naturels et d'hybrides avérés (figure 12) ont été élevés au rang d'espèces par Tanaka (1954). Originaires d'Asie (du nord du Vietnam au Japon, en passant par la Chine) les mandariniers sont des arbres de taille variable, présentant généralement de petits rameaux avec épines. Les feuilles sont lancéolées, les fleurs peuvent être isolées ou groupées en inflorescences. Les fruits sont généralement sphéroïdes, pouvant être aplatis aux deux pôles, à peau fine, facile à détacher des quartiers. Il existe des mandarines polyembryonnées et d'autres monoembryonnées<sup>4</sup>. Il semblerait, d'après les observations de Garcia-Lor (2013b), que les mandarines monoembryonnées soient essentiellement des mandarines hybrides, ayant des introgressions de pamplemousse qui leur conféreraient ce caractère.



Figure 12 : Exemple de diversité phénotypique du groupe des mandarines et hybrides de mandarines de la collection de l'lvia à Moncada, en Espagne ; © F. Curk-Inra

<sup>4</sup> cf. *Les facteurs limitants l'amélioration génétique chez les agrumes*, point 6.3.

### 2.3.1.3. Les pamplemoussiers

Il s'agit d'un groupe se différenciant nettement des deux précédents et ne comprenant que des variétés monoembryonnées. Les arbres sont de taille variable pouvant dépasser les 10 m en zone tropicale. Les jeunes pousses ainsi que les feuilles sont souvent pubescentes. Les feuilles sont très grandes, ovées à obovales avec un pétiole ailé. Les fleurs sont grandes et allongées et les fruits sont de tailles, de couleurs et de formes variées mais généralement gros, sphéroïdes, aplatis ou piriformes (figure 13) pouvant dépasser les 20 cm de diamètre, de couleur externe comme interne verte, jaune ou rose. Ils ont généralement un albedo très épais et ont des vésicules à jus très allongées qui se détachent facilement (Swingle et Reece, 1967). Ces fruits, originaires d'Asie du Sud-Est y sont toujours très cultivés pour la consommation en frais (figure 14) et sont utilisés, encore aujourd'hui, comme offrandes religieuses, au Vietnam et en Chine notamment.

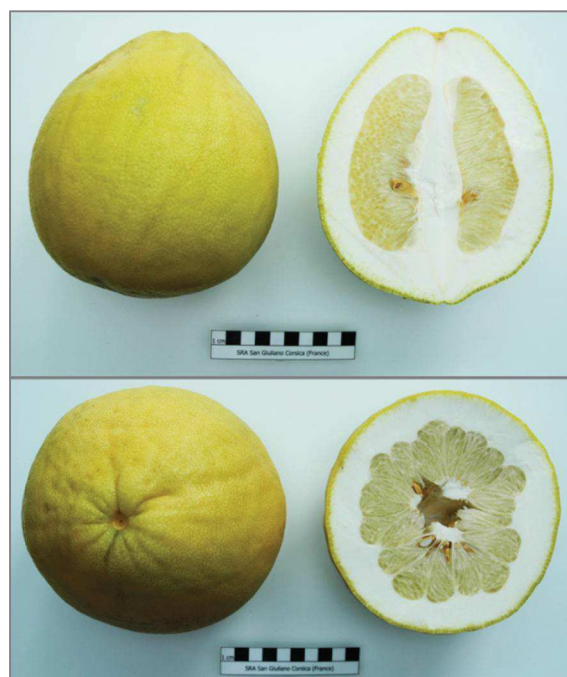


Figure 13 : Pamplemousse chinois ;  
© F. Curk-Inra



Figure 14 : Pamplemousses en exposition sur un étalage à Wuhan en Chine ; © F. Curk-Inra.

### 2.3.1.4. Les papedas

Ce groupe comprend un grand nombre d'espèces sauvages. Les feuilles sont de tailles variables mais généralement bilobées, avec un pétiole ailé presque aussi grand que le limbe. Les fleurs sont petites, avec des étamines non soudées. Les fruits, de tailles variables, ont des gouttelettes d'huile âcres dans la pulpe qui la rend particulièrement amère et désagréable à consommer. Certaines variétés, comme le combava (*C. hystrix* D.C., figure 15), sont pourtant particulièrement appréciées pour leur huile essentielle à fort parfum de citronnelle qui entre dans la composition de nombreux plats et cosmétiques traditionnels de l'océan indien (île de la Réunion, Mayotte, Madagascar...). Le *C. micrantha*, utilisé dans cette thèse comme taxon de base, représentant sauvage des papedas, serait originaire du sud des Philippines, où il est parfois cultivé pour des usages en cosmétique traditionnelle (Swingle et Reece, 1967).



Figure 15 : Combava (*C. hystrix* D.C.) ; © F. Curk-Inra

### 2.3.2. Origine génétique des espèces cultivées

Toute la communauté scientifique travaillant sur les agrumes adhère à l'hypothèse déjà évoquée dans les chapitres précédents selon laquelle il y aurait 4 taxons de base (mandariniers, pamplemoussiers, cédratiers et papedas). Par différentes combinaisons sexuées, ces taxons de base auraient généré les hybrides initiaux fondateurs des espèces dites secondaires (*C. sinensis*, orangers ; *C. paradisi*, pomelos ; *C. limon*, citronniers ; *C. aurantium*, bigaradiers ; *C. clementina*, clémentiniers...). Les limettiers auraient des origines multiples, l'une d'entre elles implique une hybridation entre un cédratier (*C. medica*) et un représentant du sous-genre *Papeda* (*Citrus micrantha*) (Nicolosi *et al.*, 2000) (figure 16).

L'apomixie facultative de toutes les espèces secondaires a permis de fixer et multiplier de façon conforme ces structures fortement hétérozygotes et a conduit les taxinomistes à considérer ces groupes comme des espèces (Scora, 1975; Barrett et Rhodes, 1976). Comme cela a pu être démontré pour les caroténoïdes ainsi que pour l'acidité, on sait qu'une grande partie de la diversité phénotypique et de sa structuration au sein du genre *Citrus* est liée à la différenciation entre les 4 taxons ancestraux avant la genèse des espèces secondaires (Ollitrault *et al.*, 2003; Fanciullino *et al.*, 2006). Le genre *Citrus* a connu une période de différenciation allopatrique entre les quatre taxons ancestraux, suivie d'événements de réticulations dont les produits directs (ou ceux issus d'un nombre réduit de méioses interspécifiques) ont souvent été fixés par l'apomixie facultative (ce qui a conduit les taxinomistes à les considérer comme de nouvelles espèces : l'oranger, *C. sinensis* ou le citronnier, *C. limon*, par exemple). Il en résulte des structures génotypiques fortement hétérozygotes formées d'une mosaïque de grands fragments d'origines phylogénétiques différentes comme c'est le cas pour *C. sinensis* qui est une mosaïque de fragments d'ADN hérités de *C. reticulata* (les mandariniers) et *C. maxima* (les pamplemoussiers). Les études basées sur le génome chloroplastique (Green *et al.*, 1986; Nicolosi *et al.*, 2000) et mitochondrial (Froelicher *et al.*, 2011) démontrent que le cytoplasme des orangers, bigaradiers et pomelos vient des pamplemoussiers (*C. maxima*). Les théories sur l'origine

des espèces secondaires d'agrumes sont renforcées par de nombreux travaux réalisés à partir de différents types de marqueurs moléculaires tels que les isozymes (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003), RLFP (Yamamoto *et al.*, 1993; Federici *et al.*, 1998), RAPD, SCAR (Nicolosi *et al.*, 2000), AFLP (Liang *et al.*, 2007), SSR (Luro *et al.*, 2001; Barkley *et al.*, 2006; Ollitrault *et al.*, 2010), ou SNP (Ollitrault *et al.*, 2012b).

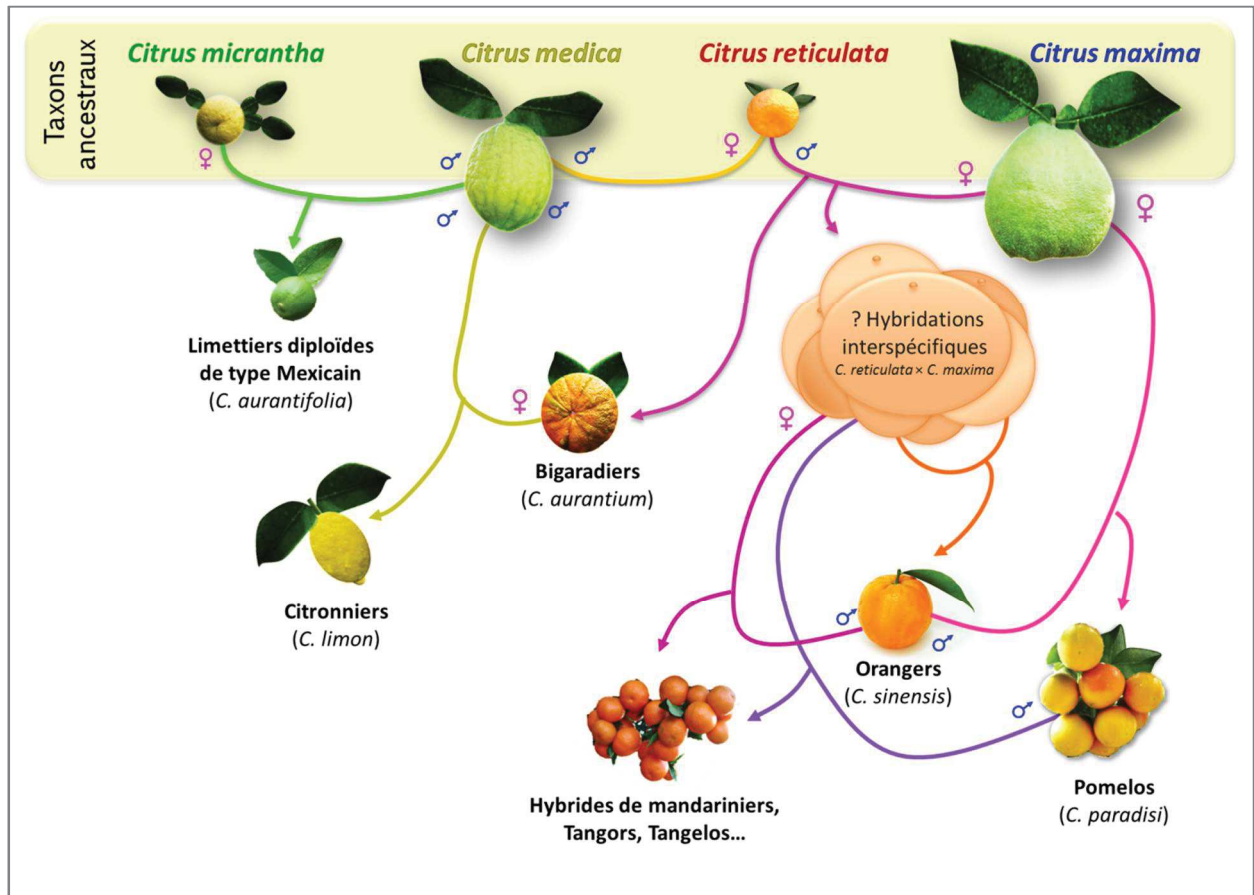


Figure 16 : Origines phylogénétique supposées des cultivars modernes d'agrumes (photos et photo-montage : © F. Curk-Inra)

### 2.3.2.1. L'oranger (*Citrus sinensis* (L.) Osb.).

L'oranger est génétiquement et phénotypiquement proche des mandariniers (*C. reticulata*) mais présente des introgressions de pamplemoussier (*C. maxima*) (Nicolosi *et al.*, 2000). La plus grande proximité de l'oranger avec les mandariniers suggère qu'il ne s'agirait pas d'un hybride direct mais d'un *backcross* de première ou de deuxième génération avec le génome du mandarinier (Barrett et Rhodes, 1976; Nicolosi *et al.*, 2000). Roose *et al.* (Roose *et al.*, 2009) décrivent l'oranger comme un *backcross* 1 (BC1) [(*C. maxima* x *C. reticulata*) x *C. reticulata*]. Cette hypothèse est reprise par Xu *et al.* (2013) en s'appuyant sur le séquençage complet du génome de l'oranger. Toutefois, la présence de portions de génome en homozygotie *C. maxima* et d'autres en homozygotie *C. reticulata*, mises en évidence par du séquençage ciblé (Garcia-Lor *et al.*, 2013b) et plus récemment par reséquençage complet du génome (Wu *et al.*, 2014) démontre que les deux parents de l'oranger étaient d'origine interspécifique (*C. maxima* / *C. reticulata*).

### 2.3.2.2. Le bigaradier ou oranger amer (*C. aurantium* L)

Il semble qu'il s'agisse d'un hybride direct entre un mandarinier (*C. reticulata*) et un pamplemoussier (*C. maxima*) (Swingle et Reece, 1967; Scora, 1975; Barrett et Rhodes, 1976; Green *et al.*, 1986; Scora, 1988; Yamamoto *et al.*, 1993; Nicolosi *et al.*, 2000; Uzun *et al.*, 2009).

### 2.3.2.3. Le pomelo (*C. paradisi* Macf.)

Le pomelo (figure 1 et 17) est une espèce proche des pamplemoussiers et serait issu d'une hybridation spontanée entre un pamplemoussier (*C. maxima*) et un oranger (*C. sinensis*) (Barrett et Rhodes, 1976; Scora *et al.*, 1982; de Moraes *et al.*, 2007; Li *et al.*, 2010; Ollitrault *et al.*, 2012b) découverte à la fin du XIX<sup>e</sup> siècle dans les Caraïbes (Webber, 1943).

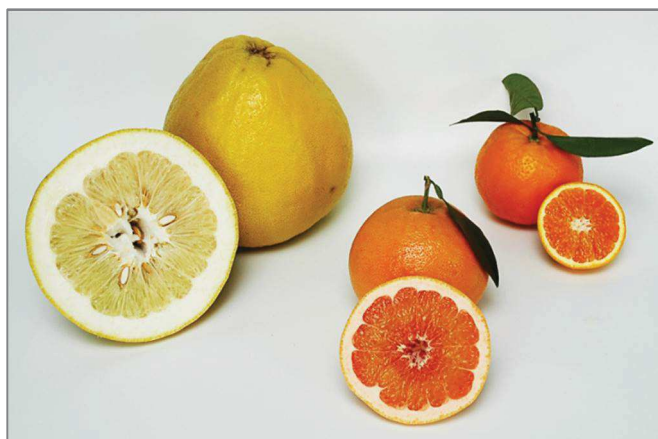


Figure 17 : Pomelo Star Ruby (*C. paradisi*), entre un pamplemousse (*C. maxima*), à gauche, et une orange (*C. sinensis*), à droite ; © F. Curk-Inra

### 2.3.2.4. Le clémentinier (*C. clementina* Hort. ex Tan.)

C'est paradoxalement grâce à l'une des introductions les plus récentes d'agrumes dans le bassin méditerranéen que la culture des petits agrumes s'y est intensifiée. Ce n'est qu'en 1902, peu après l'introduction de la mandarine en Italie (au XIX<sup>e</sup> siècle), que le Professeur Trabut (1902) décrit pour la première fois la clémentine (*C. clementina*). Elle est considérée aujourd'hui comme un hybride naturel entre un mandarinier commun (*C. deliciosa*) et du pollen d'oranger (*C. sinensis*) (Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b).

### 2.3.3. Le cas particulier des limettiers et des citronniers

Il est admis que les citronniers et les limettiers ont un lien de parenté avec le cédratier (*C. medica*) (Scora, 1975; Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Bayer *et al.*, 2009; Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013b). Mais en ce qui concerne les origines des différents génotypes de ces deux groupes, les avis divergent. Suivant les taxinomistes (Tableau 2), les citronniers et les limettiers sont classés en deux espèces, respectivement *Citrus limon* (L.) Burm. et *Citrus aurantifolia* (Christm.) Swing. (Swingle et Reece, 1967) ou séparés en plus de 30 espèces (Tanaka, 1977).

Tableau 2 : Correspondances entre les classifications de Swingle et Reece (1967) et de Tanaka (1977) pour les limettiers et les citronniers.

Classification selon Swingle et Reece (1967)			Classification selon Tanaka (1977)	
Groupe	Genre	Espèce	Genre	Espèce
Limes	<i>Citrus</i>	<i>aurantifolia</i> (Christm.) Swing.	<i>Citrus</i>	<i>aurantifolia</i> (Christm.) Swing.
			<i>Citrus</i>	<i>bergamia</i> Risso & Poit.
			<i>Citrus</i>	<i>davaoensis</i> Tan.
			<i>Citrus</i>	<i>excelsa</i> Wester
			<i>Citrus</i>	<i>hyalopulpa</i> Tan.
			<i>Citrus</i>	<i>javanica</i> Blume
			<i>Citrus</i>	<i>latifolia</i> Tan.
			<i>Citrus</i>	<i>limettioïdes</i> tan.
			<i>Citrus</i>	<i>longispina</i> Wester
			<i>Citrus</i>	<i>macrophylla</i> Wester
			<i>Citrus</i>	<i>montana</i> Tan.
			<i>Citrus</i>	<i>obversa</i> Hassk.
			<i>Citrus</i>	<i>ovata</i> Hassk.
			<i>Citrus</i>	<i>papaya</i> Hassk.
			<i>Citrus</i>	<i>pennevesiculata</i> (Lush.) Tan.
Citrons	<i>Citrus</i>	<i>limon</i> (L.) Burm.	<i>Citrus</i>	<i>pseudolimonum</i> Wester
			<i>Citrus</i>	<i>webberri</i> Wester
			<i>Citrus</i>	<i>assamensis</i> Dutta & Bhatt
			<i>Citrus</i>	<i>aurata</i> Risso
			<i>Citrus</i>	<i>balotina</i> Poit. & Turp.
			<i>Citrus</i>	<i>duttae</i> Tan.
			<i>Citrus</i>	<i>jambhiri</i> lush.
			<i>Citrus</i>	<i>karna</i> Raf.
			<i>Citrus</i>	<i>limetta</i> Risso
			<i>Citrus</i>	<i>limon</i> (L.) Burm.
			<i>Citrus</i>	<i>limonia</i> Osb.
			<i>Citrus</i>	<i>longilimon</i> Tan.
			<i>Citrus</i>	<i>lumia</i> Risso & Poit.
			<i>Citrus</i>	<i>macrolimon</i> Tan.
			<i>Citrus</i>	<i>megaloxycarpa</i> lush.
			<i>Citrus</i>	<i>mellarosa</i> Risso
			<i>Citrus</i>	<i>meyeri</i> Y. Tan.
			<i>Citrus</i>	<i>peretta</i> Risso
			<i>Citrus</i>	<i>pseudolimon</i> Tan.
			<i>Citrus</i>	<i>pyriformis</i> Hassk.
<i>Citrus</i>	<i>rissoi</i> Risso			
<i>Citrus</i>	<i>sarbati</i> Tan.			

### 2.3.3.1. Les citronniers (*C. limon*)

L'origine géographique des citrons jaunes bien connus en Méditerranée est encore assez mystérieuse. En effet, on retrouve peu de citrons de type « Eureka » (figure 18) dans les zones d'origine supposées que sont le sud de la Chine et l'Inde (Tanaka, 1929; Biraghi, 1935). Webber (1967) évoque pourtant comme centre d'origine possible le sud de la Chine et le nord du Myanmar. Les incertitudes viennent du fait que les sources historiques sont rares et que très souvent, la description du fruit ne permet pas de faire la différence entre un cédrat, un citron ou tout autre hybride (Malik *et al.*, 1974).



Figure 18 : Citron type Eureka [*Citrus limon*]; © F. Curk -Inra

D'anciennes études (Scora, 1975; Barrett et Rhodes, 1976; Federici *et al.*, 1998) affirment que le citronnier serait un hybride direct entre un cédratier et un limettier alors que de plus récentes (Nicolosi *et al.*, 2000; Gulsen et Roose, 2001c; Li *et al.*, 2010) décrivent plutôt le citronnier comme un hybride direct entre bigaradier et cédratier. Le bigaradier serait l'origine maternelle et le cédratier l'origine paternelle (Gulsen et Roose, 2001c; Froelicher *et al.*, 2011).

### 2.3.3.2. Les limettiers (*C. aruatifolia*)

Le limettier de type lime mexicaine (*Citrus aruatifolia*, figure 19) serait originaire de l'Insulinde (Webber *et al.*, 1967). Il est considéré par certains auteurs comme un hybride direct entre *Citrus micrantha* et *Citrus medica* (Scora, 1975; Nicolosi *et al.*, 2000). Hypothèse proposée dès 1969 par Tanaka qui évoquait déjà la possibilité que les limettiers dériveraient du groupe des papedas. Cette variété produit des petits citrons verts à peau fine et à pulpe vert pâle avec un à deux pépins par fruit



Figure 19 : Lime Mexicaine (*C. aruatifolia*)  
© F. Curk-Inra

Il existe également des limettiers triploïdes dont nous pouvons distinguer deux types :

- Ceux sans pépins, de type Tahiti (figure 20), Bears ou de Perse, qui semblent avoir une origine hybride avec un lien de parenté avec le cédratier au même titre que les autres limettiers et citronniers (Mabberley, 2004; Bayer *et al.*, 2009). D'après Webber (1967) l'origine de ces limes est mal connue. Tout ce que l'on sait c'est que le nom de Tahiti viendrait du fait que ce fruit fut introduit en Californie depuis Tahiti entre

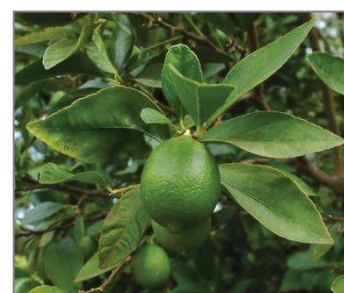


Figure 20 : Lime de Tahiti (*C. latifolia*);  
© F. Curk-Inra.



1850 et 1880 et que la lime de Perse tirerait son nom du fait qu'elle serait arrivée en Méditerranée par la Perse, même si Chapot (1965) n'en retrouva aucune en Iran en 1965. Enfin Reece et Childs (1962) pensent avoir identifié au moins un des deux parents, la lime mexicaine et proposent pour l'autre deux candidats possibles, soit le citron jaune, soit le cédrat. Ces variétés se caractérisent par la production de fruits plus gros que ceux de la lime mexicaine, à peau fine et pulpe verte mais sans pépins.

- Et ceux avec pépins de type Tanepao (ou Copenrhad). Le caractère triploïde de ces limettiers est longtemps passé inaperçu du fait de l'association de ce caractère à l'absence même de pépins (Ollitrault *et al.*, 2008). Leur origine est très mal connue.

Enfin, nous pouvons également citer une lime tétraploïde naturelle : la lime Giant Key qui serait un autotétraploïde spontané issu d'un semi de pépins de lime Key diploïde (de type Mexicaine, *C. aurantifolia*) sélectionné en 1973 par HC Barrett de l'US Horticultural Research Laboratory à Orlando aux Etats-Unis.

### 2.3.3.3. Les autres limettiers et citronniers

Il existe une grande diversité de limettiers et de citronniers ; l'origine des différents groupes reste très controversée. Ainsi, par exemple, le limettier Rangpur (*Citrus limonia* Osb. figure 21) et le citronnier *Rough lemon* (*Citrus jambhiri* Lush., figure 22), auraient une origine commune, différente de celle du *Citrus limon* (Carvalho *et al.*, 2005), les deux étant considérés comme originaires d'Inde et du Pakistan ((Webber *et al.*, 1967). Une étude récente confirmerait l'hypothèse de Scora (1975) selon laquelle il s'agirait de croisements entre un cédratier et un mandarinier (Li *et al.*, 2010).



Figure 21 : Lime Rangpur, (*C. limonia*) ; © F. Curk-Inra.



Figure 22 : *Rough lemon*, (*C. jambhiri*) ; © F. Curk-Inra.

Il existe également différents groupes de limes douces, dont les *C. limetta* (limonette de Marrakech, limette à melon (figure 23) ou bergamote de Tunis), cultivées dans le bassin méditerranéen et qui d'après Webber (1967) seraient des mutants doux de limes acides, les formes acides de *C. limetta* sont courantes mais moins connues, il en existe quelques exemplaires dans le CRB Citrus Inra-Cirad de San Giuliano. Enfin, les *C. limettioïdes* (limes douces de Perse, du Brésil ou de Palestine) sont probablement originaires d'Inde (Webber *et al.*, 1967).

L'origine du bergamotier (*Citrus bergamia* Risso & Poit., figure 24) est controversée. Galesio (1811) affirmait que le bergamotier était un hybride entre un oranger et un citronnier. Chapot (1962) tout en expliquant que le bergamotier serait apparu à partir de semis dans la région de Naples ou de Calabre, en Italie, entre le XIV<sup>e</sup> et le XV<sup>e</sup> siècle, insiste sur la nature hybride de cette variété en précisant que les pépins monoembryonnés de la bergamote sont une marque d'hybridation. Il évoque, dans le même article (Chapot, 1962), le bigaradier et un limettier comme parents possibles. Il écarte le citronnier car, selon lui, la coloration pourpre des jeunes rameaux et des fleurs, bien plus prononcée chez le citronnier que chez les limettiers, ne se retrouve pas chez le bergamotier qui a des fleurs blanches et des jeunes rameaux verts. Webber (1967) évoque, lui, une possible mutation de bigarade. Les études récentes ne sont pas toutes d'accord. Certaines présentent le bergamotier comme un hybride possible entre cédratier et limettier (Chen *et al.*, 1991) d'autres entre un bigaradier et un limettier doux (Herrero *et al.*, 1996; Federici *et al.*, 2000) ou enfin entre un bigaradier et un cédratier (Nicolosi *et al.*, 2000; Li *et al.*, 2010). Il semble aujourd'hui admis que le bigaradier et le bergamotier ont la même origine cytoplasmique et donc la même origine maternelle (Gulsen et Roose, 2001b) mais l'origine paternelle n'est toujours pas déterminée.

Le macrophylla ou Alemow (*Citrus macrophylla* Wester, figure 25) est lui rangé parmi les citronniers et les limettiers avec des liens admis avec le groupe des papedas (Barkley *et al.*, 2006). En 1967, Webber considère l'Alemow, probablement originaire de l'île de Cebu au Philippines, comme un possible hybride entre un représentant du groupe des papedas (*Citrus celebica*) et un pamplemousse (*C. grandis*). En 1976, Barrett et Rhodes pensent qu'Alemow pourrait être un hybride entre un cédratier et un papeda. Dans cette même étude de 1976, le combava (*Citrus hystrix* D.C. figure 15) est même été évoqué comme pouvant être le papeda à l'origine d'Alemow. Des études plus récentes indiquent que, tout comme la lime mexicaine, *C. macrophylla* est certainement issu d'une hybridation directe ou indirecte entre un papeda et un cédratier mais privilégiant *C. micrantha* par rapport à *C. hystrix* (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b).



Figure 23 : Limette à mamelon (*Citrus limetta*) ; © F. Curk-Inra.



Figure 24 : Bergamote castagnaro (*Citrus bergamia*) ; © F. Curk-Inra.



Figure 25 : Alemow (*Citrus macrophylla*) ; © F. Curk-Inra.

### 3. Les agrumes : ressources génétiques

#### 3.1. Les centres de ressources phytogénétiques

Le 29 juin 2004, après que quarante gouvernements l'ont ratifié, est entré en vigueur le traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture (Organisation des Nations Unies pour l'Alimentation et l'Agriculture, 2009). Ce traité définit les ressources phytogénétiques comme « le matériel génétique d'origine végétale ayant une valeur effective ou potentielle pour l'alimentation et l'agriculture ». Les objectifs de ce traité sont « la conservation et l'utilisation durable des ressources phytogénétiques pour l'alimentation et l'agriculture et le partage juste et équitable des avantages découlant de leur utilisation en harmonie avec la Convention sur la diversité biologique, pour une agriculture durable et pour la sécurité alimentaire ». Cela sous-entend des actions pour améliorer les politiques et les cadres juridiques de la gestion de cette diversité agricole, d'intégrer les ressources génétiques et la biodiversité dans les programmes de développement et de renforcer les conservatoires et garantir une utilisation durable en facilitant l'accès aux ressources génétiques et aux connaissances s'y rattachant (FAO, 2014). De plus, La sélection humaine de variétés d'intérêt économique, malgré les programmes de création variétale, et la multiplication végétative de ces variétés d'intérêt a paradoxalement fait disparaître beaucoup de génotypes sauvages originaux alors que la diversité génétique sauvage, potentiellement présente dans les sites d'origine, est constamment menacée par la destruction de leur habitat naturel (déforestation, pression démographique, tourisme...) comme c'est le cas en Inde ou en Chine. Dans ce contexte, les collections *ex situ* jouent un rôle important dans la conservation de la diversité tout en en y permettant l'accès pour la création de nouvelles variétés dans un contexte socio-économique en constante évolution.

Tout au long de son histoire l'Institut national de la recherche agronomique (Inra) a constitué de nombreuses collections de ressources génétiques végétales pour ses besoins de recherche et, de fait, a participé ainsi à la préservation de la biodiversité. L'Inra s'est associé dans les années 2000 à la mise en place des centres de ressources biologiques (CRB) en liaison avec les autres instituts de recherche et d'enseignement supérieur et les ministères concernés. Les CRB sont des centres spécialisés qui rassemblent, valident, étudient, sécurisent et distribuent des collections d'organismes vivants (semences, greffons...) et des parties « répliquables » de ces organismes (banques d'ADN, plasmides...) dans des conditions rigoureuses de qualité et de traçabilité ; ils maintiennent aussi les bases de données associées à ces collections. Les CRB sont des éléments essentiels du dispositif international de recherche, notamment dans le domaine des biotechnologies.

Les agrumes, grâce à la collection Inra-Cirad de San Giuliano en Corse, ont été intégrés dans cette démarche de sécurisation des processus de conservation, de caractérisation et de diffusion (du matériel végétal et de l'information qui l'accompagne) permettant, comme le préconise la FAO, de garantir une production d'aliments variés, durables et nutritionnellement diversifiés. Le processus de certification du CRB Citrus, débutée en janvier 2013, a abouti à la certification NF S96-900. La collection d'agrumes Inra-Cirad de San Giuliano, dont la création remonte à la fin des années 50, est aujourd'hui riche de plus de 1100 accessions plantées en plein champ et réparties sur 14 ha. Cette collection, dans laquelle le genre *Citrus* est représenté par plus de 900 accessions, est une des plus riches en mandariniers et hybrides de mandariniers (figure 26).

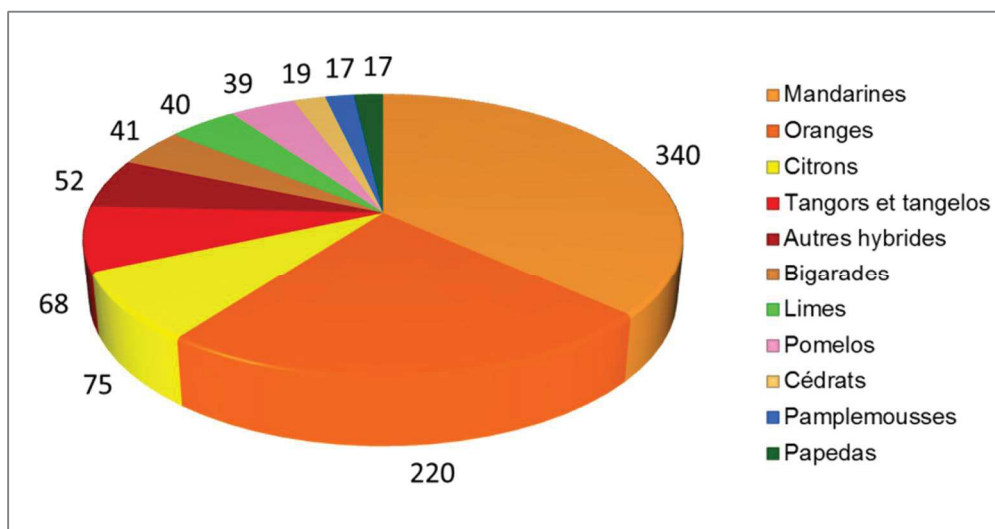


Figure 26 : Nombre d'accèsions du genre *Citrus*, par groupe variétal, de la collection d'agrumes Inra-Cirad de San Giuliano en Corse en 2014.

D'autres genres sont présents en collection et, parmi plus de 200 accèsions n'appartenant pas au genre *Citrus*, on trouve des représentants des genres *Poncirus*, *Fortunella*, *Microcitrus*, *Eremocitrus*, *Severinia*, *Clausena* et *Murraya* ainsi que de nombreux hybrides entre espèces et entre genres différents.

La collection Inra-Cirad fait partie des plus importantes collections d'agrumes qui existent aujourd'hui dans le monde, les autres se trouvent en Espagne, en Chine, au Japon et aux États-Unis.

La principale collection espagnole est à Moncada (Valence) et est gérée par l'*Instituto Valenciano de Investigaciones Agrarias* (Ivia). Elle regroupe plus de 600 génotypes dont les principales variétés commerciales de mandariniers et hybrides de mandariniers, un grand nombre de représentants du genre *Citrus* ainsi que des représentants des genres proches de la sous famille des *Aurentioideae*. Cette collection est d'une part, cultivée en plein champ et, d'autre part, conservée sous forme d'arbres en pots entreposés sous serres *insect-proof* afin d'éviter toute contamination des pieds mère par des insectes vecteurs (figure 27).



Figure 27 : Plants d'agrumes de la collection de l'ivia (Moncada, Valence, Espagne) sous serre *insect-proof* ; © F. Curk-Inra

Les collections française et espagnole sont la source du matériel végétal utilisé dans les programmes de certification fruitière pour la production de plants destinés aux agriculteurs. Les différentes variétés certifiées sont régulièrement vérifiées sur le plan sanitaire et pomologique.

La collection officielle chinoise regrouperait, *ex situ*, plus de 1000 accessions (Liu et Deng, 2007). La Chine est un des rares pays qui a des ressources génétiques sauvages et originales *in situ* mais nous n'avons malheureusement que peu d'informations à ce sujet. Le Japon gère au sein de 6 collections réparties sur tout son territoire près de 1200 génotypes dont un grand nombre de mandariniers de type satsuma (Krueger et Navarro, 2007) et une grande diversité des genres apparentés prospectés dans le sud-est asiatique.

Enfin, aux États-Unis, la *Citrus Variety collection* de l'Université de Riverside en Californie, regroupe près de 1000 génotypes : <http://www.citrusvariety.ucr.edu/> (Barkley *et al.*, 2006).

### 3.2. La conservation des ressources génétiques agrumes

Les agrumes ont des graines dites récalcitrantes (Chin et Roberts, 1980) c'est-à-dire qu'elles ne tolèrent pas la dessiccation extrême ni la conservation longue durée au froid. Le pouvoir germinatif des graines de la plupart des variétés d'agrumes chute considérablement à partir de quelques mois de conservation à une température comprise entre 4 et 5°C., à une humidité relative de 85 % (Aubert et Vullin, 1997; Jacquemond *et al.*, 2013b).

La cryoconservation (-196°C.) pourrait être une solution mais des études supplémentaires sont encore nécessaires (Engelmann *et al.*, 2009). Cette caractéristique empêche toute conservation à long terme de semences, moyen pourtant le plus utilisé pour une grande partie de la biodiversité végétale (Rao *et al.*, 2006).

Malgré les possibles avancées dans les techniques de cryoconservation, la conservation de la biodiversité « agrumes » ne se fait pour l'instant que sous la forme de plantes. La sélection humaine et la mise en monoculture de variétés choisies ont conduit à une perte de diversité importante et à la mise en danger de cette même diversité. Dans le cas des agrumes, la biodiversité sauvage des centres d'origine que sont la Chine, l'Inde et l'Indonésie par exemple, est menacée par la pression démographique, la déforestation, l'agriculture intensive et l'important développement du tourisme de masse pour ne citer que ces exemples. Il est donc de plus en plus judicieux, en plus d'essayer de mettre en place des mesures de protection des habitats naturels, de maintenir et de développer les conservatoires *ex situ* des ressources phytogénétiques en général et des agrumes en particulier. La conservation des ressources génétiques agrumes, comme nous venons de le voir, ne peut se faire que sous la forme de plantes, ce qui génère des coûts importants. Ces coûts peuvent être décuplés si ces plants sont conservés sous abris *insect-proof* comme c'est le cas dans de nombreux pays où des maladies transmises par insectes vecteurs sont présentes à l'état endémique. Depuis quelques années déjà les chercheurs travaillent sur des alternatives à la conservation de la diversité génétique des agrumes sous forme de plants, comme la cryoconservation de semences, déjà citée plus haut, mais aussi la cryoconservation de cals embryogènes (Duran-Vila *et al.*, 1997; González-Amao *et al.*, 2003), la cryoconservation d'apex caulinaires (Volk *et al.*, 2012) ou la conservation *in vitro* (Engelmann, 1997).

### **3.3. La gestion des ressources génétiques agrumes**

La conservation des ressources génétiques agrumes passe tout d'abord par la prospection et la localisation de nouveaux géotypes susceptibles d'augmenter la diversité génétique des centres de ressources biologiques, puis par l'introduction, la conservation, la caractérisation et l'évaluation de ces ressources et enfin, par la gestion des données ainsi générées grâce notamment à la mise en place de bases de données. C'est globalement le cahier des charges de la mise en place des CRB végétaux dont les procédures sont décrites dans les documents rédigés pour la certification du CRB Citrus Inra-Cirad de San Giuliano par exemple. Il existe différents processus (introduction, conservation, caractérisation, diffusion...) dont les procédures sont rédigées suivant la norme NF S96-900 qui certifie le CRB Citrus Inra-Cirad depuis juillet 2014.

## 4. Les agrumes : importance socio-économique

### 4.1. Le marché des agrumes

#### 4.1.1. La production

Aujourd'hui les agrumes représentent la première production fruitière au monde (si l'on ne prend pas en compte la tomate généralement rangée dans la production de légumes) avec plus de 131 millions de tonnes devant la banane, la pomme ou la vigne (tableau 3 ; FAOSTAT 2014). Les oranges représentent plus de 50% de la production (figure 28) avec 68,2 millions de tonnes suivi par les mandarines et leurs hybrides (27 millions de tonnes), les citrons et limes (15 millions de tonnes) Les cédrats, les kumquats et autres agrumes (12,8 millions de tonnes) et enfin les pamplemousses et pomelos (8 millions de tonnes) dont la consommation est plus limitée, complètent la gamme (tableau 3).

Tableau 3 : Production mondiale 2012 (FAOSTAT; © OAA Division de la Statistique 2014)

Production	Tonnes	Rang mondial
<b>Total agrumes</b>	<b>131 283 333</b>	<b>1</b>
Pastèques	105 372 341	2
Bananes	101 992 743	3
Pommes	76 378 738	4
<b>Oranges</b>	<b>68 223 759</b>	<b>5</b>
Raisins	67 067 129	6
Mangues, mangoustans et goyaves	42 139 837	7
Bananes plantains	37 162 205	8
Melons, cantaloups	31 925 787	9
<b>Tangerines, mandarines, clémentines, satsumas</b>	<b>27 060 756</b>	<b>11</b>
Poires	23 580 845	12
Ananas	23 333 886	13
Pêches et nectarines	21 083 151	14
<b>Citrons et limes</b>	<b>15 118 462</b>	<b>16</b>
<b>Autres agrumes</b>	<b>12 840 318</b>	<b>17</b>
Papayes	12 411 566	18
Prunes et prunelles	10 702 774	19
<b>Pamplemousses et pomelos</b>	<b>8 040 038</b>	<b>20</b>
Fraises	4 516 810	24
Abricots	3 956 640	27
Cerises	2 256 519	30
Kiwis	1 412 455	33

Depuis les années 1970, la production d'agrumes n'a cessé d'augmenter et le commerce mondial agrumicole a connu une incroyable croissance ces 40 dernières années. Les échanges internationaux d'oranges ont augmenté de presque 50 % de 1970 à 2000, les

pomelos (hors pamplemousses) de 99 %, les citrons (hors limes) de plus de 121 % et les petits agrumes de plus 670 % (Imbert, 2013). Depuis l'an 2000 ces envolées se sont ralenties et la course aux rendements dans les pays occidentaux, du fait des coûts élevés de production, laisse la place à la recherche de la qualité en s'appuyant de plus en plus sur de nouvelles variétés (sélections, créations).

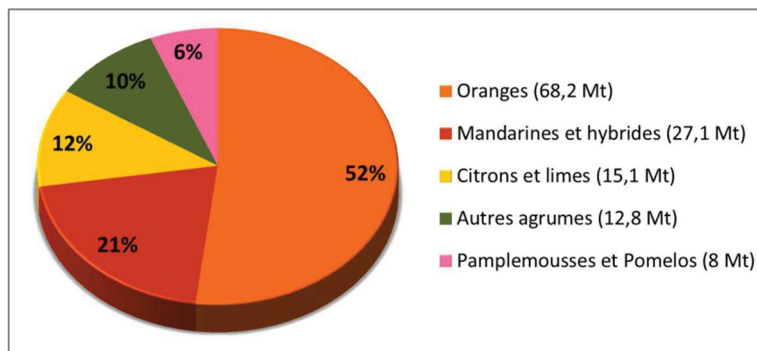


Figure 28 : Parts en pourcentage des différents groupes d'agrumes dans la production mondiale totale de 2012 (131 millions de tonnes [Mt]). (FAOSTAT ; © OAA Division de la Statistique 2014)

La production se concentre sur quatre zones géographiques majeures : la Chine est le premier producteur mondial d'agrumes avec près de 24 millions de tonnes suivie du Brésil, des États-Unis, et du bassin méditerranéen. A eux seuls ils contrôlent les deux tiers de la production mondiale d'agrumes (oranges, petits agrumes, citrons, limes, pamplemousses et pomelos hors autres agrumes). L'Espagne, au sixième rang mondial, est le plus important producteur d'agrumes européen et méditerranéen (tableau 4).

Tableau 4 : Pays producteurs d'agrumes 2012 (FAOSTAT; © OAA Division de la Statistique 2014)

Pays producteurs	Tonnes d'agrumes	Rang Mondial	Rang en Méditerranée	Rang en Europe
Chine, continentale	31 700 000	1		
Brésil	20 258 507	2		
États-Unis d'Amérique	10 619 510	3		
Inde	8 000 000	4		
Mexique	6 750 161	5		
Espagne	5 501 500	6	1	1
Égypte	3 980 151	7	2	
Nigéria	3 900 000	8		
Turquie	3 556 407	9	3	
Italie	2 904 946	10	4	2
Iran (République islamique d')	2 832 500	11		
Argentine	2 815 000	12		
Afrique du Sud	2 314 379	13		
Pakistan	2 032 000	14		



Maroc	1 890 507	15	5	
Indonésie	1 611 784	16		
Thaïlande	1 280 000	17		
Colombie	1 248 187	18		
Algérie	1 088 382	19	6	
Pérou	1 005 518	20		
Japon	1 004 000	21		
Grèce	976 500	22	7	3
Israël	618 412	26	8	
Australie	516 408	31		
Tunisie	454 900	32	9	
Portugal	234 031	44	10	4
Liban	228 500	46	11	
Chypre	102 681	61	12	5
France	46 607	79	15	6

#### 4.1.2. Les surfaces

Les surfaces plantées en agrumes représentent plus de 8,7 millions d'hectares en production (tableau 5). Les agrumes sont cultivables et cultivés tout autour de la planète du 40<sup>e</sup> parallèle nord au 40<sup>e</sup> parallèle sud.

Tableau 5 : Surfaces mondiales de production en hectares 2012 (FAOSTAT ; © OAA Division de la Statistique 2014)

Production	Hectares
Oranges	3 816 692
Tangerines, mandarines, clémentines, satsumas	2 345 020
Autres agrumes	1 353 762
Citrons et limes	980 949
Pamplemousses et pomelos	289 126
<b>Total agrumes</b>	<b>8 785 549</b>

Les agrumes sont très sensibles au terroir et de ce fait, tous les pays producteurs d'agrumes ne produisent pas n'importe quelle variété mais bien celles les mieux adaptées à leur zone de production. Par exemple, la coloration externe des fruits est fortement influencée par la température, c'est le froid qui fait que le zeste des oranges, mandarines et clémentines accumulent caroténoïdes et anthocyanes, c'est donc en zones tempérées à froides que se concentre leur production pour les fruits frais. À l'inverse, en zones chaudes et tropicales seront cultivées les limes qui pourront garder leur coloration verte et ne jauniront pas comme elles le font en Méditerranée où les hivers sont plus froids. Les oranges tropicales restent vertes et ont une pulpe plutôt pâle (Dhuique-Mayer *et al.*, 2005, 2009), elles sont donc essentiellement destinées à l'industrie du jus. Enfin, à part de rares

exceptions, les pamplemousses et pomelos ont besoin de chaleur pour mûrir correctement et se colorer intérieurement (production de lycopène). Le Brésil et les États-Unis (Floride) se sont spécialisés dans l'industrie du jus et représentent 85 % de la production mondiale de jus d'orange. Deux tiers de la production mondiale de jus de citron vient d'Argentine, du Mexique et des États-Unis. 50 % de la production de jus de pomelos vient des États-Unis. À l'inverse, la moitié des agrumes frais exportés à travers le monde vient de Méditerranée, soit 60 % du commerce mondial des oranges et des citrons frais (hors limes) et plus de 70 % des petits agrumes (Imbert, 2013).

#### 4.1.3. Les exportations

De nombreux pays producteurs ont un marché intérieur important et consomment une grande partie de leur production (64 % du marché des agrumes frais est autoconsommé, figure 29) alors que d'autres destinent l'essentiel de leur production à l'exportation (12 % de la production mondiale). Ainsi l'Espagne, au sixième rang de pays producteurs d'agrumes, est le plus important pays exportateur (tableau 6).



Figure 29 : Oranges (production locale) en vrac sur un marché d'Agadir, Maroc ; © F. Curk-Inra.

Tableau 6 : Pays exportateurs d'agrumes en tonnes 2011 (FAOSTAT ; © OAA Division de la Statistique 2014)

Pays exportateurs d'agrumes	Tonnes
Espagne	3 116 943
Afrique du Sud	1 083 239
Égypte	1 058 409
Turquie	837 260
États-Unis d'Amérique	806 882
Chine, continentale	792 720
Maroc	520 124
Grèce	468 900
Pakistan	331 775
Italie	248 285
Argentine	241 823
Brésil	34 232

#### 4.1.4. Les importations

Les pays importateurs sont généralement des pays non producteurs, à l'exception de certains pays tels que les États-Unis qui se sont largement spécialisés dans la production d'agrumes de gros calibres (type pomelos) ou à jus (oranges) et importent d'importantes quantités de petits agrumes de type mandarines et hybrides de mandariniers (tableau 7). La consommation européenne et occidentale en générale a plutôt tendance à se stabiliser et ce sont les pays de l'Est qui dynamisent les échanges. Les pays émergents augmentent leur consommation d'où la forte croissance des marchés intérieurs (autoconsommation).

Tableau 7 : Pays importateurs d'agrumes en tonnes 2011 (FAOSTAT ; © OAA Division de la Statistique 2014)

Pays importateurs d'agrumes	Tonnes	Rang mondial
Fédération de Russie	1 307 709	1
Allemagne	866 766	2
France	747 137	3
Pays-Bas	653 142	4
Royaume-Uni	549 661	5
Arabie saoudite	419 880	6
Canada	332 048	7
Ukraine	325 513	8
Pologne	295 174	9
Iran (République islamique d')	249 248	10
États-Unis d'Amérique	245 075	11
Italie	227 793	12

#### 4.2. Cas particulier des citrons et des limes

Comme nous l'avons vu précédemment les bassins de production de la lime et du citron ne sont pas les mêmes. Le citron frais est principalement cultivé en zones tempérées, en Méditerranée et en Argentine, alors que la lime, qui nécessite un climat plutôt tropical pour rester verte, est principalement produite au Mexique et au Brésil (figure 30).



Figure 30 : Citrons de Corse et d'Espagne et limes importées du Brésil sur un étalage de supermarché, Haute-Corse, 2014 ; © F. Curk-Inra. (N. B. : Les citrons d'Espagne sont emballés [sachets en bas et filets en haut] ce qui permet aux citrons de Corse, en vrac, d'être référencés. Ils sont ainsi facilement différenciables au moment de passer en caisse).

### 4.2.1. Le citron

Le marché du citron frais, qui n'a cessé d'augmenter jusqu'en 2007, connaît depuis une certaine stagnation (figure 31) (Duportal *et al.*, 2013).

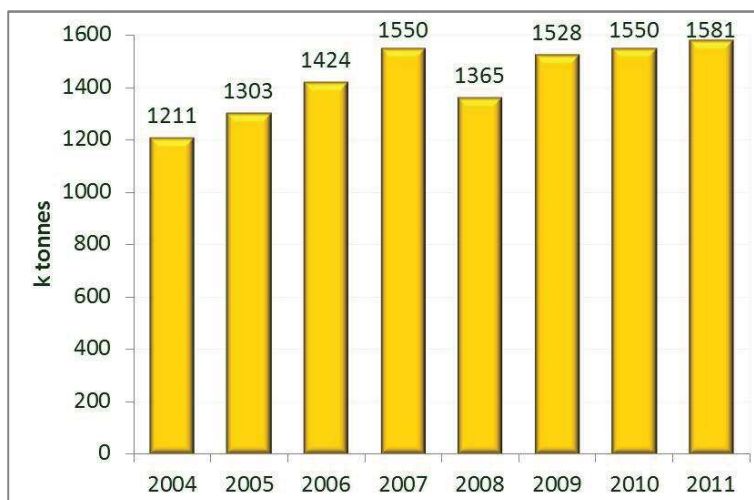


Figure 31 : Marché mondial du citron, en millions de tonnes de fruits échangés par an.

Avec presque 1,6 millions de tonnes de citrons échangés, il est à la 11<sup>ème</sup> place derrière notamment les bananes, pommes, oranges, raisins, pêches et nectarines, mais devant le kiwi. Le marché du citron frais est un marché particulier car ce fruit, de par sa nature particulièrement acide, n'est pas réellement consommé en frais mais utilisé en cuisine comme ingrédient (jus, zeste) ou comme décor (tranche, quartier ou rondelle). Le citron frais est donc particulièrement concurrencé par des produits transformés plus chers mais plus simples d'utilisation (Figure 32). Enfin, le citron est également concurrencé par la lime, dont le marché est en constante augmentation depuis les années 80.



Figure 32 : jus de lime et de citron industriels, prêts à l'emploi, dans un rayon de supermarché, Haute-Corse, 2014 ; © F. Curk-Inra.

L'Europe est le premier débouché du citron avec plus de 50 % du commerce mondial, or les échanges stagnent depuis 2006/2007 et la consommation plafonne en moyenne à 1,6 kg/habitant/an en Europe, les pays du nord de l'Europe étant les plus gros consommateurs (figure 33) (Duportal *et al.*, 2013).

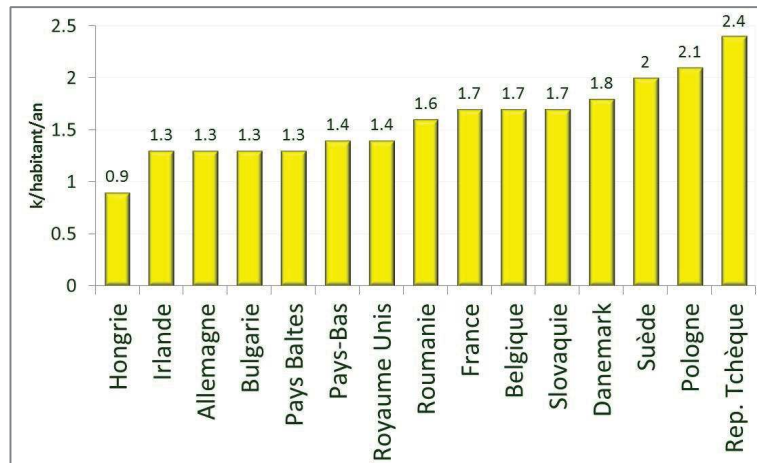


Figure 33 : consommation en kg/habitant/an de citron frais

Face à cette situation un certain nombre de marchés ont choisi de se différencier pour continuer d'exister, les labels de qualité de type IGP et/ou Agriculture Biologique ne semblent pas avoir permis l'essor des ventes à l'international mais semblent porter leurs fruits sur les marchés locaux en augmentant la valeur ajoutée des produits du terroir. C'est le cas de certaines appellations comme celles du citron de Syracuse et de Sorrento en Italie ou du citron de Menton en France (figure 34).



Figure 34 : Fête du citron de Menton, Menton 2007 ; © F. Curk-Inra

Alors que le marché du fruit frais stagne, la consommation mondiale de produits dérivés du citron explose. Ainsi, le prix de l'huile essentielle de citron a quintuplé de la fin des années 90 à 2008 et celui du jus de citron concentré a triplé. C'est un quart de la production mondiale de citrons qui est transformé. Les trois pays qui se partagent l'essentiel du marché sont les États-Unis, en troisième position avec 248 000 tonnes de citrons transformés, puis l'Espagne, en seconde position avec 287 000 tonnes, et enfin l'Argentine, loin devant, avec près de 1 221 000 tonnes. Ces trois pays représentent 85 % des volumes. C'est l'extraordinaire dynamisme du marché des boissons rafraîchissantes sans alcool (BRSA) qui tire le marché des produits dérivés du citron vers le haut. Près de 98 % du jus de citron concentré est utilisé comme ingrédient de boissons comprenant de moins de 5 % à 15 % de jus de fruits. Le sucre étant de moins en moins à la mode en Europe (taxation des boissons sucrées en France par exemple) et aux États-Unis (interdiction des formats de soda *jumbo* dans certains états), ce sont les pays émergents qui sont les principaux moteurs de ce dynamisme (Duportail *et al.*, 2013).

### 4.2.2. Les limes

Contrairement à celui du citron, le marché des limes ou « citrons verts » est en constante augmentation depuis les années 80 (figure 35)

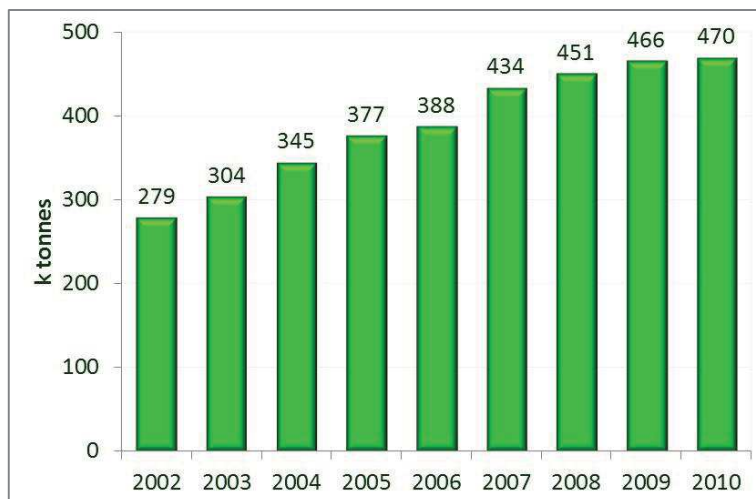


Figure 35 : Marché mondial de la lime, en millions de tonnes de fruits échangés par an.

Le marché est dominé par la lime triploïde, sans pépins de type lime de Tahiti (figure 20 et 36), lime de Perse ou lime Bears (*Citrus latifolia*).



Figure 36 : Lime de type Tahiti, *Citrus latifolia* ; © F. Curk-inra

La lime de type lime Mexicaine, *Key lime* ou lime antillaise (*Citrus aurantifolia*, figure 19 et 37), plus petite et avec un ou deux pépins par fruit, ne représente qu'une part négligeable du marché.

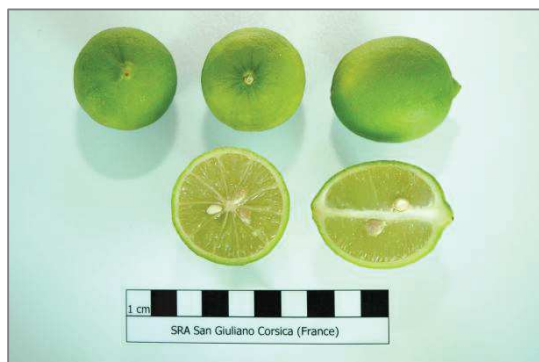


Figure 37 : Lime de type Mexicaine *C. aurantifolia* ; © F. Curk-Inra

Les limes douces (lime douce de l'Inde, lime douce du Brésil, limette à mamelon [figure 23], limonette de Marrakech...) ne sont produites que pour des marchés de niche très localisés dans les zones même de production (figure 38).



Figure 38 : Limonette de Marrakech (*C. limetta* Risso) en saumure sur un marché d'Agadir, Maroc ; © F. Curk-Inra.

C'est le Mexique qui est de loin le premier exportateur de limes (du Mexique mais pas mexicaines) au monde, avec près de 450 000 tonnes exportés en 2012 soit plus de 85 % du marché mondial. Les États-Unis, l'Europe et le Canada représentent plus de 99 % des importations de limes (figure 39) (Duportal *et al.*, 2013).

Au vue du faible nombre de pays importateurs, il existe encore une grande marge de progression pour le marché de ce fruit encore mal connu dans des zones comme l'Europe de l'est par exemple.

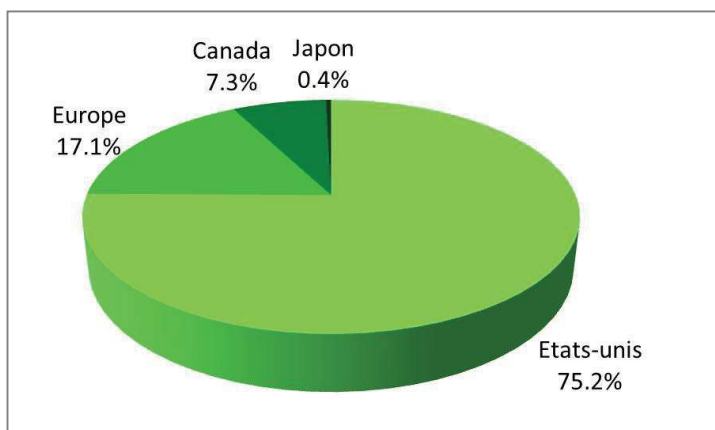


Figure 39 : Importation mondiale de limes par pays, en % (Duportal *et al.*, 2013).

Malgré ses extraordinaires potentialités ce marché n'est pas à l'abri d'une crise importante. En effet, le marché de la lime, en constante augmentation, va devoir faire face à une chute annoncée de la production mexicaine du fait de mauvaises conditions climatiques à répétition (froid) et de l'arrivée récente du huanglongbing<sup>5</sup> en provenance de la Floride et qui menace grandement le verger du premier exportateur de limes au monde. Cette baisse de la production, déjà amorcée, génère une augmentation telle des prix que la consommation nord-américaine se tourne déjà, aux grand dam des consommateurs de cocktails en tous genres, vers du citron jaune dont les prix sont plus bas (Kurtzleben, 2014; Mosends, 2014).

<sup>5</sup> cf. *les stress biotiques*, point 5.2.



## 5. Agrumiculture moderne : les contraintes

L'aire de culture des agrumes est l'une des plus vastes de toutes les plantes cultivées : on les rencontre tout autour du globe du 40<sup>e</sup> parallèle nord au 40<sup>e</sup> parallèle sud. Ils sont donc confrontés à des types de sols et de climats très différents, allant du très sec, au tropical humide. Les agrumes sont donc soumis à une grande diversité de stress biotiques et abiotiques.

### 5.1. Les stress abiotiques

Parmi les stress abiotiques, la salinité (Storey et Walker, 1999) et la sécheresse sont des freins importants au développement de la culture des agrumes. En zones tempérées, comme le bassin méditerranéen, la sécheresse perturbe considérablement le développement végétatif : chute des feuilles (Tudela et Primo-Millo, 1992), diminution du potentiel hydrique et de la conductance stomatique (Gomez-Cadenas *et al.*, 1996), baisse de rendements et de la qualité des fruits (Yakushiji *et al.*, 1998; Jacquemond *et al.*, 2013a). La chlorose ferrique, associée aux sols calcaires et basiques, touche également une grande partie des plants d'agrumes cultivés en Méditerranée (20 à 50 % des arbres (Jaeger *et al.*, 2000)). A l'opposé, les sols trop acides peuvent également être un problème (Ollitrault et Navarro, 2012; Jacquemond *et al.*, 2013a).

### 5.2. Les stress biotiques

Les agrumes sont aussi soumis à de nombreux stress biotiques causés par différents agents pathogènes et ravageurs. Sont considérés comme ravageurs les oiseaux, rongeurs, insectes, etc. qui détruisent les plantes cultivées, notamment en les consommant ou en leur transmettant diverses maladies. Les agrumes subissent les attaques de nombreux agresseurs souvent naturellement présents dans leur environnement de culture ou introduits artificiellement par l'homme. Les insectes sont les principaux ravageurs des agrumes, on peut citer dans l'ordre des Hémiptères les cochenilles, les pucerons et les aleurodes qui créent des dégâts directs (par prélèvement de sève et/ou sécrétion de toxines) ou indirects (par sécrétion de miellat [sur lequel se développe la fumagine, champignons qui créent des taches superficielles veloutées et noirâtres sur les feuilles et les fruits] ou comme agent vecteur de maladies) (Jacquemond *et al.*, 2013c). Dans l'ordre des lépidoptères la mineuse des agrumes (*Phyllocnistis citrella* Stainton) ralentit la croissance en détruisant les jeunes pousses. Enfin, citons la mouche méditerranéenne des fruits (ordre des Diptères) ou cératite (*Ceratitis capitata* Wiedman), qui peut provoquer d'importantes pertes de récolte en pondant dans les fruits dans lesquels se développent ses larves (figure 40).

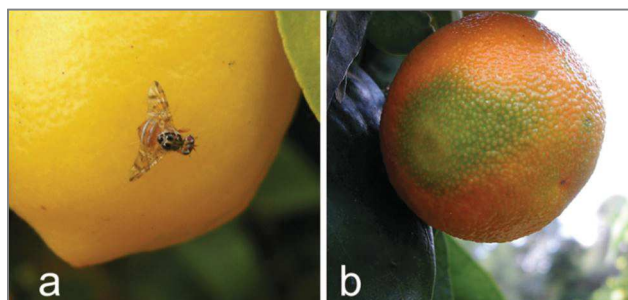


Figure 40 : Cératite ou mouche méditerranéenne des fruits (*Ceratitis capitata* Wiedman) : a. adulte sur fruit ; b. dégât sur clémentine (piqûre de ponte de cératite) ; © F. Curk-Inra.

Comme nous venons de le voir certains insectes sont considérés comme ravageurs non pas pour les dégâts qu'ils provoquent directement sur les plantes cultivées mais du fait qu'ils transmettent un certain nombre de maladies pouvant provoquer le dépérissement de cultures entières. Parmi les maladies à virus transmises par vecteur la tristezza (*Citrus tristeza virus* [CTV]) est la plus connue. Cette affection virale, transmise par les pucerons, est qualifiée de maladie d'association car ne dépérissent que certaines variétés (de type mandarine, orange ou pomelo par exemple) quand elle sont greffées sur bigaradier (*C. aurantium*) (Moreno *et al.*, 2008; Bové et Duran-Vila, 2013) (figure 41).



Figure 41 : Clémentinier greffé sur bigaradier, tué par la tristezza (plaine de la Mitidja, Algérie) ; © F. Curk-Inra

Il existe également des maladies à virus non transmises par insectes vecteurs mais diffusées et transmises par l'homme lors du greffage ou de la taille. C'est le cas de la psorose écailleuse [*Citrus psorosis virus* (CPSV)] qui est une maladie à virus très répandue dans les vieux vergers et dont les symptômes sont toujours localisés sur la partie variété de l'arbre greffé quel que soit le porte-greffe et ne se transmet que par greffage (Martín *et al.*, 2004; Achachi *et al.*, 2014). Parmi les maladies transmises par l'action de l'homme, mais à viroïdes cette fois, nous pouvons citer l'exocortis (*Citrus exocortis viroid* [CEVd]) qui provoque des écailllements de l'écorce des porte-greffe sensibles et des baisses considérables de rendement. Cette maladie est transmise par greffage et par outils de taille (Bové et Duran-Vila, 2013). En général, les porte-greffe résistants à la tristezza sont particulièrement sensibles à l'exocortis (Jacquemond *et al.*, 2013a).

D'autres maladies, dues à des bactéries endogènes, sont aujourd'hui une des préoccupations majeures du monde agrumicole. Le huanglongbing ou maladie des pousses jaunes ou du dragon jaune (HLB) est présent depuis longtemps en Afrique et en Asie, il commence à prendre une importance considérable dans les zones majeures de production des agrumes sur le continent américain (Floride, Brésil, Mexique). Cette maladie est causée par diverses espèces de bactéries du genre *Liberibacter* et est transmise par greffage et par insectes vecteurs (psylles). Il n'existe aujourd'hui aucun moyen de lutte efficace. Le bassin méditerranéen ne semble pas à l'abri de l'arrivée prochaine de cette maladie du fait de la présence officielle, à l'ouest (à Madère et aux îles Canaries), du psylle africain (*Trioza*

*erytrae* Del Guercio), et à l'Est (en Arabie Saoudite et en Iran), du psylle asiatique (*Diaphorina citri* Kuwayama) (Bové et Duran-Vila, 2013). Une autre bactérie du xylème, de la famille des *Xanthomonadaceae*, *Xylella fastidiosa* responsable de la chlorose variéguée des agrumes (CVC, *Citrus Variegated Chlorosis*) présente au Brésil depuis la fin des années 80 vient de faire son apparition en Europe, en Italie, en octobre 2013, après deux alertes en 2011 et 2012 en France (Anses, 2012).

Citons également les maladies à bactéries exogènes non transmises par greffage dont le chancre citrique répandu dans les vergers d'orangers, de pomelos et de citronniers des grands pays producteurs mais pour l'instant absent du bassin méditerranéen. Cette maladie est due à la bactérie *Xanthomonas axonopodis* pv. *Citri* (Hasse) (Das, 2003; Bové et Duran-Vila, 2013; Jacquemond *et al.*, 2013c).

Enfin, citons les maladies à champignons, dont le mal secco (figure 42) récemment détecté en Corse mais présent sur la Côte d'Azur depuis plusieurs années et qui s'attaque spécifiquement aux citronniers et cédratiers (Fredon Corse, 2014) ou la gommosse à *Phytophthora*, maladie présente dans toutes les régions de production des agrumes. Cette dernière maladie, propagée à travers le monde par l'intensification de la culture des agrumes au XIX<sup>e</sup> siècle, est responsable de la mort de la plupart des plants d'agrumes cultivés francs de pied (non greffés) (Cohen *et al.*, 2003; Bové et Duran-Vila, 2013) d'où la généralisation de la culture des agrumes sous forme d'arbres greffés.



Figure 42 : Cédratier tué par le mal secco ; © F. Curk-Inra

## 6. Agrumiculture moderne : les apports de l'amélioration des plantes en réponse aux enjeux et contraintes

Les agrumes sont cultivés sous la forme d'arbres greffés (figure 43), l'amélioration génétique vise donc à l'amélioration des performances des scions comme des porte-greffe. Les objectifs d'amélioration portent essentiellement sur des critères de qualité, de production fruitière ainsi que de résistances ou de tolérances à des contraintes biotiques et abiotiques, aériennes et sous-terraines.



Figure 43 : Clémentiniers greffés sur citrange en Corse ; © F. Curk-Inra.

### 6.1. Les objectifs d'amélioration et de sélection des porte-greffe

Les porte-greffe jouent un rôle clef dans la résistance à certaines maladies et influencent de manière importante le comportement d'une variété non seulement en lui permettant de s'adapter aux conditions pédoclimatiques mais également en agissant sur ses performances agronomiques.

La principale raison du développement du greffage chez les agrumes à partir des années 1870 fut la lutte contre la gombose à *Phytophthora* (Bové et Duran-Vila, 2013). On sait aujourd'hui que cette gombose est provoquée par un ensemble de champignons du sol appartenant au genre *Phytophthora*. La principale méthode de lutte contre cette maladie est l'utilisation d'un porte-greffe résistant. Le bigaradier (*C. aurantium*), utilisé comme porte-greffe, fut la toute première solution proposée avec succès. En 1930, 90 % des agrumes cultivés dans le monde étaient greffés sur bigaradier. Malheureusement, cette solution miracle contre les *Phytophthora* s'est révélée inefficace contre un des virus les plus destructeurs qu'ait connu l'agrumiculture : la tristeza. Dès 1931 on signale des dépérissements d'agrumes greffés sur bigaradier, d'abord en Argentine, puis au Brésil, puis en Espagne. Entre 1930 et 1945, 16 millions d'arbres sont mort en Argentine, 10 millions au Brésil et, entre 1957 et 1986, 10 millions rien que dans la région de Valence en Espagne (Bové et Duran-Vila, 2013). C'est cette maladie qui a conduit à la mise en place de programmes de sélection de nouveaux porte-greffe dans le monde entier.

Aujourd'hui, en plus des conditions *sine qua non* à la sélection d'un porte-greffe d'agrumes que sont la tolérance ou la résistance aux *Phytophthora* du sol, aux nématodes et à la tristeza, se rajoutent d'autres critères tels que l'adaptation aux conditions environnementales (type de sols, salinité, sécheresse, excès d'eau, tolérance au froid...) et une influence positive sur la production en quantité et en qualité. Enfin, notamment en régions agrumicoles tropicales où les arbres ont tendance à se développer très rapidement,

rendant la récolte plus compliquée, ou dans les régions plus tempérée, pour augmenter les densités de plantation, des porte-greffe nanisant, limitant la vigueur des arbres, sont également recherchés (Ollitrault et Navarro, 2012).

Un autre élément important pour la sélection d'un porte-greffe a longterm été sa capacité à produire des graines polyembryonnées permettant sa propagation à l'identique (apomixie partielle<sup>6</sup>) par semis. Ce caractère qui est un obstacle à l'obtention d'hybrides ou de descendants par autofécondation dans les programmes d'amélioration variétale (Ollitrault et De Rocca Serra, 1992) est un atout pour la multiplication clonale des porte-greffe. Aujourd'hui, la propagation *in vitro* de porte-greffe semble permettre de s'affranchir de ce caractère.

## 6.2. Les objectifs de l'amélioration des scions

Les objectifs de l'amélioration variétale varient en fonction des caractéristiques du marché visé (fruits frais, transformés, marché local, export...) et des conditions environnementales de la zone de production, du terroir (type de sol, climat, disponibilité et qualité de l'eau, savoir-faire et pratiques culturelles...). L'industrie du jus, par exemple, cherche à améliorer les taux de jus et de sucre des fruits, alors que pour la production de fruits frais les qualités pomologiques (calibre, couleur, facilité d'épluchage, résistance au transport...) et organoleptiques (arômes, goût, taux de jus, acidité, taux de sucre...) des fruits sont des caractéristiques prépondérantes. Parmi les critères importants de sélection variétale d'agrumes destinés au marché du fruit frais, l'absence de pépin est un caractère essentiel. L'auto-incompatibilité, comme chez le clémentinier, la stérilité mâle et femelle, comme chez l'oranger navel, et la parthénocarpie sont donc particulièrement recherchés. Paradoxalement, comme nous le verrons dans le chapitre suivant, ces caractères sont aussi un frein à la mise en place de programme d'amélioration variétale par hybridation sexuée. Aujourd'hui, de nouveaux objectifs apparaissent dans les projets de création variétale comme l'étalement de la production et l'amélioration de la qualité nutritionnelle des fruits (teneur en vitamine C, en caroténoïdes et composés phénoliques) (Ollitrault et Navarro, 2012).

Au-delà de la qualité au sens large, l'amélioration variétale cherche également à augmenter la résistance ou la tolérance aux maladies tels que le huanglongbing, le chancre citrique, le mal secco ou l'alternariose (Cuenca *et al.*, 2013a).

## 6.3. Les facteurs limitant l'amélioration génétique chez les agrumes

L'amélioration variétale des espèces secondaires d'agrumes présente un intérêt économique majeur (plus de 80% de la production mondiale des agrumes). Cependant, à ce jour, elle est difficilement envisageable par des schémas d'amélioration conventionnelle (recombinaisons sexuées) en raison de diverses caractéristiques biologiques, telles que l'apomixie, la stérilité, l'incompatibilité sexuelle, et des structures génomiques d'origine interspécifique hautement hétérozygotes (Grosser et Gmitter, 1990).

---

<sup>6</sup> cf. *Les facteurs limitants l'amélioration génétique chez les agrumes*, point 6.3.

L'apomixie facultative, c'est-à-dire la formation d'embryons d'origine nucellaires (sans méiose, ni fécondation), aboutit à des plants génétiquement identiques à la mère. Comme leur nom l'indique, les pépins polyembryonnés renferment plusieurs embryons. Un embryon zygotique, résultant de la fécondation, est présent dans la graine et est accompagné par un, à parfois plus d'une dizaine d'autres embryons provenant des cellules du tissu nourricier de l'ovaire (le nucelle). Le développement du, ou des embryons nucellaires (figure 44), est généralement plus rapide que celui de l'embryon zygotique qui souvent avorte (Rebour, 1966). Ce phénomène pose évidemment un problème au moment d'obtenir des populations de descendants hybrides en nombre suffisant pour réaliser de la sélection.



Figure 44 : Plantules d'agrumes se développant à partir du semis d'une seule graine ; © F. Curk-Inra.

Un certain nombre de variétés commercialement intéressantes (orange navel ou mandarines satsuma par exemple) posent problème au moment de les utiliser comme parents dans des programmes d'hybridation du fait de leur stérilité gamétique mâle et/ou femelle. Ces stérilités peuvent être d'ordre génétique ou physiologique (en raison du froid par exemple dans certaines régions). Il existe également des incompatibilités sexuelles qui compliquent encore les programmes d'hybridation (Soost et Cameron, 1969; Soost et Cameron, 1975). En effet, de nombreuses variétés, telles que les clémentiniers et d'autres hybrides de types tangors (*C. reticulata* x *C. sinensis*) et tangelos (*C. reticulata* x *C. paradisi*), ainsi que tous les pamplemoussiers (*C. maxima*) sont auto-incompatibles. Certaines variétés auto-incompatibles sont également inter-incompatibles et il existe même des cas d'inter-incompatibilité unilatérale (Ollitrault et De Rocca Serra, 1992).

Le fort niveau d'hétérozygotie, que l'on rencontre chez la plupart des espèces d'agrumes, génère des descendances hybrides très hétérogènes (Herrero *et al.*, 1996; Ollitrault *et al.*, 2003). Il est donc particulièrement difficile de réunir chez un hybride tous les caractères intéressants de ses parents, sans parler des problèmes d'endogamie qu'il n'est pas rare de rencontrer dans les descendances hybrides (Barrett et Rhodes, 1976) compte tenu de l'apparentement de beaucoup d'agrumes.

À toutes ces spécificités biologiques qui compliquent l'amélioration génétique classique par voie sexuée, il faut rajouter une période de juvénilité particulièrement longue chez les agrumes (4 à 9 ans en fonction des espèces et du climat de culture ; Aubert et Vullin, 1997), la méconnaissance des modes de transmission de la majorité des caractères d'intérêts agronomiques, la rareté des marqueurs liés à ces caractères et la présence de pépins dans la grande majorité des hybrides obtenus.

## 6.4. Les principales stratégies d'amélioration

Compte tenu de l'ensemble des contraintes génétiques que nous venons de citer chez les agrumes, les chercheurs développent, en plus de la recombinaison sexuée classique au niveau diploïde, d'autres pistes d'amélioration. Chez les agrumes commerciaux de type oranges, citrons et pomelos, la faible diversité génétique au sein de ces groupes et le fort niveau d'hétérozygotie empêchent l'amélioration variétale par hybridation sexuée et obligent à la sélection de mutations, spontanées ou induites, ou à la transformation génétique. Chez les mandarines, beaucoup plus diversifiées, en plus de sélection de mutants (au sein du groupe des clémentines ou des satsumas par exemple) les programmes de créations variétales par hybridations sexuées ont permis d'obtenir de nouvelles variétés diploïdes et triploïdes dont certaines se retrouvent déjà sur les marchés (Russo *et al.*, 2004; Williams et Roose, 2004; Tokunaga *et al.*, 2005; Aleza *et al.*, 2010; Cuenca *et al.*, 2010; Jacquemond *et al.*, 2013a).

### 6.4.1. La sélection de mutants spontanés et la mutagenèse induite

La diversité variétale connue des orangers, des mandariniers Satsuma et des clémentiniers est issue de mutations spontanées sélectionnées et multipliées par l'homme. Fort de ce constat les sélectionneurs ont mené de nombreux programmes de mutagenèse induite par irradiations aux rayons gamma sur bourgeons et semences ont été conduits depuis les années 60 dans différents pays (Roose et Williams, 2007). De tous ces programmes sont principalement sorties des variétés sans pépins obtenues à partir de variété fertiles. Le pomelo Star Ruby, issue d'une irradiation de semences de pomelo Hudson réalisée au Texas et diffusé à partir de 1970 (Hensz, 1971, 1977), est un exemple commercial qui a ouvert la voie à d'autres projets. Plus récemment, de nouveaux programmes essentiellement dirigés vers l'amélioration variétale de la gamme des petits agrumes de type mandarines et hybrides de mandarines ont donné quelques résultats intéressants tels que les variétés Tango (mutation induite par irradiation de la variété Afourer ou Nadorcott aux Etats-Unis), Orri (mutation induite par irradiation de la mandarine Orah [Mandarine Temple x Mandarine Dancy] en Israël) ou Nero et Nulesin (mutations induites par irradiation de la clémentine Nules en Espagne) (Asíns *et al.*, 2002; Jacquemond *et al.*, 2013a). Les variétés obtenues par mutagenèse induite ne sont pas considérées comme des organismes génétiquement modifiés et ne sont donc pas soumises à la réglementation sur les OGM.

### 6.4.2. La recombinaison sexuée au niveau diploïde

Malgré l'organisation en genres et espèces des différentes classifications déjà citées<sup>7</sup>, l'inter-fertilité est largement répandue chez les agrumes au niveau intra et inter spécifique mais également au niveau inter générique (figure 45). Même si l'agrumiculture continue d'exploiter majoritairement des mutants et des hybrides spontanés, du fait des obstacles biologiques déjà cités, qui rendent plus difficile l'hybridation dirigée, l'inter-fertilité des agrumes continue, encore aujourd'hui, d'être exploitée par les sélectionneurs.

---

<sup>7</sup> cf. *Les agrumes : botanique et origines génétiques*, point 2

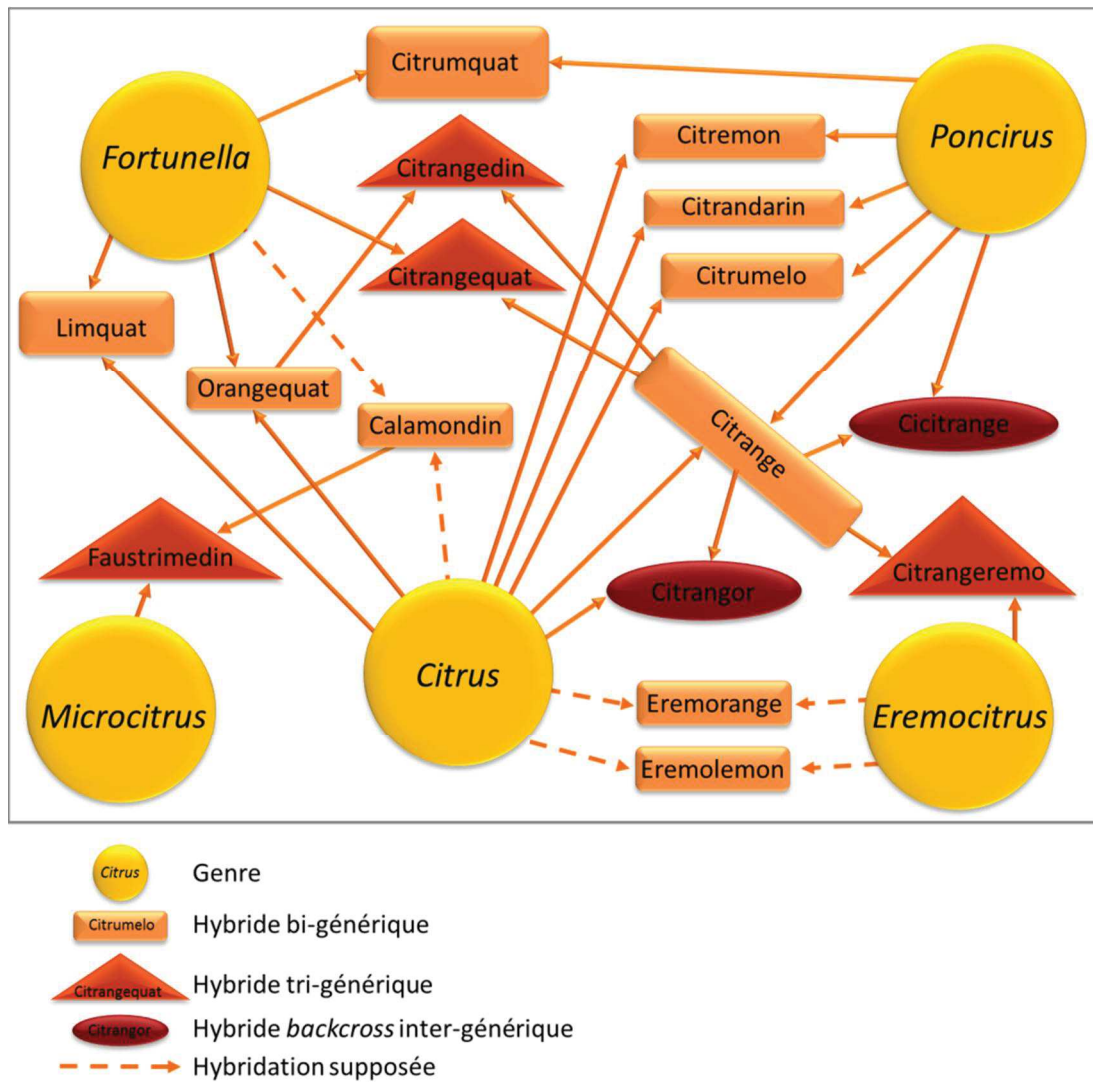


Figure 45 : Possibilités d'hybridation inter-générique chez les agrumes (d'après Swingle et Reece, 1948)

Il est donc possible de créer des hybrides par pollinisation contrôlée même entre genres botaniques différents chez les agrumes. Les premiers hybrides réalisés de manière dirigée le furent, semble-t-il, à la fin du XIX<sup>e</sup> siècle, par Walter Tennyson Swingle, entre un *Poncirus trifoliata* et un oranger (Swingle et Reece, 1967). Les citranges ainsi obtenus ouvrirent la voie à l'obtention d'un grand nombre d'hybrides très différents (Swingle et Reece, 1967). Ces premières hybridations avaient pour objectif de créer de nouvelles variétés ayant les qualités gustatives de l'orange et les propriétés de résistance au froid du *Poncirus*. Au final, aucune ne fut intéressante pour la consommation mais beaucoup sont encore utilisées aujourd'hui comme porte-greffe des agrumes. Il existe d'autres hybrides inter-génériques créés par l'homme et notamment entre les genres *Fortunella* et *Citrus* : orangequat ou limequat par exemple (figure 46).



Ces hybrides inter-génériques fertiles peuvent eux-mêmes entrer dans des programmes d'hybridation et donner ainsi soit des hybrides tri-génériques (Citrangequat par exemple, figure 47) ou des croisements en retour (*backcross*) avec un des parents (figure 48). Les hybridations interspécifiques au sein d'un même genre sont tout de même plus fréquentes et plus recherchées pour l'amélioration génétique des variétés d'agrumes. Le groupe des petits agrumes, de type mandarine, est un des plus intéressants et des plus riches en diversité créée par pollinisations dirigées. La mandarine commune puis la clémentine dans le bassin méditerranéen, ou les satsumas au Japon notamment, avaient un tel succès commercial qu'un grand nombre de programmes de création variétale a cherché à diversifier et à améliorer ces fruits colorés et parfumés mais pas toujours sans pépins. C'est ainsi que dans les années 40 un grand nombre d'hybridations a été réalisé en intra spécifique, entre mandariniers, et en interspécifique, entre mandariniers et d'autres hybrides interspécifiques naturels tels que les orangers et les pomelos avec pour objectif l'augmentation des calibres notamment (figure 48).



Figure 46 : Limequat Eustis ; © F. Curk-Inra.



Figure 47 : Citrangequat 4 saisons ; © F. Curk-Inra.

Sont arrivés sur le marché des hybrides interspécifiques entre mandariniers telles que Fortune et Kinnow par exemple ainsi que des tangors, hybrides entre mandariniers et orangers dont les tangors Ortanique (Alonso et Howell, 1983), ou Kiyomi (Combrink *et al.*, 2013) par exemple, et des tangelos, hybrides entre mandariniers et pomelos (ou pamplemoussiers) dont les tangelos Orlando et Minneola (Swingle et Reece, 1967) eux même réutilisés comme géniteurs pour d'autres hybridations (figure 48). Le suffixe tang- vient de *tangerine*, nom générique des mandarines en anglais, venant lui-même de l'origine tangeroise des mandarines commercialisées en Angleterre, via Gibraltar, tandis que les « or » et « gelo » de tangor et tangelo renvoient réciproquement à ORange et *pummELO*.

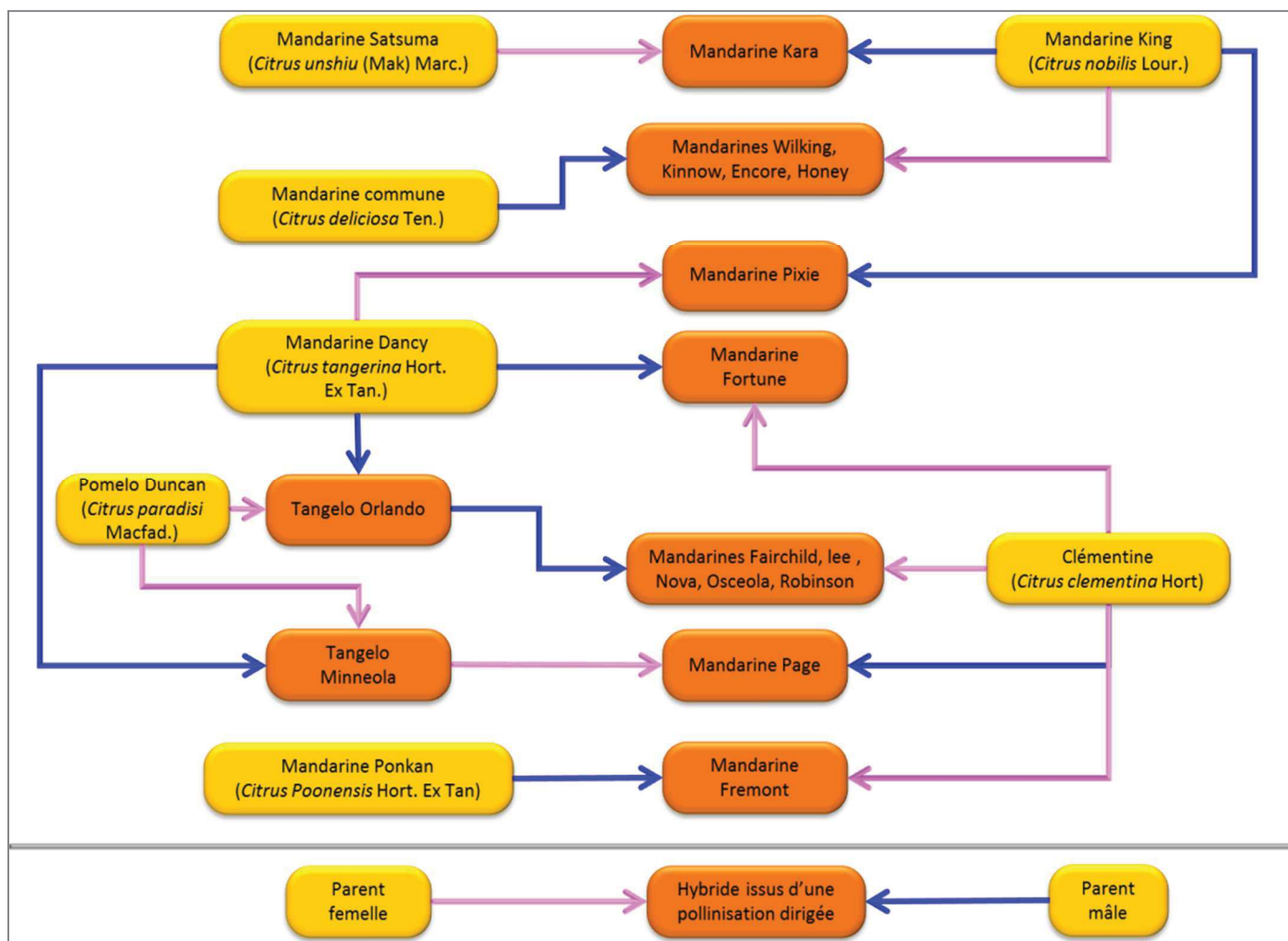


Figure 48 : Hybrides inter-spécifiques (espèces selon Tanaka, 1954) issus de pollinisations dirigées de mandariniers d'après Blondel (Blondel, 1978).

### 6.4.3. La polyploïdie

Les agrumes sont généralement diploïdes et ont 9 paires de chromosomes pour une taille du génome estimé à 367 Mb (Ollitrault *et al.*, 2003). Chez de nombreuses espèces végétales, il existe des individus polyploïdes portant trois, quatre, cinq, six (et même davantage) lots de chromosomes homologues. Dès 1941 Jean Rostand proposait pour expliquer la polyploïdisation une anomalie survenue au cours de la formation des cellules sexuelles, ou bien tout au début du développement du zygote (Rostand, 1941). On sait aujourd'hui que la triploïdie chez les agrumes découle d'une méiose incomplète générant des gamètes diploïdes (mâle ou femelle) au lieu d'haploïdes. Ces gamètes, quand ils sont viables, s'associent à des gamètes normaux (haploïdes) et donnent des zygotes triploïdes (Aleza *et al.*, 2010). La triploïdie conduit généralement à des méioses déséquilibrées ayant des conséquences importantes sur la fertilité des individus, comme la stérilité mâle et femelle. C'est le cas pour les limes de type Tahiti (*C. latifolia*) triploïdes naturels stériles mâle et femelle et donc sans pépins (figure 36).

Les programmes d'amélioration variétale se sont donc rapidement intéressés à cette caractéristique (stérilité) pour générer des variétés notamment de petits agrumes de types mandarines et hybrides de mandarines stériles, sans pépins et non pollinisateurs d'autres variétés pour éviter, en cas de culture proche, la présence de pépins dans les clémentines par exemple (Russo *et al.*, 2004; Williams et Roose, 2004; Cuenca *et al.*, 2010). Les voies d'obtention de triploïdes peuvent varier (Ollitrault *et al.*, 2008) : la sélection de triploïdes spontanés après pollinisation dirigée entre deux parents diploïdes et sauvetage d'embryons *in vitro* (Aleza *et al.*, 2010), l'hybridation dirigée entre diploïdes et tétraploïdes (Recupero *et al.*, 2005; Tokunaga *et al.*, 2005; Aleza *et al.*, 2012), et la fusion de protoplastes entre haploïdes et diploïdes (Ollitrault *et al.*, 2000; Vilorio et Grosser, 2005).

La tétraploïdie quant à elle, découle généralement, chez les agrumes, d'un doublement du stock chromosomique de cellules nucellaires, à l'origine des embryons surnuméraires chez les variétés polyembryonnées (Aleza *et al.*, 2011). Le kumquat Hongkong (*Fortunella hindsii*, figure 8) est un exemple de tétraploïde naturel chez les agrumes. Contrairement aux triploïdes, les tétraploïdes sont généralement fertiles. En revanche, leur comportement diffère du diploïde dont ils sont issus en terme de vigueur des arbres et de tolérance aux stress abiotiques par exemple (Allario *et al.*, 2013). Ces caractéristiques font que les programmes d'amélioration génétique des porte-greffe s'intéressent de près à ces phénomènes de tétraploïdisation dans l'espoir d'améliorer les tolérances à certaines contraintes abiotiques ainsi que pour l'obtention d'arbres moins vigoureux à volume réduit. Des hybrides tétraploïdes peuvent également être générés dans des croisement interploïdes (2x x 4x) par la fécondation d'un ovule 2n par un pollen tétraploïde (Ollitrault *et al.*, 2008) ou par hybridation somatique symétrique entre deux individus diploïdes (Dambier *et al.*, 2011; Grosser et Gmitter, 2011).

#### 6.4.4. Les cybrides

La fusion de protoplastes permet également de générer des individus dit cybrides, possédant les organites cytoplasmiques (mitochondries et chloroplastes) d'une espèce et le noyau d'une autre. Les interactions nucléocytoplasmiques peuvent permettre de modifier le phénotype (Bassene *et al.*, 2008; Bassene *et al.*, 2011) sans totalement remettre en cause les structures génomiques complexes des espèces secondaires comme le ferait la recombinaison sexuée. Ainsi, le cytoplasme mâle stérile de la mandarine satsuma, (système de stérilité nucléocytoplasmique), a été utilisé dans certains programmes d'amélioration génétique des mandariniers par fusion de protoplastes (Guo et Grosser, 2004). L'intérêt des cybrides est également étudié pour l'amélioration des porte-greffe (Dambier *et al.*, 2011).

#### 6.4.5. La transgénèse

L'objectif de la transgénèse est l'introduction d'un ou plusieurs gènes dans un génotype afin de modifier un ou des caractères sans modifier significativement son phénotype. Cette technique, peu utilisée en Europe sur agrumes, du fait de la défiance des consommateurs vis-à-vis des plantes transgéniques, est une voie explorée par d'autres pays (Chine, Brésil, États-Unis) comme outils de sélection (Pasquali *et al.*, 2009) ou pour la création de variétés et porte-greffe résistants à des maladies aussi dévastatrices que la tristeza (Orbovic *et al.*, 2011), le chancre citrique ou le huanglongbing (Grosser *et al.*, 2011;

Stover *et al.*, 2013). Elle constitue par ailleurs un outil très puissant pour les études de physiologie moléculaire (Cervera *et al.*, 2010; Agüero *et al.*, 2012; Pons *et al.*, 2012).

#### **6.4.6. La mutagenèse dirigée**

Il s'agit, connaissant le génome d'une variété, de le modifier ponctuellement par mutagenèse dite dirigée (Heckman et Pease, 2007). Non encore utilisée sur agrumes, cette technique a permis, sur colza par exemple, l'obtention de mutants nains plus résistants et permettant d'optimiser les rendements à la récolte (Liu *et al.*, 2010). Cette technique semble complémentaire des études récentes de séquençage et les nouvelles connaissances génomiques acquises sur agrumes (Xu *et al.*, 2013; Wu *et al.*, 2014). Il s'agit en fait d'une application intéressante du décryptage fin du génome des variétés cultivées et de l'identification des gènes d'intérêts. Il deviendra possible de modifier des caractères simples par mutagenèse dirigée. La frontière entre transgénèse et mutagenèse dirigée est relativement fine et les éventuelles variétés obtenues par mutagenèse dirigée peuvent être, en fonction des évolutions législatives, soumises à la réglementation sur les OGM.

#### **6.4.7. La reconstruction des structures génomiques hybrides**

Il devrait être possible, dans un futur proche, de reconstruire des structures génomiques hybrides à partir de combinaisons de génotypes des taxons ancestraux qui constituent les principaux réservoirs de diversité des formes cultivées d'agrumes. Cette approche pourrait être envisagée à condition de connaître très finement les structures du génome des espèces secondaires et de repérer les points de recombinaison entre chromosomes des taxons ancestraux.

## 7. Les agrumes : évolution des outils moléculaires

Une bonne connaissance de la diversité génétique des agrumes est indispensable à la bonne gestion des ressources phytogénétiques mais également à l'efficacité des programmes d'amélioration génétique. La caractérisation moléculaire est donc une priorité pour toute personne ou organisme en charge de la gestion de ressources génétiques végétales. Différents outils, en constante évolution, sont d'ores et déjà disponibles, certains utilisés à grande échelle et d'autre en cours d'évaluation.

En 20 ans, l'évolution des marqueurs moléculaires, comme outils d'étude de la diversité et de la phylogénie des plantes, a été spectaculaire. Rien que sur agrumes nous pouvons citer les exemples suivants (tableau 8) :

### 7.1. Analyse des génomes nucléaires

#### 7.1.1. AFLP (*Amplification Fragment Length Polymorphism*)

La technique de l'AFLP est basée sur la mise en évidence conjointe de polymorphisme de site de restriction et polymorphisme d'hybridation d'une amorce de séquence arbitraire. Elle combine les techniques RFLP (*Restriction Fragment Length Polymorphism*) et l'amplification d'ADN par PCR, il s'agit d'une amplification sélective de fragments de restriction d'ADN (Zabeau et Vos, 1993; Vos *et al.*, 1995). Parmi les avantages de cette technique la plus importante est l'abondance des polymorphismes mis en évidence. En effet, plusieurs dizaines de fragments (entre 50 et 100 fragments de 100 à 500 pb) peuvent être visualisés en une seule PCR, et le tout sans connaissance préalable des séquences. Ses inconvénients majeurs sont : une résolution parfois imparfaite des profils de bandes, le risque de générer des fragments de même taille à partir de séquences non homologues (Pang *et al.*, 2007) et le caractère majoritairement dominant de ce type de marqueur. Cette technique a pourtant été utilisée en association avec des microsatellites pour le repérage de plantules zygotiques dans des semis de pépins de citrons (Scarano *et al.*, 2003), pour des études de phylogénie du genre *Citrus* et apparentés (Pang *et al.*, 2007; Li *et al.*, 2010; Biswas *et al.*, 2011; Nartvaranant et Nartvaranant, 2011; Pessina *et al.*, 2011; Zhang *et al.*, 2011), de cartographie génétique de certaines variétés (Carlos de Oliveira *et al.*, 2007), de caractérisation de génotypes face à certain stress (Al-Sadi *et al.*, 2012; Zhou *et al.*, 2013) ou encore l'étude d'expression génique (Bernardi *et al.*, 2010; Fang *et al.*, 2010; Kepiro et Roose, 2010; Li *et al.*, 2010; Distefano *et al.*, 2011; Licciardello *et al.*, 2011; Xiao *et al.*, 2011; Kacar *et al.*, 2012).

#### 7.1.2. CAPS (*Cleaved Amplified Polymorphic Sequences*) ou PCR-RFLP

Les polymorphismes sont révélés par digestion enzymatique de fragments d'ADN issus de PCR (d'où l'appellation PCR-RFLP) puis séparés par migration sur gel de polyacrylamide. SNP et Indels peuvent être mis en évidence, il s'agit de marqueurs facilement reproductibles, codominants et nécessitant peu de quantité d'ADN. Cette technique a été utilisée pour l'étude du génome nucléaire des agrumes (Omura *et al.*, 2000; Jiang Dong *et al.*, 2010; Amar *et al.*, 2011; Yamamoto *et al.*, 2011; Shimada *et al.*, 2014).

### 7.1.3. Indel (*Insertions - délétions*)

Le mot Indel, ou polymorphisme d'insertion-délétion, désigne une insertion ou une délétion présente dans une séquence d'ADN repérable par comparaison de la longueur de fragments d'une même zone du génome, amplifiés par PCR. Les Indels apparaissent suite à l'insertion d'un retrotransposon ou tout autre élément mobile, suite à un décalage d'une séquence simple lors de la réplication ou d'événements de rétrocroisements inégaux. Comme pour les SNP, les probabilités d'homoplasie sont faibles ce qui les rend très adaptés aux études de phylogénie (Britten *et al.*, 2003). De plus, ils sont facilement mis en évidence par électrophorèse de produits de PCR ciblée (Vasemagi *et al.*, 2010). Les indels ont été utilisés lors d'études sur le blé (Raman *et al.*, 2006) et le riz (Hayashi *et al.*, 2006) et plus récemment sur agrumes (Ollitrault *et al.*, 2011; Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2012; Ollitrault *et al.*, 2012a; Snoussi *et al.*, 2012; Garcia-Lor *et al.*, 2013b).

### 7.1.4. IRAP (*Inter-Retrotransposon Amplified Polymorphism*)

Il s'agit d'une technique qui se base sur l'amplification par PCR de fragments de génome situés entre deux rétrotransposons (séquences ADN endogènes capables de se déplacer, présentes en grande quantité et distribuées aléatoirement le long du génome des plantes) à l'aide d'amorces spécifiques homologues à ses séquences. Cette technique a l'avantage de mettre en évidence d'importants polymorphismes même si beaucoup sont dominants. Sur agrumes, au sein des espèces secondaires, il a été observé que le polymorphisme basé sur les éléments transposables sont plus nombreux que ceux basés sur les marqueurs de séquences aléatoires ou les microsatellites (Bretó *et al.*, 2001). Ces marqueurs ont été utilisés dans une analyse génétique plus récente de 48 variétés du genre *Citrus* et apparentés (Biswas *et al.*, 2010), ainsi que dans des études de diversité et de cartographie (Bernet *et al.*, 2010; Santos *et al.*, 2013).

### 7.1.5. ISSR (*Inter Simple Sequence Repeat*)

Il s'agit d'amplifier une séquence d'ADN comprise entre deux microsatellites relativement proches. Cette technique peut être utilisée sans connaissance préalable de la zone à amplifier et est très facilement reproductible. Toutefois, l'homologie entre les bandes est parfois douteuse. Ce sont de plus des marqueurs dominants. Sur agrumes ces marqueurs ont été utilisés dans des études de la diversité génétique et des relations phylogénétiques entre espèces du genre *Citrus* (Fang *et al.*, 1997; Fang *et al.*, 1998; Gulsen et Roose, 2001a; Gulsen et Roose, 2001b; Nematollahi *et al.*, 2009; Biswas *et al.*, 2010; Yang *et al.*, 2010; Golein *et al.*, 2011; Uzun *et al.*, 2011a; Lombardoa *et al.*, 2012; Kumar *et al.*, 2013; Santos *et al.*, 2013; Tripolitsiotis *et al.*, 2013).

### 7.1.6. RAPD (*Random Amplified Polymorphic DNA*)

Il s'agit d'une amplification de segments d'ADN au hasard à l'aide d'amorces courtes de séquences choisies arbitrairement. Ce furent les tout premiers marqueurs d'ADN utilisés sur agrumes pour analyser des accessions appartenant au genre *Citrus*, *Microcitrus*, *Poncirus*, *Fortunella* et *Eremocitrus* (Luro *et al.*, 1995; Federici *et al.*, 1998). Ces marqueurs furent également utilisés pour le repérage de plantules zygotiques dans un semis de mandarines (Bastianel *et al.*, 1998), de bigarades (Rao *et al.*, 2007; Rao *et al.*, 2008) et de

Rough lemon (Savita *et al.*, 2012), ainsi que pour de la cartographie génétique (Ruiz et Asins, 2003; de Oliveira *et al.*, 2004; Biswas *et al.*, 2010; Gulsen *et al.*, 2010). Il s'agit en fait de marqueurs peu intéressants aujourd'hui du fait de leur faible reproductivité et de leur caractère dominant. Ils restent malgré tout utilisés sur agrumes par certaines équipes (Pessina *et al.*, 2011; Savita *et al.*, 2012; Şahin-Çevik et Moore, 2012; Kim *et al.*, 2013; Malik *et al.*, 2013; Tripolitsiotis *et al.*, 2013; Naz *et al.*, 2014)

### 7.1.7. RFLP (*Restriction Fragment Length Polymorphism*)

Cette méthode met principalement en évidence des polymorphismes affectant des sites de restriction. Après digestion de l'ADN total, et migration par électrophorèse, les polymorphismes de régions spécifiques du génome sont révélés par hybridation avec des sondes radioactives ou fluorescentes, Il s'agit de marqueurs intéressants du fait de leur codominance, de la possibilité de les développer sur l'ensemble du génome et de leur important polymorphisme. Pourtant, aujourd'hui l'utilisation de cette technique est pratiquement abandonnée compte tenu de sa lourdeur comparativement à d'autres méthodes. Ces marqueurs furent utilisés sur agrumes pour établir les premières cartes génétiques (Durham *et al.*, 1992; Jarrell *et al.*, 1992; Cai *et al.*, 1994; Liou *et al.*, 1996; Luro *et al.*, 1996; Kijas *et al.*, 1997; de Simone *et al.*, 1998; Cristofani *et al.*, 1999; Garcia *et al.*, 1999; Ling *et al.*, 1999; Roose *et al.*, 2000; Sankar et Moore, 2001). Associés avec des RAPD, ils ont été également utilisés pour différencier des plantules zygotiques et nucellaires ainsi que pour évaluer la variabilité génétique intraspécifique des orangers et des mandariniers (Luro *et al.*, 1995). L'analyse de 88 représentants du genre *Citrus* ainsi que d'hybrides et apparentés fut réalisée à l'aide de ces marqueurs en 1998 (Federici *et al.*, 1998).

### 7.1.8. SNP (*Single Nucleotide Polymorphisms*)

On parle de polymorphisme nucléotidique pour la variation allélique d'une seule base entre deux haplotypes d'un même individu ou entre chromosomes homologues d'une population d'individus. Ce sont les variations de structure d'ADN les plus fréquentes, présentes indifféremment en zones codantes et non codantes (Brookes, 1999). Ces polymorphismes permettent de développer des marqueurs codominants nécessitant une connaissance préalable de la séquence étudiée. L'abondance de ces variations et leur distribution sur l'ensemble du génome font de ces marqueurs des outils très puissants (Taylor, 2013). Par ailleurs, la probabilité de convergence vers une forme identique à partir d'événements indépendants est très faible, limitant les risques d'homoplasie. La recherche de SNP à grande échelle a été entreprise sur le génome de nombreuses espèces dont l'homme (Garg *et al.*, 1999; Sachidanandam *et al.*, 2001), *Arabidopsis thaliana* (Jander *et al.*, 2002) et la drosophile *Drosophila melanogaster* (Hoskins *et al.*, 2001) mais également de plantes cultivées telles que l'orge (Rostoks *et al.*, 2005), le maïs (Ching *et al.*, 2002), le riz ((Shen *et al.*, 2004; McNally *et al.*, 2006), le soja (Zhu *et al.*, 2003), le blé (Ablett *et al.*, 2006; Ravel *et al.*, 2006), la pomme de terre (Rickert *et al.*, 2003), la pomme (Han *et al.*, 2009) ainsi que pour des plantes polyploïdes telles que la canne à sucre (Bundock *et al.*, 2009) ou l'arachide (Bertioli *et al.*, 2014). Les techniques de séquençage se démocratisant, les SNP sont de plus en plus utilisés pour le développement de cartes génétiques, l'identification de variétés, la détection d'associations entre génotypes et phénotypes ou la sélection assistée par marqueurs (Morales *et al.*, 2004; Xing *et al.*, 2005; Lijavetzky *et al.*, 2007; Duval *et al.*, 2014;

Fu *et al.*, 2014; Kobayashi *et al.*, 2014; Liu *et al.*, 2014a; Liu *et al.*, 2014b; Manrique-Carpintero *et al.*, 2014; Muys *et al.*, 2014).

Sur agrumes, comme pour les autres cultures, les SNP sont de plus en plus utilisés pour divers objectifs dont :

- l'étude des niveaux d'hétérozygotie et de diversité (Novelli *et al.*, 2004; Jiang Dong *et al.*, 2010; Ollitrault *et al.*, 2012b; Bernardi *et al.*, 2013; Chen et Gmitter, 2013; Cuenca *et al.*, 2013a; Distefano *et al.*, 2013; Fujii *et al.*, 2013; Garcia-Lor *et al.*, 2013a et b),
- l'association de marqueurs et l'identification de gènes de résistance (Cuenca *et al.*, 2013b),
- la discrimination de zygotes dans des semis (Zhu *et al.*, 2013),
- de la cartographie (Ollitrault *et al.*, 2012a),
- de la génomique fonctionnelle (Ye JunLi *et al.*, 2010; Kaçar *et al.*, 2014).

Les SNP identifiés sur agrumes à partir de données de séquençage peuvent faire l'objet de génotypage haut débit à l'aide de puces à ADN comme le système Goldengate d'Illumina (Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b) ou pour des études plus ciblées par PCR compétitive allèle-spécifiques [KBioscience Competitive Allele-Specific Polymerase chain reaction (KASPar) assay (Cuenca *et al.*, 2013a; Garcia-Lor *et al.*, 2013a)].

Enfin, en l'absence de connaissance préalable sur les séquences, il est également possible de révéler des polymorphismes SNP par la méthode SSCP (*Single Strand Conformation Polymorphisms*) basée sur les différences de structure tertiaire de l'ADN provoquées par le changement de nucléotide affectant la migration lors de l'électrophorèse (Olivares-Fuster *et al.*, 2007; Luro *et al.*, 2011; Simsek *et al.*, 2011).

### 7.1.9. SSR (*Simple Sequence Repeats*)

Les SSR ou microsatellites, sont de courtes séquences d'ADN qui se répètent en tandem, un nombre variable de fois. Ces marqueurs, utilisés pour la première fois sur agrumes par Kijas *et al.* (Kijas *et al.*, 1995) puis repris pour des études génétiques plus larges (Gulsen et Roose, 2001a; Luro *et al.*, 2001; Barkley *et al.*, 2006; Ollitrault *et al.*, 2011; Uzun *et al.*, 2011b) sont considérés comme des outils puissants, car ils sont très polymorphes, codominants, généralement *locus-spécifiques* et dispersés de façon aléatoire dans le génome de la plante. Pourtant, Barkley *et al.* (Barkley *et al.*, 2009) ont démontré qu'en raison de phénomènes d'homoplasie (des formes alléliques peuvent être identiques alors qu'elles sont issues d'événements génétiques indépendants), l'utilité de ces marqueurs pourrait présenter des limites pour comprendre l'origine phylogénétique des fragments d'ADN. De plus, ces marqueurs, qu'ils soient génomiques ou inclus dans des EST, ne permettent souvent pas d'atteindre le niveau de spécificité recherché. En effet, les études de génotypage attestent de la présence très fréquente d'allèles communs au moins à deux taxons ancestraux (mandariniers et pamplemoussiers) (Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2012a; Garcia-Lor *et al.*, 2013b).

Aujourd'hui plusieurs centaines de marqueurs SSR ont été publiés (Novelli *et al.*, 2000; Roose *et al.*, 2000; Corazza-Nunes *et al.*, 2002; Liu *et al.*, 2002; Ahmad *et al.*, 2003; Golein *et al.*, 2005; Jiang Dong *et al.*, 2006; Novelli *et al.*, 2006; Caruso *et al.*, 2008; Chen *et al.*, 2008; Luro *et al.*, 2008; Nematollahi *et al.*, 2009; Nematollahi *et al.*, 2009; Ghorabaie *et al.*,



2010; Ghorabaie *et al.*, 2010; Gulsen *et al.*, 2010; Ollitrault *et al.*, 2010; Amar *et al.*, 2011; Biswas *et al.*, 2011; Cristofani-Yaly *et al.*, 2011; El-Mouei *et al.*, 2011; Kamiri *et al.*, 2011; Ollitrault *et al.*, 2011; Singh *et al.*, 2011; Uzun *et al.*, 2011b; Biswas *et al.*, 2012; Garcia-Lor *et al.*, 2012; Golein *et al.*, 2012; Ollitrault *et al.*, 2012a; Polat *et al.*, 2012; Snoussi *et al.*, 2012; Chai *et al.*, 2013; Kacar *et al.*, 2013; Liu *et al.*, 2013; Yildiz *et al.*, 2013; Hou *et al.*, 2014) et sont disponibles pour les diverses recherches conduites en génétique des agrumes.

## 7.2. Marqueurs cytoplasmiques

Des études de phylogénie maternelle, dont certaines déjà citées, ont pu être réalisées sur agrumes à l'aide de marqueurs chloroplastiques RFLP : (Gulsen et Roose, 2001b; Abkenar *et al.*, 2004), Indels (de Araújo *et al.*, 2003), CAPS (Nicolosi *et al.*, 2000; Lotfy *et al.*, 2003; Yamamoto *et al.*, 2013), SSR chloroplastiques (cpSSR) (Cheng *et al.*, 2005; Li *et al.*, 2006; Deng *et al.*, 2007) et comparaison de séquences de gènes chloroplastiques (Jung *et al.*, 2005; Bayer *et al.*, 2009; Lu *et al.*, 2011; Wali *et al.*, 2013). Des marqueurs mitochondriaux ont également été développés (Yamamoto *et al.*, 1993; Froelicher *et al.*, 2011; Snoussi *et al.*, 2012). Des caractérisations cytoplasmiques ont aussi été conduites à l'aide de SSCP, *Single Strand Conformation Polymorphisms* (Cheng *et al.*, 2005; Olivares-Fuster *et al.*, 2007). La séquence complète du génome chloroplastique de l'oranger est disponible depuis 2006 (Bausher *et al.*, 2006).

Tableau 8 : Récapitulatif bibliographique des études utilisant des marqueurs moléculaires sur agrumes

Marqueur	Analyse de la diversité	Discrimination zygote / nucléaire	Phylogénie	Phylogénie maternelle (Marqueurs cytoplasmiques)	Cartographie	Association marqueur / phénotype
AFLP	Dominant				Carlos de Oliveira <i>et al.</i> , 2007	Al-Sadi <i>et al.</i> , 2012; Zhou <i>et al.</i> , 2013; Bernardi <i>et al.</i> , 2010; Fang <i>et al.</i> , 2010; Kepiro and Roose, 2010; Li <i>et al.</i> , 2010; Distefano <i>et al.</i> , 2011; Licciardello <i>et al.</i> , 2011; Xiao <i>et al.</i> , 2011; Kacar <i>et al.</i> , 2012
IRAP	Dominant				Bernet <i>et al.</i> , 2010	
ISSR	Dominant					
RAPD	Dominant					
INDEL	Codominant					

CAPS (PCR- RFLP)	Codominant	Omura <i>et al.</i> , 2000; Jiang Dong <i>et al.</i> , 2010; Amar <i>et al.</i> , 2011; Yamamoto <i>et al.</i> , 2011; Shimada <i>et al.</i> , 2014			Nicolosi <i>et al.</i> , 2000 ; Lotfy <i>et al.</i> , 2003 ; Yamamoto <i>et al.</i> , 2013 ; Gulsen and Roose, 2001b ; Abkenar <i>et al.</i> , 2004;			
RFLP	Codominant	Federici <i>et al.</i> , 1998; Ramadugu <i>et al.</i> , 2013	Luro <i>et al.</i> , 1995		Yamamoto <i>et al.</i> , 1993		Durham <i>et al.</i> , 1992; Jarrell <i>et al.</i> , 1992; Cai <i>et al.</i> , 1994; Liou <i>et al.</i> , 1996; Luro <i>et al.</i> , 1996; Kijas <i>et al.</i> , 1997; de Simone <i>et al.</i> , 1998; Cristofani <i>et al.</i> , 1999; Garcia <i>et al.</i> , 1999; Ling <i>et al.</i> , 1999; Roose <i>et al.</i> , 2000; Sankar and Moore, 2001	
SNP	Codominant	Novelli <i>et al.</i> , 2004; Jiang Dong <i>et al.</i> , 2010; Ollitrault <i>et al.</i> , 2012b; Bernardi <i>et al.</i> , 2013; Chen and Gmitter, 2013; Cuenca <i>et al.</i> , 2013a; Distefano <i>et al.</i> , 2013; Fujii <i>et al.</i> , 2013; Garcia- Lor <i>et al.</i> , 2013a et b; Olivares-Fuster <i>et al.</i> , 2007; Luro <i>et al.</i> , 2011; Simsek <i>et al.</i> , 2011	Zhu <i>et al.</i> , 2013		Bausher <i>et al.</i> , 2006; Lu <i>et al.</i> , 2011; Wali <i>et al.</i> , 2013	Ollitrault <i>et al.</i> , 2012a et 2011	Cuenca <i>et al.</i> , 2013b; Ye JunLi <i>et al.</i> , 2010; Kaçar <i>et al.</i> , 2014	
SSR	Codominant	Nematollahi <i>et al.</i> , 2009; Uzun <i>et al.</i> , 2011b; Kijas <i>et al.</i> , 1995; Barkley <i>et al.</i> , 2006; Garcia-Lor <i>et al.</i> , 2013b; Novelli <i>et al.</i> , 2000; Roose <i>et al.</i> , 2000; Corazza-Nunes <i>et al.</i> , 2002; Liu <i>et al.</i> , 2002; Ahmad <i>et al.</i> , 2003; Golein <i>et al.</i> , 2005; Jiang Dong <i>et al.</i> , 2006; Novelli <i>et al.</i> , 2006; Caruso <i>et al.</i> , 2008; Chen <i>et al.</i> , 2008; Luro <i>et al.</i> , 2008; Ghorabaie <i>et al.</i> , 2010; Gulsen <i>et al.</i> , 2010; Ollitrault <i>et al.</i> , 2010; Amar <i>et al.</i> , 2011; Biswas <i>et al.</i> , 2011; Cristofani-Yaly <i>et al.</i> , 2011; El-Mouei <i>et al.</i> , 2011; Kamiri <i>et al.</i> , 2011; Ollitrault <i>et al.</i> , 2011; Singh <i>et al.</i> , 2011; Biswas <i>et al.</i> , 2012; Garcia-Lor <i>et al.</i> , 2012; Golein <i>et al.</i> , 2012; Ollitrault <i>et al.</i> , 2012a; Polat <i>et al.</i> , 2012; Snoussi <i>et al.</i> , 2012; Chai <i>et al.</i> , 2013; Kacar <i>et al.</i> , 2013; Liu <i>et al.</i> , 2013; Yildiz <i>et al.</i> , 2013; Hou <i>et al.</i> , 2014			Gulsen and Roose, 2001a; Luro <i>et al.</i> 2001	Snoussi <i>et al.</i> , 2012; Cheng 2005	Kijas <i>et al.</i> , 1997; Ollitrault <i>et al.</i> , 2011 et 2012a	

### 7.3. Données génétiques et génomiques disponibles

Les premières données de génomiques obtenues sur agrumes sont compilées dans différentes synthèses comme celles de Gmitter *et al.* 2007, Talon et Gmitter en 2008 ou Tadeo *et al.* 2008. Il existe aujourd'hui différentes bases de données regroupant les EST agrumes, dont celles de l'*International Citrus Genome Consortium* (2014), qui regroupent plus de 500 000 EST enregistrées dans la base de données GenBank (Benson *et al.*, 2014) avec plus d'un tiers d'EST provenant de l'oranger (37,7 %), un tiers partagé entre le clémentinier et le *Poncirus* (31,9 %), et le reste (30,4 %) provenant de 22 autres agrumes (tableau 9) (Forment *et al.*, 2005; Terol *et al.*, 2008; Shimizu *et al.*, 2009; Delseny *et al.*, 2010).

Tableau 9 : EST agrumes enregistrées dans la base NCBI (2014)

Organismes	Nombre d'EST	%
<i>Citrus sinensis</i> (Valencia orange)	214 598	37.7 %
<i>Citrus clementina</i>	118 365	20.8 %
<i>Poncirus trifoliata</i>	62 941	11.1 %
<i>Citrus reticulata</i>	56 041	9.9 %
<i>Citrus unshiu</i>	19 072	3.4 %
<i>Citrus aurantium</i>	14 584	2.6 %
<i>Citrus limonia</i>	11 173	2.0 %
<i>Citrus latifolia</i>	8 756	1.5 %
<i>Citrus aurantiifolia</i>	8 219	1.4 %
<i>Citrus limettioides</i>	8 188	1.4 %
<i>Citrus x paradisi</i>	8 039	1.4 %
<i>Citrus x paradisi x Poncirus trifoliata</i>	7 954	1.4 %
<i>Citrus reticulata x Citrus temple</i>	5 823	1.0 %
<i>Citrus reshni</i>	5 768	1.0 %
<i>Citrus sunki</i>	5 216	0.9 %
<i>Fortunella margarita</i>	2 924	0.5 %
<i>Citrus macrophylla</i>	1 929	0.3 %
<i>Citrus clementina x Citrus tangerina</i>	1 843	0.3 %
<i>Citrus sinensis x Poncirus trifoliata</i>	1 837	0.3 %
<i>Citrus limon</i>	1 505	0.3 %
<i>Citrus medica</i>	1 115	0.2 %
<i>Citrus jambhiri</i>	1 017	0.2 %
<i>Citrus nobilis x Citrus kinokuni</i>	645	0.1 %
<i>Citrus tamurana</i>	358	0.06 %
<i>Citrus natsudaikai</i>	202	0.04 %
<i>Citrus sinensis x Citrus reticulata</i>	160	0.03 %
<i>Citrus hassaku</i>	154	0.03 %
<i>Citrus trifoliata</i>	139	0.02 %
<i>Citrus clementina x Citrus reticulata</i>	74	0.01 %
<i>Citrus aurantiifolia/Candidatus Phytoplasma aurantifolia mixed EST library</i>	55	0.01 %
<i>Citrus maxima</i>	44	0.01 %
Total	568 738	

[http://www.ncbi.nlm.nih.gov/genbank/dbest/dbest\\_summary/](http://www.ncbi.nlm.nih.gov/genbank/dbest/dbest_summary/)

Les données du séquençage du génome complet d'un haploïde de clémentinier séquencé par le consortium international ICGC (*International Citrus Genome Consortium*) est disponible sur Phytozome (Department of Energy's Joint Genome Institute and Center for Integrative Genomics, 2014) ainsi que sur la base de donnée du consortium (ICGC, 2014). Ces données sont aujourd'hui utilisées comme référence pour la majorité des séquençages de variétés d'agrumes, qui sont généralement mis en correspondance (ancrés) sur cet assemblage original. Ainsi, les données de reséquençage complet d'autres variétés d'agrumes sont aujourd'hui libérées et l'analyse de l'origine de ces espèces, à partir de ces

données, a été publié par l'ICGC en 2014 (Wu *et al.*, 2014). D'autres équipes ont réalisé des séquençages complets de génomes d'agrumes dont le premier publié fut celui de l'orange (*Citrus sinensis*) en 2013 (disponible en ligne dès 2012 ; Xu *et al.*, 2013), suivi de différentes mandarines (Satsuma et Ponkan) et hybrides (Shimizu *et al.*, 2012). Enfin, le consortium espagnol CITRUSEQ / CITRUSGENN (<http://www.citrusgenn.es/>) est en possession de données de séquençage complet de plusieurs variétés d'agrumes dont l'ancrage a également été réalisé sur l'assemblage de l'haploïde de clémentinier de l'ICGC (Terol *et al.*, 2012). Ce consortium CITRUSEQ/CITRUSGENN est un projet de collaboration entre deux instituts publics (*l'Instituto Valenciano de Investigaciones Agrarias* [IVIA] et *l'Instituto de Recursos Naturales y Agrobiología de Sevilla* [IRNASE]), deux fondations (*la Fundación de la Comunidad Valenciana Centro de Investigación Príncipe Felipe* [CIPF] et *la Fundación Ruralcaja Valencia*) et 5 compagnies privées (*Eurosemillas*, *Source Citrus Geensis* SL, *GMC Variedades Vegetales* A.I.E., *Anecoop*, et *ICCSA*). L'IVIA est porteur du projet du consortium espagnol *Genómica de Cítricos* (SCGC), (intégré au consortium international de génomique des agrumes (ICGC, 2014), organisme dont l'objectif est de favoriser la recherche génétique sur les agrumes et de stimuler la collaboration entre les différents programmes de recherches. Les différents acteurs du consortium décidèrent de s'unir en 2008 pour développer des outils et des méthodologies génétiques permettant d'accélérer les programmes d'amélioration génétique des agrumes. Ces outils étant indispensables à la rationalisation de la citriculture espagnole, principale motivation du consortium CITRUSEQ / CITRUSGENN. Le consortium a permis de séquencer le génome de 250 espèces, variétés et porte-greffe d'agrumes et a généré une base de données regroupant toute l'information nécessaire à l'identification des variants structurels tels que les SNP, indels et recombinaisons méiotiques des génotypes séquencés. Cette base de données est sous contrôle strict du consortium et les données ne peuvent être utilisées pour aucun autre objectif que ceux définis et approuvés par le consortium. C'est dans ce contexte que nous avons eu accès aux données qui ont permis la rédaction du Chapitre IV de cette thèse.

Il existe d'autres bases de données génomiques comme la *Citrus Genome Database* (USDA, 2014) qui regroupent les données de cartographie, libraires BAC, EST et reprend les données de séquençage complet de l'orange (diploïde) de d'un haploïde de clémentine produite par l'ICGC, ou celle de *Huazhong Agricultural University* (Wang *et al.*, 2014), donnant accès au séquençage complet de l'orange (*C. sinensis*) publié par Xu *et al.* en 2013 (Huazhong Agricultural University, 2014).

Des puces à ADN ou ARN de haute densité ont également été développées pour des études d'expression ou de génotypage (Shimada *et al.*, 2005; Terol *et al.*, 2007; Martinez-Godoy *et al.*, 2008; Khalaf *et al.*, 2011; Shimizu *et al.*, 2011; Ollitrault *et al.*, 2012a). Il existe également des banques BAC génomique de clémentinier (bacterial artificial chromosome) (Terol *et al.*, 2008). Une carte physique ainsi que des cartes génétiques de l'oranger, du clémentinier et du pamplemoussier sont également disponibles (Ollitrault *et al.*, 2012a; ICGC, 2014).

## 7.4. Les promesses des NGS

Alors qu'au début des années 2000 le séquençage en méthode Sanger (Sanger et Coulson, 1975; Sanger *et al.*, 1977) coûtait environ 10 000 dollars pour un million de paires de bases et que l'on ne pouvait séquençer que quelques centaines de milliers de paires de base en un seul coup (réalisation d'un processus complet, *run*), en 2010 on parlait déjà de séquençage de nouvelle génération produisant des millions de séquences en un seul coup, pour un coût estimé à 1 dollar le séquençage d'un million de paires de base (Nature, 2010). Cette évolution a été considérée comme une véritable *révolution* avec l'apparition de technologies qualifiées de *Next generation sequencing* ou nouvelles générations de séquençage à haut débit (*high throughput sequencing*). Cette deuxième génération de séquenceurs à haut débits (454, Illumina, Applied Biosystems) produit des lectures de longueurs limitées (35 pb à 200 pb pour Illumina contre 300 pb à 600 pb pour 454) mais permet une multitude de lectures en parallèle. Les réactions de séquençage peuvent être relativement longues mais elles génèrent en une seule fois une quantité très importante de données (tableau 10).

Après les séquenceurs de deuxième génération, nécessitant une phase d'amplification de l'ADN par PCR, on parle aujourd'hui de séquenceur de « nouvelle-nouvelle » génération ou de troisième génération ou TGS (*third-generation sequencing*) décryptant directement une seule molécule d'ADN (*Oxford Nanopore Technologies*) (Thudi *et al.*, 2012).

### 7.4.1. Séquençage à haut débit, de deuxième génération

#### 7.4.1.1. Pyroséquençage (454 Roche)

Commercialisé pour la première fois en 2004, le séquenceur 454 de *Life Science* (Roche) est basé sur la technique du pyroséquençage (Margulies *et al.*, 2005). Le séquençage se fait par détection du signal lumineux, provoqué par la libération de pyrophosphate à chaque incorporation d'une nouvelle base, à l'aide d'une caméra DTC (Dispositif à Transfert de Charge ou CCD : *Charge-Coupled Device*). Cela permet d'obtenir jusqu'à 1 million de séquences par processus complet (*run*) pouvant aller jusqu'à 400 bases. Cette technique a notamment été utilisée avec succès pour la détection de SNP (Barbazuk *et al.*, 2007; Maughan *et al.*, 2009; Fu et Peterson, 2011; Pootakham *et al.*, 2011; Mercati *et al.*, 2013; Pootakham *et al.*, 2013; Fu *et al.*, 2014). Roche a développé un système d'adaptateurs (MID : *Multiplex Identifier*) permettant le multiplexage (454 Life Sciences Corp., 2009). Cette technique permet de séquençer jusqu'à 600 millions de bases en un cycle de 10 heures. Même si son prix est relativement intéressant, ramené à la base séquencée, il reste élevé pour un cycle complet, ce qui limite son utilisation pour un nombre réduit d'échantillons. Son principal défaut reste malgré tout son taux d'erreur élevé au moment de séquençer des zones riches en homopolymères (répétition de la même base).

#### 7.4.1.2. Séquençage à l'aide de terminateurs réversibles (Illumina)

Cette technique fait également partie des toutes premières plateformes mises sur le marché. Elle permet d'amplifier des échantillons d'ADN fixés à un support solide, support sur lequel l'ADN est directement séquencé. Le séquençage se fait par ajout d'un mélange contenant toutes les bases, chacune associée à un fluorophore différent. Une lecture laser

permet de détecter toutes les positions incorporées. Cette méthode permet de séquencer en parallèle plus de 3 milliards de séquences de 100 bases de long (Hillier *et al.*, 2008; Van Tassell *et al.*, 2008). Les erreurs, du fait de la lecture base à base, sont provoquées par la fréquente substitution d'une base par une autre.

#### 7.4.1.3. Séquençage par ligature (*ligation*) (Applied Biosystems)

L'amplification des échantillons d'ADN se fait, comme pour le pyroséquençage, par emPCR, PCR en émulsion, mais les séquences amplifiées sont ensuite fixées sur un support solide. Le séquençage se fait par cycles de ligature (*ligation*) et de clivage. Cette technique permet de corriger d'éventuelles erreurs d'incorporation. Un milliard et demi de séquences de 75 bases de long peuvent être lues par cycle (Shendure *et al.*, 2005). Le système de correction ainsi que l'utilisation de ligase fiabilisent considérablement cette technique.

#### 7.4.1.4. Séquençage par semi-conducteur (*Ion Torrent*)

Cette technique ne nécessite aucun fluorophore associé à l'ADN. Elle est basée sur la variation de pH que provoque l'ajout d'un nucléotide à la séquence. Après amplification de l'ADN à séquencer sur une perle de capture, ces perles sont déposées individuellement dans des puits qui sont successivement remplis d'un mélange ne contenant qu'un seul type de nucléotides. A chaque fois qu'une base correspond à la base complémentaire attendue, le pH de la solution change. Chaque puits correspond à un pH-mètre miniature. En plus d'être facilement miniaturisable, cette technique qui génère des séquences de 100 à 500 pb est relativement peu chère (matériel de départ et réactifs) mais semble avoir un taux d'erreur encore très élevé (Mascher *et al.*, 2013). Il est annoncé en 2014 une puce « PIII » pour le séquenceur Ion Proton, permettant de générer environ 64Gb.

Tableau 10 : récapitulatif des capacités des principaux séquenceurs de deuxième génération des trois plus gros fournisseurs du marché d'après Blervaque (Blervaque, 2013)

Société	Roche		Illumina				Life Technologies						
Plateforme	GS Junior	454	Mi Seq	GAllx	HiSeq 1000/1500	HiSeq 2000/2500	ion Torrent PGM			Ion Proton		SOLiD 5500xl	SOLiD 5500
Technologie	Titanium	FLX+					Chip 314	Chip 316	Chip 318	Chip PI	Chip PII		
Méthode d'amplification	PCR en émulsion		PCR sur plaque				PCR en émulsion						
Méthode de séquençage	Synthèse (Pyroséquençage)		Synthèse				Ligation						
Capacité séquençage /run	35 Mb	700Mb	8 Gb	95 Gb	300 Gb	600 Gb	100 Mb	1 Gb	2 Gb	10 Gb	32 Gb	95 Gb	48 Gb
Taille moyenne des reads	400	700	2x300	2x150	2x100/150	2x100/150	400	400	400	200	100	2x60	2x60
Coût en kUSD machine + annexes (2013)	125	550	125	300	590	690	50+20	50+20	50+20	149	149	600	350
Coût en kUSD / run (2013)	1	6	1	17	11	23	350	550	750	1	1	10	5
Durée de séquençage / run	10 h	23 h	27 h	14 j	8,5 j	11 j	4 h	5 h	7 h	4 h	4 h	6 j	6 j

### 7.4.2. Séquençage de troisième génération (TGS)

Le développement de nouvelles technologies de séquençage a généralement eu pour objectif de permettre de diminuer les coûts et d'augmenter le rendement et la qualité des réactions de séquençage. Contrairement aux technologies à haut débit dites de deuxième génération, qui requièrent une étape d'amplification des molécules d'ADN avant le séquençage, celles dites de troisième génération permettraient le décryptage direct d'une seule molécule d'ADN. C'est ce que propose l'entreprise Oxford Nanopore Technologies (ONT). La technologie annoncée repose sur l'utilisation de nanopores, des protéines transmembranaires modifiées, qui ont la capacité d'émettre un signal électrique spécifique lorsqu'elles se trouvent au contact de certaines molécules (Yuhui *et al.*, 2013). De l'ADN, non ou peu préparé, est censé émettre un signal électrique spécifique au moment de traverser ces nanopores imbriquées dans des membranes hybrides polymères / lipides. Le séquenceur doit interpréter ce signal électrique pour savoir quel nucléotide en est responsable. La technique ne nécessite que peu à pas de préparation d'ADN et ne connaît virtuellement pas de limite de longueur de lecture. Le taux d'erreur annoncé est encore très élevé (4 %) mais rapidement améliorable selon l'entreprise (Hayden, 2012). Parmi les séquenceurs annoncés le plus surprenant est le MinION, une unité de séquençage par nanopores à usage unique intégrés dans une clef USB. Cet outil, directement branché sur un port USB dont il tirera son alimentation, permettrait de lire un échantillon d'ADN en milieu aqueux et de directement transférer les séquences lues sur l'ordinateur (Oxford Nanopore Technologies, 2014) ; le tout pour moins de 1 000 dollars. D'autres voies sont à l'étude telles que la combinaison de la détection optique et la technologie multi-pore (McNally *et al.*, 2010), l'imagerie directe de l'ADN par microscopie électronique (Gentile *et al.*, 2012) ou encore le séquençage à l'aide de transistor (*Transistor-mediated DNA sequencing* ; (Polonsky *et al.*, 2007; Luan *et al.*, 2010). Pourtant, aujourd'hui, après de nombreux effets d'annonces, l'une des seules technologies dite de troisième génération réellement utilisée est celle de Pacific Bioscience, le Pacbio RS basée sur la technique de séquençage *Single Molecule Real Time sequencing* (SMRT). Avec 200 à 300 Mb délivrés par SMRT-cell, le séquençage de gros génomes par cette technique demande encore un investissement important.

### 7.5. Exploiter et intégrer les données des séquençages

Décrypter l'information du génome est un enjeu majeur pour la recherche scientifique. Cela permet de mieux comprendre et d'appréhender le fonctionnement des êtres vivants. Chez l'humain comme chez les plantes cela permet de comprendre l'origine de certaines maladies et par conséquent les soigner. C'est aussi un outil qui a permis l'essor de la phylogénie moderne. Quelle que soit la technique de séquençage utilisée elle génère une multitude de données. Ces données ne sont utiles que si elles sont analysées. Tout commence par une étape d'évaluation de la qualité des données, une validation après « nettoyage » et, s'il s'agit de séquençages de génome complet, soit un assemblage *de novo* (ce qui a été le cas pour le premier génome séquencé de l'haploïde de clémentine) soit un ancrage sur un génome de référence (pour les agrumes, la première référence disponible fut l'assemblage *de novo* du génome de l'haploïde de clémentine de l'ICGC (2014)). Enfin, une fois les séquences obtenues assemblées ou ancrées, il faut être capable d'identifier et de caractériser les variations structurelles (SNP, indels, inversions, translocations, transpositions...). Toutes ces actions requièrent des outils informatiques puissants et des compétences en bio-informatique spécifiques. Les données brutes ainsi que les données

générées par l'analyse doivent être stockées et accessibles pour permettre de réaliser toutes les études nécessaires aux différentes stratégies de caractérisation, d'étude de phylogénie, et de compréhension des mécanismes d'expression par exemple. Un grand nombre de navigateurs de génomes et de pages Internet ont été élaborés spécifiquement pour mettre à disposition le maximum de données ainsi que les outils qui permettent de fouiller et travailler sur ces données. Ces quantités, ou volumes, de données générées ne sont généralement pas gérables sous un environnement informatique classique. Le nuage informatique ou l'infonuagique (termes français préconisés par le grand dictionnaire terminologique de l'office québécois de la langue française (Office québécois de la langue française, 2014) pour remplacer les termes anglais de *cloud* et *cloud-computing*) semble être une solution intéressante pour les utilisateurs finaux de ces données qui n'ont plus forcément besoin de connaître le site physique de stockage des données ni la configuration des systèmes qui fournissent ces données. Le nuage offre donc d'incroyables possibilités d'accès à de grandes quantités de données, à de grandes capacités de stockage de données, à des logiciels mais également à de puissantes possibilités de calcul (Stein, 2010; Thudi *et al.*, 2012). Il existe aujourd'hui différents outils et plateformes permettant d'aider la communauté scientifique à faire face à cette arrivée massive de données créées par la démocratisation du séquençage :

#### 7.5.1. Biopython

Il s'agit d'un projet gratuit et open source (Cock *et al.*, 2009) regroupant un ensemble de programmes et de procédures en ligne de commande sous Linux pour le traitement et l'analyse de données biologiques en langage de programmation Python permettant l'accès à de nombreux formats de fichier utilisés par les principales bases de données (Blast out, Fasta, GenBank, UniGene...)

#### 7.5.2. South Green

Il s'agit d'une plateforme regroupant plusieurs instituts et organismes dont le Cirad, l'Inra, l'IRD, SupAgro et qui développe et donne accès à des programmes et propose des solutions informatiques dédiées à la génétique et la génomique des plantes tropicales et méditerranéenne (SouthGreen, 2014).

#### 7.5.3. France Génomique

Créée grâce à un financement « Investissements d'Avenir », France Génomique est une infrastructure née de la volonté d'optimiser et de renforcer les capacités françaises dans le domaine de la génomique à haut débit et de la bioinformatique associée (France Génomique, 2014). Les objectifs sont d'intégrer à l'échelon national des capacités importantes d'analyse génomique (séquençage/génotypage et traitements bio-informatique) et de rassembler la masse critique et les expertises nécessaires pour renforcer la compétitivité de la communauté nationale (recherche publique ou industrielle). La gouvernance de France Génomique devrait à terme être renforcée grâce à la création d'une UMS (Unité Mixte de Service) qui regroupera les quatre principaux organismes de recherche publics impliqués : CEA, CNRS, INRA et INSERM.



## 8. Objectifs et activités de thèse

Comme nous venons de le voir, la connaissance des structures génétiques et des mécanismes impliqués dans l'évolution des espèces est un préalable, non seulement à la rationalisation de la gestion des collections de ressources génétiques, mais également à la maîtrise de leur exploitation dans des schémas d'amélioration variétale. Les agrumes, première production fruitière au monde avec plus de 130 millions de tonnes<sup>8</sup>, sont des arbres produisant des fruits de tailles, de formes et de couleurs très variables, caractérisés par une surface de peau (le zeste) riche en glandes à huiles essentielles et une organisation interne originale appelée *hesperidie* (présence de quartiers, comprenant des pépins et de nombreux poils succulents ou sacs à jus<sup>9</sup>). Les agrumes dits vrais regroupent plusieurs espèces appartenant à 6 genres botaniques inter-fertiles (*Citrus*, *Fortunella*, *Poncirus*, *Clymenia*, *Microcitrus* et *Eremocitrus*<sup>10</sup>). Le genre *Citrus* regroupe l'ensemble des variétés de grande consommation, alors que le genre *Poncirus* est exploité pour la production de porte-greffe. Les agrumes sont généralement diploïdes et ont 9 paires de chromosomes pour une taille du génome estimé à 367 Mb (Ollitrault *et al.*, 2003).

Toute la communauté scientifique travaillant sur les agrumes adhère à l'hypothèse selon laquelle, il y aurait 4 taxons de base (*C. reticulata*, mandariniers ; *C. maxima*, pamplemoussiers, *C. medica*, cédratiers et *C. micrantha*, papédas) à l'origine des agrumes cultivés du genre *Citrus* (Nicolosi *et al.*, 2000; Nicolosi, 2007; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b; Ramadugu *et al.*, 2013). Comme cela a pu être démontré pour les caroténoïdes ainsi que pour l'acidité, on sait qu'une grande partie de la diversité phénotypique et de sa structuration au sein du genre *Citrus* est liée à la différenciation entre les 4 taxons ancestraux avant la genèse des espèces secondaires (Ollitrault *et al.*, 2003; Fanciullino *et al.*, 2007).

Par différentes combinaisons sexuées, ces taxons de base auraient généré les hybrides initiaux fondateurs des espèces dites secondaires (*C. sinensis*, orangers ; *C. paradisi*, pomelos ; *C. limon*, citronniers ; *C. aurantium*, bigaradiers ; *C. clementina*, clémentiniers ; figure 16). Les limettiers et citronniers auraient des origines multiples. Ainsi la lime Mexicaine pourrait résulter d'une hybridation entre un représentant du sous-genre papéda (*Citrus micrantha*) et un cédratier (*C. medica*) alors que les citronniers largement cultivés dans le bassin méditerranéen seraient issus d'une hybridation entre *C. aurantium* (une bigaradiers) et *C. medica* (Nicolosi *et al.*, 2000). L'apomixie facultative de toutes les espèces secondaires a permis de fixer et multiplier de façon conforme ces structures fortement hétérozygotes et a conduit les taxonomistes à considérer ces groupes comme des espèces (Scora, 1975; Barrett et Rhodes, 1976). De fait, les génomes des espèces secondaires et des variétés modernes peuvent être considérés comme des mosaïques interspécifiques de grands fragments des génomes des taxons de base.

---

<sup>8</sup> cf. La production ; point 4.1.2.

<sup>9</sup> cf. Ordre des Sapindales, famille des Rutaceae ; point 2.1.

<sup>10</sup> cf. Les agrumes vrais : six genres botaniques ; point 2.2.

De nombreuses études génétiques, et notamment de marquage moléculaire, ont contribué à identifier les mécanismes de diversification au cours de l'évolution (Herrero *et al.*, 1996; Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Bretó *et al.*, 2001; Luro *et al.*, 2001; Barkley *et al.*, 2006; Nicolosi, 2007; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b). Ainsi, la sexualité a été le moteur principal de la diversification intraspécifique des taxons ancestraux, tandis que la mutation (au sens large, incluant éléments transposables et variations épigénétiques) a été celui des espèces secondaires. La plupart des agrumes cultivés sont donc issus de la combinaison des génomes de quatre taxons ancestraux, résultant d'un nombre limité d'événements de méiose interspécifique. Ces espèces secondaires présentent ainsi une diversité génétique intra spécifique très étroite portant principalement sur des caractères sélectionnés par l'homme (période de production, qualité des fruits...). Elles sont donc particulièrement fragiles vis-à-vis des maladies émergentes, comme l'illustrent bien les ravages causés par la tristeza sur les vergers d'agrumes greffés sur bigaradier (*C. aurantium*) et les difficultés actuelles de l'agrumiculture brésilienne (monoculture de l'oranger, *C. sinensis*), confrontée à deux maladies bactériennes, la chlorose variéguée (*Xylella fastidiosa*) et le huanglongbing ou encore de l'agrumiculture mexicaine avec la culture de la lime qui vient de subir plusieurs hivers rigoureux et vient de voir arriver le huanglongbing<sup>11</sup>.

L'amélioration variétale des espèces secondaires présente un intérêt économique majeur (plus de 80% de la production mondiale des agrumes). Cependant, comme nous avons pu le voir au point 6.3. « *Les facteurs limitant l'amélioration génétique chez les agrumes*<sup>12</sup> », à ce jour, elle n'est pas envisageable par des schémas d'amélioration conventionnelle (recombinaisons sexuées) pour plusieurs motifs :

- leur structure très hétérozygote d'origine interspécifique ne permet pas un maintien des caractéristiques phénotypiques de base (le morphotype) après recombinaison sexuée,
- la multiplication clonale a contribué à l'accumulation sous forme récessive d'allèles néfastes ou délétères, qui, lors d'un croisement par un génotype proche, se trouvent à l'état homozygote et en s'exprimant entraînent un effet dépressif ou létal,
- de nombreux « cultivars » ont été sélectionnés sur la perte de leur capacité à produire des pépins (aspermie) et sont donc stériles,
- chez les cultivars produisant des fruits à pépins, les pépins sont le plus souvent polyembryonnés avec de multiples embryons somatiques limitant fortement la possibilité de régénération de l'embryon zygotique.

Compte tenu de l'ensemble de ces contraintes génétiques, actuellement, les seules possibilités d'amélioration de ces variétés sont, soit la mutagenèse, soit la transformation génétique. Néanmoins une autre voie pourrait être envisagée : celle de la reconstruction des structures génomiques hybrides à partir de combinaisons de génotypes des taxons ancestraux qui constituent les principaux réservoirs de diversité des formes cultivées. Cette approche pourrait être envisagée à condition de connaître très finement les structures du génome des espèces secondaires et de repérer les points de recombinaison entre chromosomes des taxons ancestraux.

---

<sup>11</sup> cf. *Les stress biotiques* ; point 5.2.

<sup>12</sup> cf. *Les facteurs limitant l'amélioration génétique chez les agrumes* ; point 6.3.

L'importance pour l'industrie agrumicole de ces thématiques de recherche fait que de nombreuses équipes travaillent sur le décryptage du génome des principales espèces commerciales. Une version préliminaire du génome de l'oranger (*Citrus sinensis*) issue de séquençage 454 a été disponible en ligne en 2012 (Xu *et al.*, 2013). Le génome de référence des agrumes issu de séquençage Sanger d'un haploïde de clémentinier a été proposé par le consortium international ICGC (2014) qui a également reséquéncé diverses variétés de mandariniers, pamplemoussiers ainsi qu'un bigaradier, un oranger et le clémentinier diploïde (Wu *et al.*, 2014). Enfin, le consortium espagnol CITRUSEQ / CITRUSGENN (<http://www.citrusgenn.es/>) a financé le séquençage complet de plus de 250 variétés d'agrumes (Terol *et al.*, 2012).

C'est dans ce contexte que ce travail de thèse a pu explorer différentes approches d'analyse de la diversité des génomes avec pour objectif d'identifier l'origine phylogénétique des différents groupes d'agrumes cultivés et de décrypter la structure phylogénomique de leurs génomes. Cette thèse a bénéficié de l'évolution particulièrement rapide des nouvelles générations de séquençage ou NGS (*New Generation Sequencing*) et peut ainsi proposer une utilisation raisonnée des outils disponibles en fonction des questions de recherches.

Le groupe des limettiers et citronniers, généralement moins connu et moins étudié que la plupart des autres agrumes, malgré leur importance économique (forte croissance du marché et rôle important dans l'amélioration des porte-greffe), a fait l'objet d'une analyse plus poussée. Il s'agit d'un groupe particulièrement complexe et intéressant car il présente la particularité de regrouper des individus diploïdes, triploïdes et tétraploïdes naturels, issus d'hybridations impliquant probablement les quatre taxons de base (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b).

Le séquençage parallèle d'amplicons en 454 (Pyroséquençage, Roche), qui permet d'obtenir une information haplotypique multi-loci sur des fragments d'environ 500 bases, a été utilisé dans l'objectif de décrypter la structure en mosaïques interspécifiques du génome des agrumes cultivés. Le chromosome 2 a été pris comme exemple et 50 variétés ont été haplotypées (Chapitre I). Ces travaux ont fait l'objet d'un article<sup>13</sup> accepté à *BMC Genetics* (Curk *et al.*, accepted,) et de deux communications à congrès (Curk *et al.*, in press, Annexe I.1) et Curk *et al.*, 2011, Annexe I.2). Une méthode simplifiée, permettant de préparer des banques d'amplicons 454 « taguées » pour chaque génotype a par ailleurs été validée (Annexe I.3).

Cette même technique de pyroséquençage 454 a permis d'identifier des polymorphismes de type SNP, diagnostiques des taxons ancestraux. Des marqueurs basés sur de la PCR compétitive entre allèles (technologie Kaspar) ont été développés à partir de ces SNP diagnostiques d'espèces. Leur transférabilité au sein des *Aurantioideae* a été testée et leur potentiel pour révéler les structures interspécifiques des espèces secondaires et des variétés modernes a été validé (Chapitre II). Ces travaux feront l'objet d'un article soumis à *PLOS ONE* (Curk *et al.*, submitteda).

---

<sup>13</sup> Pour un confort de lecture, nous avons fait le choix d'uniformiser la mise en page de l'ensemble des chapitres de la thèse sans respecter la mise en page spécifique des revues auxquels ont été soumis les articles correspondants.

Ces marqueurs diagnostiques d'espèce, en association avec des marqueurs SSR et indels permettant de révéler davantage de polymorphismes intraspécifiques, ont apporté un nouvel éclairage sur l'origine des citronniers et des limettiers diploïdes, triploïdes et tétraploïdes, à la suite d'un génotypage exhaustif des collections Inra-Cirad de San Giuliano en Corse et Ivia de Valence en Espagne (Chapitre III). Les résultats feront l'objet d'un article soumis à *Annals of Botany* (Curk *et al.*, prepared for submission). Par ailleurs, une analyse préliminaire sur un nombre restreint de variétés d'agrumes acides, mais intégrant des données de séquençage Sanger de fragments de gènes impliqués dans l'élaboration de la qualité et dans l'adaptation au stress a été présenté au Congrès de la société internationale de citriculture de 2012 (Curk *et al.*, in press ; proceeding, Annexe I.4).

Enfin, les données du consortium espagnol CITRUSEQ / CITRUSGENN (<http://www.citrusgenn.es/>) de reséquençage complet en Illumina de six variétés diploïdes de citronniers et de limettiers, et d'une variété triploïde de limettier, en comparaison avec celles de représentants des taxons ancestraux, ont été utilisées pour reconstituer la structure interspécifique de leurs génomes complets. L'objectif final étant de pouvoir schématiser, pour la première fois, leurs caryotypes phylogénomiques complets (Chapitre IV).

---

# Chapitre I

---

## Next generation haplotyping to decipher nuclear genomic interspecific admixture in Citrus species: analysis of Chromosome 2

Franck Curk<sup>12</sup>, Gema Ancillo<sup>2</sup>, Andres Garcia-Lor<sup>2</sup>, François Luro<sup>1</sup>, Xavier Perrier<sup>3</sup>, Jean-Pierre Jacquemoud-Collet<sup>3</sup>, Luis Navarro<sup>2</sup>, Patrick Ollitrault<sup>23</sup>

<sup>1</sup>UMR AGAP, Institut National de la Recherche Agronomique (Inra), Centre Inra de Corse, F-20230 San Giuliano, France

<sup>2</sup>Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada (Valencia), Spain

<sup>3</sup>UMR AGAP, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), TA A-108/02, 34398 Montpellier, Cedex 5, France

BMC Genetics, 2014

The electronic version of this article can be found online at:

<http://www.biomedcentral.com/1471-2156/15/152>

## Titre et Résumé en français

### L'haplotypage de nouvelle génération, un outil pour déchiffrer les structures en mosaïque interspécifique du génome des agrumes ; analyse du chromosome 2

#### État de l'art

Les principales espèces d'agrumes cultivés proviennent d'hybridations interspécifiques naturelles entre quatre taxons ancestraux (*Citrus reticulata*, mandariniers ; *C. maxima*, pamplemoussiers ; *C. medica*, cédratiers ; *C. micrantha*, papedas). Il en résulte des structures génotypiques complexes, généralement fixées par l'apomixie, fortement hétérozygotes et formées d'une mosaïque de grands fragments chromosomiques d'origines phylogénétiques différentes. Par ailleurs, la structuration de la variabilité phénotypique suggère que la différenciation initiale des quatre taxons de base supporte une part essentielle de la « variabilité utile » des agrumes. Une connaissance approfondie de l'origine des différentes formes cultivées et de leur structure phylogénomique est ainsi indispensable à la bonne gestion des ressources biologiques et à l'optimisation des programmes d'amélioration génétique. L'haplotypage de fragments de gènes le long du génome semble être une approche intéressante pour mettre en évidence la structure génomique des espèces cultivées et pour comprendre les séquences évolutives ayant formatées ce complexe d'espèces. Nous avons testé l'efficacité du séquençage en parallèle 454 (pyroséquençage) pour déchiffrer la structure hybride des espèces et des variétés d'agrumes modernes le long du chromosome 2.

#### Méthode

Des bibliothèques d'amplicons 454 ont été établies à l'aide du système Fluidigm pour 48 génotypes et 16 fragments de gènes du chromosome 2. Les haplotypes ont été établis à partir des lectures de chaque variété et l'analyse phylogénétique a été réalisée à partir des données haplotypiques de chaque fragment de gènes.

#### Résultats

La longueur des lectures 454 et le niveau de différenciation entre les taxons ancestraux ont permis une assignation phylogénétique d'haplotypes efficace pour 12 fragments de gènes sur 16. L'analyse de la structure génomique en mosaïque interspécifique des espèces secondaires et cultivars modernes (i) a révélé des introgressions de *C. maxima* dans les génomes des mandariniers modernes; (ii) était en accord avec les hypothèses antérieures sur l'origine des espèces secondaires; et (iii) a fourni une nouvelle vision de l'évolution du chromosome 2.

#### Conclusions

La méthode de séquençage 454 (Pyroséquençage, Roche) est efficace pour obtenir des haplotypes pouvant être assignés phylogénétiquement chez les agrumes, offrant ainsi une nouvelle vision de la structure en mosaïque interspécifique, le long du chromosome 2, de 48 génotypes d'agrumes.

#### Mots-clés

Phylogénie, haplotype, évolution, SNP, NGS, génome

## Abstract

### Background

The most economically important *Citrus* species originated by natural interspecific hybridization between four ancestral taxa (*Citrus reticulata*, *Citrus maxima*, *Citrus medica*, and *Citrus micrantha*) and from limited subsequent interspecific recombination as a result of apomixis and vegetative propagation. Such reticulate evolution coupled with vegetative propagation results in mosaic genomes with large chromosome fragments from the basic taxa in frequent interspecific heterozygosity. Modern breeding of these species is hampered by their complex heterozygous genomic structures that determine species phenotype and are broken by sexual hybridisation. Nevertheless, a large amount of diversity is present in the citrus gene pool, and breeding to allow inclusion of desirable traits is of paramount importance. However, the efficient mobilization of citrus biodiversity in innovative breeding schemes requires previous understanding of *Citrus* origins and genomic structures. Haplotyping of multiple gene fragments along the whole genome is a powerful approach to reveal the admixture genomic structure of current species and to resolve the evolutionary history of the gene pools. In this study, the efficiency of parallel sequencing with 454 methodology to decipher the hybrid structure of modern citrus species was assessed by analysis of 16 gene fragments on chromosome 2.

### Results

454 amplicon libraries were established using the Fluidigm array system for 48 genotypes and 16 gene fragments from chromosome 2. Haplotypes were established from the reads of each accession and phylogenetic analyses were performed using the haplotypic data for each gene fragment. The length of 454 reads and the level of differentiation between the ancestral taxa of modern citrus allowed efficient haplotype phylogenetic assignments for 12 of the 16 gene fragments. The analysis of the mixed genomic structure of modern species and cultivars (i) revealed *C. maxima* introgressions in modern mandarins, (ii) was consistent with previous hypotheses regarding the origin of secondary species, and (iii) provided a new picture of the evolution of chromosome 2.

### Conclusions

454 sequencing was an efficient strategy to establish haplotypes with significant phylogenetic assignments in *Citrus*, providing a new picture of the mixed structure on chromosome 2 in 48 citrus genotypes.

### Keywords

Phylogeny, haplotype, evolution, SNP, NGS, genome admixture, citrus



## Background

World-wide production of citrus was 131 million tonnes in 2011 and 2012 (FAO, 2014). The main citrus varietal groups are sweet oranges (52%), small citrus (21%), limes and lemons (12%), and grapefruits and pummelos (6%). The inter-varietal genetic diversity of most of these varietal groups is very scarce, particularly for sweet oranges, lemons, and grapefruits, where intra-group diversity results from clonal variation/selection in vegetatively propagated material (Ollitrault and Navarro, 2012). This confers a substantial fragility of these groups against emerging diseases, as demonstrated by the ongoing major crisis in the Brazilian and Floridian citrus industries (Grosser *et al.*, 2011; Texeira *et al.*, 2005; Wang and Trivedi, 2013). Moreover, conventional breeding of these species is hampered by their complex heterozygous genomic structures that determine species phenotype and are broken by sexual hybridisation. Therefore, most breeding efforts for sweet orange, grapefruit, and lemons to date have used natural or induced mutations and somaclonal variation (Grosser *et al.*, 2007). However, important natural phenotypically useful variability exists in the citrus gene pool particularly for resistance to biotic and abiotic constraints (Krueger and Navarro, 2007). The efficient mobilization of this biodiversity in innovative breeding schemes will require prior knowledge of varietal group origins and genomic structures.

The taxonomy of *Citrus* remains controversial due to the conjunction of broad morphological diversity, total interspecific sexual compatibility within the genus, and partial apomixis of many cultivars. Fixing complex genetic structures through seedling propagation via apomixis led some taxonomists to consider clonal families of interspecific origin as new species (Scora, 1975). Two major systems are widely used to classify *Citrus* species: the Swingle and Reece (1967) classification, which identifies 16 species, and the Tanaka (1961) classification, which recognizes 156 species. More recently, Mabberley (1997) proposed a new classification of edible citrus that recognized three species and four hybrid groups. In this paper, we will use the Swingle and Reece (1967) classification system. This taxonomic system is widely used in the citrus scientific community and, as mentioned below, mostly agrees with molecular data. Despite the difficulties involved in establishing a consensus classification system for edible citrus, most authors now agree on the origins of the main cultivated forms. Molecular analyses clarified the genetic underpinnings of various cultivated species of *Citrus* (Barkley *et al.*, 2006; Federici *et al.*, 1998; Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b). Four ancestral taxa [*C. medica* L. (citron), *C. reticulata* Blanco (mandarin), *C. maxima* (Burm.) Merr. (pummelo), and *C. micrantha* Wester (papeda)] were identified as the ancestors of all cultivated *Citrus* (Garcia-Lor *et al.*, 2012; Nicolosi *et al.*, 2000). Differentiation between these sexually compatible taxa may be explained by foundation effects in three distinct geographic zones and by an initial allopatric evolution. *C. maxima* originated in the Malay Archipelago and Indonesia, *C. medica* evolved in northeastern India and the nearby region of Myanmar and China, and *C. reticulata* diversification occurred over a region including Vietnam, southern China, and Japan (Scora, 1975; Webber *et al.*, 1967). Secondary species [*C. sinensis* (L.) Osb. (sweet orange), *C. aurantium* L. (sour orange), *C. paradisi* Macf. (grapefruit), *C. limon* (L.) Burm. (lemon), and *C. aurantifolia* (Christm.) Swing. (lime)] arose from hybridizations between the four basic taxa (Garcia-Lor *et al.*, 2012; Nicolosi *et al.*, 2000). Partial apomixis of most of the secondary species has been an essential element in the limitation of the number of further interspecific meiosis. Moreover, studies considering diversity of morphological characteristics (Barrett and Rhodes, 1976; Ollitrault *et al.*, 2003), primary metabolites (Luro *et al.*, 2011), and secondary metabolites (Fanciullino *et al.*, 2006)

confirmed that a major part of the phenotypic diversity of edible citrus resulted from differentiation between the basic taxa. In this context, deciphering the phylogenomic structures of the secondary citrus species is essential before innovative conventional breeding strategies can be developed.

Reticulations pose serious challenges in phylogenetic analyses and result in evolutionary histories that cannot be adequately represented in the form of phylogenetic trees (Arnold, 1997; Doolittle, 1999; Grant, 1981; Otto and Whitton, 2000; Stebbins, 1950). For many species, these relationships resemble a network with phylogenetic incongruities observed not only between cytoplasmic and nuclear genomes, but also between different regions of nuclear genomes (Beiko and Hamilton, 2006; Linder and Rieseberg, 2004; Pamilo and Nei, 1988; Rieseberg and Soltis, 1991). In plants such as citrus, where vegetative propagation such as apomixis took place immediately or a few generations after a reticulation event, large parts of the genome remain in interspecific heterozygosity. Genome-wide molecular analyses are, therefore, needed to decipher the complex interspecific mosaic genomes resulting from such evolution. Studies based on linkage disequilibrium can provide good evidence for recent and ancient hybridization events. This was demonstrated in sunflower by Rieseberg *et al.* (2003; 1996), who showed that the genomes of hybrid sunflower species contained chromosomal segments from both parental species. When examining heterozygous structures like citrus genotypes, phased multilocus studies offer improvements over monolocus analysis for the identification of interspecific heterozygous genome fragments deriving from reticulate events. The expectation is that tightly linked markers in a hybrid species are significantly more likely to come from the same parent and, therefore, to display linkage disequilibrium (Linder and Rieseberg, 2004). Sanger sequencing after bacterial cloning to separate gene copies was used effectively for such analysis (Fortune *et al.*, 2008; Ramadugu *et al.*, 2013; Rousseau-Gueutin *et al.*, 2009). However, because this is time-consuming and expensive, and only a few individuals and genes can be investigated, this type of analysis can miss intraspecific diversity components and may lead to erroneous conclusions about the evolutionary history of related taxa (Maddison and Knowles, 2006). In recent years, massively parallel sequencing of barcoded DNA mixtures enabled rapid and relatively inexpensive DNA sequence data production and facilitated genome-wide sequence variant discovery. This analysis was applied to a wide variety of bacteria, fungi (Jumpponen and Jones, 2009; Sønstebo *et al.*, 2010), multi-copy genes (Taudien *et al.*, 2010), and polyploids. In citrus, recent whole genome sequencing projects (Wu *et al.*, 2014; Xu *et al.*, 2013a) confirmed hybridization at the origin of *C. sinensis* and *C. clementina* (clementine) and allowed the phylogenetic origin of DNA fragments in the whole genome to be determined. However, the genomic structure of other secondary species and most modern varieties remain to be studied, and no analysis of the phylogeny of DNA fragments from the whole genome has yet been undertaken.

Whole genome sequencing (WGS) in large populations remains costly and requires considerable bioinformatic analysis. Major challenges include the need to reduce genome complexity and manage orthologous sequence data for a large number of individuals. Alternatives such as targeted capture (Okou *et al.*, 2007) or targeted amplicon (Bybee *et al.*, 2011) sequencing can be valuable. In human research, deep amplicon sequencing using 454 technology yielded thousands of haplotype calls per amplicon at the beta-defensin locus, and this was considered to be an efficient method for haplotyping and copy-number estimation in small to medium-sized cohorts (Taudien *et al.*, 2010). A particular advantage of using such an approach for haplotyping heterozygous structures is that sequencing data come from

single DNA molecules, and there is no requirement for cloning. Therefore, we hypothesize that, by using a sequencing method allowing enough long reads (over 500 bp) such as 454 pyrosequencing (Taudien *et al.*, 2010), it should be possible to establish multilocus haplotypes that are phylogenetically significant when working at a sufficient level of genetic differentiation between taxa.

The objective of this work was to analyze the potential of the 454 sequencing method for efficient targeted parallel haplotyping to decipher complex interspecific genomic structures resulting from reticulate evolution in citrus. Amplicons from 48 genotypes, representative of *Citrus* ancestral taxa and secondary species, were subjected to parallel sequencing. Sixteen targeted genes distributed across chromosome 2 were sequenced. Chromosome 2 was selected due to its complex admixture structure in sweet orange, as identified in our previous research (Garcia-Lor *et al.*, 2013; Wu *et al.*, 2014).

## Methods

### Plant material

Leaves from 48 accessions of the *Citrus* genus and one accession of *Severinia buxifolia* [Poir.] Tenore were collected from the IVIA Citrus Germplasm Bank of pathogen-free plants (Valencia, Spain; accessions with IVIA identification number) and the INRA/CIRAD Citrus collection of San Giuliano (Corsica, France; accessions with SRA identification number, Annex II, Table S.I.1). In addition, *in silico* data were mined (phytozome.net (Department of Energy's Joint Genome Institute and Center for Integrative Genomics, 2014)) from the haploid clementine used to establish the first high-quality reference sequence of Citrus (Wu *et al.*, 2014). The Swingle and Reece (1967) botanical classification for scientific names was adopted (Table I.1 and Annex II, Table S.I.1).

Table I.1 - Scientific names and number of accessions per common horticultural group

	Common horticultural group name	Swingle scientific name	Number of accessions
Ancestral groups	Pummelo	<i>Citrus maxima</i> (Burm.) Merr.	10
	Mandarin	<i>Citrus reticulata</i> Blanco	12
		<i>Citrus tachibana</i> (Mak.) Tan.	2
	Citron	<i>Citrus medica</i> L.	6
	Papeda	<i>Citrus micrantha</i> Wester	1
Secondary species or genotypes arising from hybridizations between ancestral groups	Bergamot	<i>Citrus aurantifolia</i> (Christm.) Swing.	1
	Lime	<i>Citrus aurantifolia</i> (Christm.) Swing.	1
	Alemow	<i>Citrus aurantifolia</i> (Christm.) Swing.	1
	Sour orange	<i>Citrus aurantium</i> L.	2
	Lemon	<i>Citrus limon</i> (L.) Burm.	5
	Grapefruit	<i>Citrus paradisi</i> Macf.	2
	Clementine	<i>Citrus reticulata</i> Blanco	3
	Sweet orange	<i>Citrus sinensis</i> (L.) Osb.	3
Out-group		<i>Severinia buxifolia</i> (Poir.) Ten.	1

The four ancestral taxa of the *Citrus* genus were represented by 31 accessions: 14 mandarins (12 *C. reticulata* and two *C. tachibana* (Mak.) Tan.), ten pummelos (*C. maxima*), six citrons (*C. medica*), and one papeda (*C. micrantha*). Representatives of secondary citrus species or genotypes included two diploid clementines (*C. reticulata*), the haploid clementine used to establish the whole citrus genome reference sequence (*C. reticulata*), three sweet oranges (*C. sinensis*), two sour oranges (*C. aurantium*), two grapefruits (*C. paradisi*), five lemons (*C. limon*), , one bergamot ( *C. aurantifolia*), one lime (*C. aurantifolia*), and one 'Alemow' (*C. aurantifolia*). These 18 genotypes were putative hybrids derived from the four ancestral taxa. One *Citrus* genus relative (*Severinia buxifolia*) was added as an out-group.

## **DNA extraction**

High molecular weight genomic DNA was extracted from leaf samples using the DNeasy Plant Mini Kit (Qiagen S.A.; Madrid, Spain) according to the manufacturer's instructions.

## **Target genomic fragment selection**

### *Chromosome 2 targeted genomic fragments*

The reference citrus whole genome sequence, released in Phytozome (Department of Energy's Joint Genome Institute and Center for Integrative Genomics, 2014) by the International Citrus Genome Consortium (ICGC), was used to select gene fragments in this study. The annotated genes file ("Cclementina\_182\_gene.gff3" file) was used and is available at the Phytozome web page ([Department of Energy's Joint Genome Institute and Center for Integrative Genomics, 2014](#)).

Duplicated and overlapping genes were discarded. SSRs were annotated (up to tetranucleotidic motifs and at least 11 bp sequences) and all genes presenting microsatellite motifs were eliminated. Finally, the genes were sorted by length, and 415 genes were selected, each with a length of 1000–2000 bp. This length was selected to facilitate the design of primers for efficient sequencing of 500–600 bp amplicons. Sixteen genes within chromosome 2 were chosen.

## **Amplicon library preparation**

For the 16 selected gene fragments of chromosome 2 (Annex II, Table S.I.2), 16 primer pairs were designed (according to the Access Array™ System for 454 Sequencing Platform User Guide (Fluidigm Corp., 2014)) and loaded on the Fluidigm Access Array. This method employed the same approach as the two-step PCR methods proposed by Bybee *et al.* (2011) and validated by Curk *et al.* (2011) for citrus. Two successive PCR reactions produced amplicons with specific multiplex identifiers (MIDs) and directional titanium primer sequences for each variety. PCR products were generated using a 48.48 Access Array IFC (Fluidigm 48.770 Digital PCR Workflow Quick Reference Card), and amplicon quality was checked using an Agilent 2100 Bioanalyzer (Agilent DNA 1000 Kit Guide). Next, equal volumes of the PCR products were pooled together to create one PCR product library. The PCR product library was purified using AMPure beads. After purification, the PCR product library was quantified using Quant-iT PicoGreen fluorimetry (Quant-iT™ PicoGreen® User Guide) before proceeding to emulsion PCR.

454 parallel sequencing was performed using a mixture of all the amplicons for all the genotypes. DNA from each genotype carried a different MID, as defined by Roche (Life Sciences Corp., 2009b). The 454 sequencing technique requires amplicon primers to contain a directional GS FLX Titanium primer sequence (which includes a four base library "key"

sequence) at the 5' end of the oligonucleotide in addition to the gene-specific sequence at the 3' end. To allow for automated software identification of samples after pooling and sequencing, MID sequences (Annex II, Table S.I.3) were added between primer A (or B) and the gene-specific sequences (Life Sciences Corp., 2009a).

Forty-eight DNA samples were amplified and parallel-sequenced on a GS FLX Titanium system (Roche 454). Haploid clementine gene fragment sequences were obtained from the reference citrus whole genome sequence (Phytozome (Department of Energy's Joint Genome Institute and Center for Integrative Genomics, 2014). *S. buxifolia* (out-group) gene fragments were obtained by PCR amplification performed using a Mastercycler Ep gradient S thermocycler (Eppendorf). PCR was conducted in a final volume of 25  $\mu$ l containing 0.027 U Taq DNA polymerase (Fermentas), 1 ng/ $\mu$ l of genomic DNA, 10  $\times$  PCR buffer (Fermentas), 0.2 mM of each dNTP, 1.5 mM MgSO<sub>4</sub>, and 0.2  $\mu$ M of each primer. The following PCR program was applied: denaturation at 94°C for 5 min; 40 cycles of 30 s at 94°C, 1 min at 55°C, and 2 min at 72°C; and a final elongation step of 4 min at 72°C. PCR product purification was performed using a QIAquick PCR purification kit (Qiagen S.A.). Amplicons were sequenced using the Sanger method from the 5' end using fluorescently labeled dideoxynucleotides (Big Dye Terminator Cycle Sequencing Kit v3.1).

### Sequencing and sequence data analysis for SNP calling

Raw reads obtained from 454 pyrosequencing were preprocessed by removal of low-quality reads and adapter/primer sequences using PRINSEQ (2012). Short reads (<150 bases) with primer dimers were considered to be low-quality reads. Remaining reads were automatically identified and sorted by MID and specific gene primers using the SFF Tool commands of Newbler software (Life Sciences Corp., 2010).

For each variety, 454 pyrosequencing reads were aligned independently for each gene using SeqMan NGen software version 7.0 (DNASTAR Inc., 2014) with the following assembly parameters: match size, 12; minimum match percentage, 80; and minimum sequence length, 150. For each gene fragment, consensus alignments from a homozygous sequence comprised one haplotype, while those from a heterozygous sequence comprised two haplotypes.

### Genetic analysis of SNP data

Unbiased expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), fixation index values ( $F_w$  (Wright, 1978)), and  $F_{stat}$  parameters ( $F_{ST}$  and  $G_{ST}$  Index) were calculated using GENETIX v. 4.03 software (Belkhir *et al.*, 1996-2004). SNP number and location were identified with SNIPlay online software (Dereeper *et al.*, 2011; Dereeper *et al.*, 2013). Principal component analysis (PCA) was performed using XLSTAT software.

Haplotype and genotypic phylogenetic relationships were studied by (i) neighbor-joining analysis (NJA), based on the SNP data using DARwin software (Perrier and Jacquemoud-Collet, 2006) with a simple matching dissimilarity index, (ii) maximum likelihood phylogenetic analysis using Mega software (Tamura *et al.*, 2013). The simple matching dissimilarity index was also used to infer intra- and inter-taxa average differentiation.

Graphical visualization of chromosome 2 genotypes was constructed using GGT2 software (van Berloo, 2008).

Population structure was inferred using Structure (version 2.3.4) software (Pritchard Lab, 2014), which implements a model-based clustering method using genotype data (Falush *et al.*, 2003; Pritchard *et al.*, 2000). No a priori population structure was defined. The linkage model option was used, with allele frequencies correlated and compute probability of the

data for estimating K. Analyses were made with K-values (number of subpopulations) of 1–10. The statistics used to select the correct K-value were those used by Evanno *et al.* (Evanno *et al.*, 2005). Ten runs using Structure software were performed, each with 50,000 steps of burning followed by 50,000 Monte Carlo Markov Chain (MCMC) repetitions using the linkage model, knowing Map distances between loci (Annex II, Table S.I.2) (Ollitrault *et al.*, 2012a). The independent Structure-run cluster outputs were permuted and aligned to match one another as closely as possible

## Results

### Read distribution

The first round of Fluidigm amplification/454 sequencing produced 64,170 reads. Of these, 11% were short reads with primer dimers, and 57,394 reads were therefore considered useful. Useful reads were classified according to their MID and titanium sequences, and MID sequences were removed using 454 software tools. All reads were attributed to one of the 768 ( $48 \times 16$ ) amplicons according to the fragment gene sequence. The average number of reads per amplicon was 75; however, the distribution of reads per amplicon (Figure I.1) was asymmetric, resulting in a high proportion of amplicons with insufficient coverage. Based on 454 single-read sequencing data error rates and our preliminary unpublished data, we defined a threshold level of 50 reads per amplicon for confident genotype calling. However, 305 amplicons (40%) had fewer than this initial threshold number. Detailed analysis of read distribution for each amplicon (Annex II, Table S.I.4) showed that much of the heterogeneity was due to global under-representation of three gene fragments and over-representation of five fragments. The total number of reads per variety was less heterogeneous than one per gene fragment. We therefore conducted a second round of Fluidigm/454 sequencing. A total of 159,490 useful reads was obtained (average 208 reads per amplicon) from the combination of the two runs (Annex II, Table S.I.5). The distribution of the number of reads per amplicon remained highly heterogeneous, and 135 amplicons (18% of the total gene fragments/varieties) still had fewer than 50 associated reads. In cases where number or quality of reads was insufficient for genotype calling, amplicons were Sanger sequenced to complete the genotypic data set. Sanger sequence analysis also allowed inference of haplotype if only one or no heterozygous loci were observed in the Sanger sequence (Annex II, Table S.I.5).

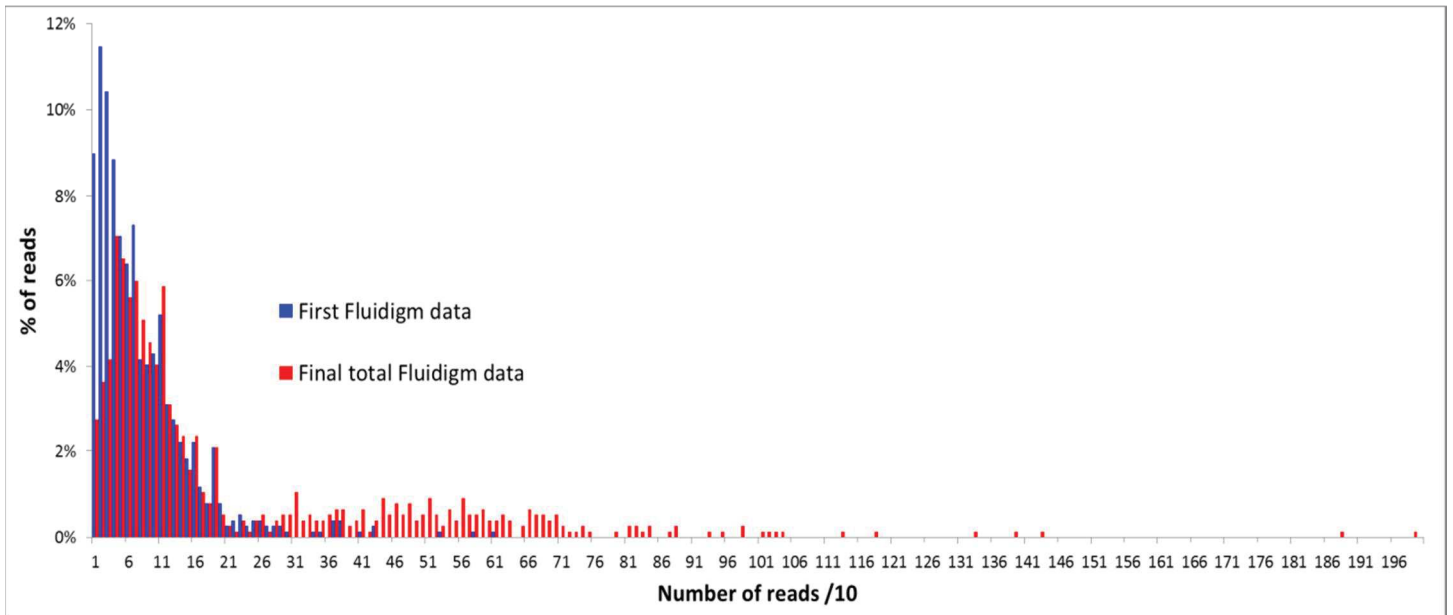


Figure I.1 - Distribution of the numbers of reads per amplicon for two rounds of Fluidigm/454 sequencing.

### Genotype calling and polymorphism of gene fragments

A total of 318 SNPs were identified from 7895 bp readable sequences for the 16 gene fragments within the 48 *Citrus* accessions (Table I.2). The web based SNIPlay tool (Dereeper *et al.*, 2011; Dereeper *et al.*, 2013) was used to analyze the intragenic location and potential impact of the different SNPs according to the whole genome annotation available at phytozome.net. The vast majority (98%) of the SNP loci was diallelic, but 2% (seven loci) were triallelic (Table I.2). The tri-allelism was validated by Sanger sequencing (data not shown). Sanger sequencing of the 2P33506778 fragment was performed for 32 *Citrus* varieties to estimate the 454 SNP-calling error rate. Only three differences between 454 and Sanger data were observed over 17,152 bp genotyping data (32 genotypes per 536 bp fragment; 0.02% error rate). The 'Clemenules' clementine was homozygous according to Sanger sequencing, but had two heterozygotic SNPs according to the 454 sequencing data. The 'Beauty' mandarin was shown to be heterozygous with the two techniques, but one of the three heterozygotic 454 SNPs was not identified in the Sanger data. The average SNP frequencies in intronic, exonic and 3' UTR regions were 53.57, 38.77, and 39.77 SNPs/kb, respectively. In addition, five indels were found in exonic regions (fragments 2P8108334, 2P26819388, and 2P32507721 contained one indel, and 2P29538734 contained two).

Table I.2 - SNP number and location for 16 gene fragments sequenced in 48 diploid Citrus genotypes

Gene fragment	Total Sequence				Intron			Exon			3'-UTR		
	Seq size	SNP	SNP / kb	Trialelic SNP	Seq size	SNP	SNP / kb	Seq size	SNP	SNP / kb	Seq size	SNP	SNP / kb
2P737170	452	22	48.67	Exon	0	0	-	452	22	48.67	0	0	-
2P3068140	421	14	33.25	-	337	12	35.61	84	2	23.81	0	0	-
2P4517048	502	12	23.90	-	0	0	-	316	4	12.66	186	8	43.01
2P8108334	502	40	79.68	Exon	0	0	-	502	40	79.68	0	0	-
2P11442721	547	21	38.39	Exon	0	0	-	547	21	38.39	0	0	-
2P13928427	502	21	41.83	-	0	0	-	336	15	44.64	166	6	36.14
2P21022460	538	11	20.45	-	0	0	-	538	11	20.45	0	0	-
2P25198627	454	12	26.43	-	128	7	54.69	326	5	15.34	0	0	-
2P26819388	535	22	41.12	Exon	0	0	-	535	22	41.12	0	0	-
2P29538734	541	36	66.54	Exon	190	12	63.16	351	24	68.38	0	0	-
2P30446231	475	28	58.95	-	216	15	69.44	259	13	50.19	0	0	-
2P32507721	463	16	34.56	-	0	0	-	463	16	34.56	0	0	-
2P33532337	459	9	19.61	-	0	0	-	459	9	19.61	0	0	-
2P33506778	536	6	11.19	Exon	0	0	-	536	6	11.19	0	0	-
2P35391362	449	19	42.32	-	108	6	55.56	341	13	38.12	0	0	-
2P36235952	519	29	55.88	Exon	141	8	56.74	378	21	55.56	0	0	-
<b>16</b>	<b>7895</b>	<b>318</b>	<b>40.28</b>		<b>1120</b>	<b>60</b>	<b>53.57</b>	<b>6423</b>	<b>244</b>	<b>37.99</b>	<b>352</b>	<b>14</b>	<b>39.77</b>

### SNP diversity differentiation

Previous molecular studies (Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2013; Wu *et al.*, 2014) showed that some varieties of the main *Citrus* cultivar groups had interspecific introgressions. Therefore, in this study, we differentiated mandarin, pummelo, and citron groups of their respective pure ancestral taxa: *C. reticulata*, *C. maxima*, and *C. medica*.

For genotypic based analyses, we refer to the modern varietal groups, while we focus on pure ancestral taxa for the haplotype phylogenetic analyses.

Only 19 of the 318 SNPs were not found in the accessions representing the four basic taxa. These rare alleles were identified in heterozygosity in secondary species ('Alemow', nine; sour oranges, four; bergamot, three, 'Volkamer' lemon, one; 'Mexican' lime, one; and grapefruit, one) and concerned 9 of the 16 gene fragments. The parameters of SNP genetic diversity given in Table I.3 (and detailed in Annex II, Table S.I.6 for each SNP position) were calculated without these 19 rare alleles.



Table I.3 - SNP genetic diversity within and between supposed ancestral varietal groups

	Whole Population			Citrons			Mandarins			<i>Citrus micrantha</i>			Pummelos			4
	Ho	He	F <sub>W</sub>	Ho	He	F <sub>W</sub>	Ho	He	F <sub>W</sub>	Ho	He	F <sub>W</sub>	Ho	He	F <sub>W</sub>	F <sub>ST</sub>
2P737170	0.11	0.23	0.52	0.02	0.03	0.33	0.01	0.01	-0.04	0.15	0.08	-1.00	0.08	0.07	-0.22	0.78
SD	0.07	0.15	0.25	0.02	0.11	0.00	0.03	0.02	0.00	0.36	0.18	0.00	0.04	0.15	0.08	0.33
CI	0.03	0.07	0.11	0.02	0.05	-	0.01	0.01	-	0.16	0.08	-	0.02	0.07	0.08	0.14
2P3068140	0.18	0.33	0.46	0.00	0.00	-	0.07	0.10	-0.01	0.00	0.00	-	0.02	0.02	-0.18	0.72
SD	0.08	0.17	0.22	0.00	0.00	-	0.12	0.10	0.13	0.00	0.00	-	0.04	0.07	-	0.40
CI	0.05	0.09	0.12	-	-	-	0.06	0.05	0.07	-	-	-	0.02	0.04	-	0.22
2P4517048	0.09	0.19	0.55	0.01	0.01	-0.09	0.10	0.08	-0.17	0.00	0.00	-	0.06	0.05	-0.14	0.48
SD	0.07	0.19	0.33	0.03	0.04	-	0.07	0.18	0.22	0.00	0.00	-	0.08	0.09	0.04	0.47
CI	0.04	0.11	0.19	0.02	0.02	-	0.04	0.10	0.25	-	-	-	0.05	0.05	0.04	0.27
2P8108334	0.12	0.20	0.42	0.01	0.04	0.80	0.04	0.06	0.38	0.21	0.10	-1.00	0.14	0.09	-0.52	0.52
SD	0.09	0.18	0.36	0.01	0.10	0.49	0.05	0.14	0.28	0.41	0.20	0.00	0.06	0.17	0.33	0.36
CI	0.03	0.06	0.11	0.01	0.03	0.39	0.03	0.04	0.18	0.13	0.06	-	0.04	0.05	0.20	0.11
2P11442721	0.15	0.18	0.18	0.02	0.02	-0.09	0.23	0.20	-0.19	0.00	0.00	-	0.06	0.06	0.04	0.32
SD	0.11	0.13	0.26	0.05	0.05	0.00	0.09	0.20	0.16	0.00	0.00	-	0.06	0.12	0.32	0.35
CI	0.05	0.05	0.11	0.04	0.02	0.00	0.05	0.08	0.09	-	-	-	0.04	0.05	0.28	0.15
2P13928427	0.11	0.17	0.38	0.04	0.05	0.17	0.06	0.05	-0.11	0.10	0.05	-1.00	0.01	0.01	-0.05	0.40
SD	0.08	0.16	0.23	0.04	0.15	0.00	0.11	0.08	0.05	0.30	0.15	0.00	0.03	0.03	0.00	0.38
CI	0.04	0.07	0.10	0.04	0.06	-	0.06	0.04	0.04	0.13	0.07	-	0.02	0.01	-	0.16
2P21022460	0.12	0.19	0.34	0.00	0.04	1.00	0.05	0.06	0.11	0.18	0.09	-1.00	0.10	0.07	-0.50	0.49
SD	0.08	0.13	0.28	0.00	0.13	-	0.07	0.13	0.11	0.39	0.20	0.00	0.03	0.16	0.35	0.41
CI	0.05	0.08	0.16	-	0.08	-	0.03	0.08	0.11	0.23	0.12	-	0.02	0.09	0.48	0.24
2P25198627	0.17	0.21	0.20	0.08	0.10	0.16	0.10	0.13	0.23	0.17	0.08	-1.00	0.20	0.15	-0.38	0.34
SD	0.14	0.16	0.20	0.08	0.18	0.61	0.14	0.18	0.22	0.37	0.19	0.00	0.10	0.22	0.05	0.33
CI	0.08	0.09	0.12	0.06	0.10	0.69	0.07	0.10	0.15	0.21	0.11	-	0.06	0.12	0.05	0.19
2P26819388	0.09	0.16	0.46	0.07	0.05	-0.33	0.08	0.13	0.45	0.32	0.17	-1.00	0.02	0.02	-0.33	0.25
SD	0.10	0.17	0.30	0.06	0.13	0.00	0.10	0.19	0.36	0.47	0.24	0.00	0.02	0.08	-	0.18
CI	0.04	0.07	0.12	0.05	0.05	-	0.07	0.08	0.21	0.19	0.10	-	0.01	0.03	-	0.07
2P29538734	0.17	0.26	0.34	0.00	0.03	1.00	0.19	0.16	-0.18	0.06	0.03	-1.00	0.05	0.05	-0.07	0.53
SD	0.11	0.16	0.28	0.00	0.09	0.00	0.16	0.17	0.09	0.23	0.11	0.00	0.06	0.10	0.37	0.39
CI	0.04	0.05	0.09	-	0.03	-	0.08	0.05	0.04	0.08	0.04	-	0.04	0.03	0.23	0.13
2P30446231	0.12	0.20	0.37	0.02	0.04	0.56	0.10	0.11	0.07	0.08	0.04	-1.00	0.10	0.16	0.39	0.47
SD	0.09	0.14	0.28	0.04	0.12	0.00	0.12	0.17	0.11	0.27	0.14	0.00	0.07	0.20	0.44	0.37
CI	0.03	0.05	0.11	0.03	0.05	-	0.06	0.07	0.07	0.10	0.05	-	0.05	0.08	0.26	0.14
2P32507721	0.14	0.21	0.31	0.00	0.00	-	0.29	0.31	0.06	0.18	0.08	-1.00	0.07	0.04	-0.60	0.43
SD	0.10	0.14	0.18	0.00	0.00	-	0.25	0.24	0.05	0.39	0.19	0	0.04	0.14	-	0.32
CI	0.06	0.08	0.10	-	-	-	0.13	0.14	0.03	0.23	0.11	-	0.02	0.08	-	0.18
2P33506778	0.18	0.29	0.37	0.00	0.00	-	0.18	0.17	-0.07	0.00	0.00	-	0.05	0.06	0.20	0.81
SD	0.13	0.17	0.11	0.00	0.00	-	0.17	0.21	0.02	0.00	0.00	-	0.08	0.15	-	0.20
CI	0.10	0.13	0.09	-	-	-	0.09	0.17	0.03	-	-	-	0.05	0.12	-	0.16
2P33532337	0.19	0.32	0.39	0.00	0.07	1.00	0.23	0.24	-0.14	0.00	0.00	-	0.04	0.09	0.52	0.65
SD	0.11	0.15	0.15	0.00	0.08	0.00	0.26	0.17	0.20	0.00	0.00	-	0.09	0.19	0.00	0.31
CI	0.07	0.10	0.10	-	0.05	-	0.14	0.11	0.15	-	-	-	0.06	0.12	-	0.21
2P35391362	0.22	0.37	0.41	0.01	0.01	-0.09	0.24	0.20	-0.24	0.00	0.00	-	0.02	0.06	-0.15	0.73
SD	0.09	0.13	0.16	0.02	0.04	-	0.24	0.18	0.10	0.00	0.00	-	0.03	0.09	0.09	0.24
CI	0.04	0.06	0.07	0.02	0.02	-	0.13	0.08	0.06	-	-	-	0.02	0.04	0.06	0.11
2P36235952	0.13	0.28	0.55	0.00	0.00	-	0.02	0.04	0.45	0.04	0.02	-1.00	0.08	0.07	-0.28	0.55
SD	0.08	0.21	0.34	0.00	0.00	-	0.04	0.11	0.35	0.20	0.10	0.00	0.02	0.13	0.15	0.47
CI	0.03	0.08	0.14	-	-	-	0.02	0.04	0.26	0.08	0.04	-	0.01	0.05	0.12	0.19
Total	0.14	0.23	0.29	0.02	0.03	0.46	0.12	0.12	-0.02	0.09	0.05	-1.00	0.07	0.07	-0.08	0.51
SD	0.10	0.17	0.28	0.01	0.10	0.53	0.05	0.17	0.24	0.19	0.15	0.00	0.02	0.14	0.37	0.38
CI	0.03	0.02	0.03	0.01	0.01	0.19	0.03	0.02	0.04	0.09	0.02	-	0.01	0.02	0.09	0.04

Ho: observed heterozygosity; He: expected heterozygosity; F<sub>W</sub>: fixation index; F<sub>ST</sub>: fixation index within population; SD: standard deviation; CI: confidence interval estimated with alpha = 0.05.

The whole population displayed a diversity index ( $H_e$ ) of 0.23 and a fixation index ( $F_w$ ) value of 0.29, suggesting an important population genetic structure of the analysed varietal sample. Mandarin and pummelo intra-diversity  $F_w$  values were close to zero, but intra-group polymorphism was higher in mandarin ( $H_e = 0.12 \pm 0.02$ ) than in pummelo ( $H_e = 0.07 \pm 0.02$ ). Citron displayed low heterozygosity ( $H_o = 0.02 \pm 0.01$ ) and diversity ( $H_e = 0.03 \pm 0.01$ ). Only one *C. micrantha* representative was available: the observed heterozygosity value ( $0.09; \pm 0.09$ ) was, therefore, calculated between the pummelo and mandarin values.

The average numbers of SNPs/kb between two varieties within and between the four supposed basic taxa were 1.26–3.93 SNPs/kb within groups and 10.41–14.56 SNPs/kb at the inter-group level (Table I.4).

Table I.4 - Intra- and inter-varietal group dissimilarities (average number of SNP/kb between two varieties)

	Mandarins	Pummelos	Citrons
Mandarins	3.93*		
Pummelos	10.41	2.06*	
Citrons	14.56	11.21	1.26*
<i>C. micrantha</i>	13.49	10.61	12.24

\*average number of SNP/kb at intra-specific level

For secondary species, no intraspecific polymorphism was observed for sweet oranges, grapefruits, and sour oranges, represented, respectively, by three, two, and two varieties. The two clementine cultivars were also found to be identical. Polymorphism was found between regular lemons and the other ones; however, the two regular lemons ('Eureka' and 'Lisbon') and 'Sweet' lemon were found to be identical. Acid citrus types (lemons, limes, 'Alemow', and bergamot) and sour orange displayed high  $H_o$  values ( $0.26\text{--}0.34 \pm 0.05$ ). Sweet orange ( $0.15 \pm 0.04$ ), clementine ( $0.19 \pm 0.04$ ), and grapefruit ( $0.12 \pm 0.04$ ) displayed comparatively lower heterozygosity levels (Annex II, Table S.I.7).

Structure software analysis was performed in the absence of a prior hypothesis for group number. Analysis of  $\Delta K$  identified  $K = 4$  as the optimal population number. The ten runs for  $K = 4$  displayed very homogeneous results (as shown by the average values (Figure I.2, Annex II, Table S.I.8). *C. medica*, *C. maxima*, and *C. micrantha* defined three populations, and five mandarins defined a fourth population. The magnitude of genetic differentiation between the groups was statistically confirmed by the pairwise  $F_{ST}$  values, which ranged from  $0.499 \pm 0.091$  for *C. maxima/C. micrantha* to  $0.719 \pm 0.087$  for *C. micrantha/C. medica* (Table I.5). Eight of the additional mandarins appeared to belong chiefly to this last group but exhibited introgression from the *C. maxima* group. 'Shekwasha' mandarin displayed a possible introgression of *C. micrantha*. Some cultivars displayed more pronounced genetic mixing. 'Alemow' and 'Mexican' lime had half their features from the *C. micrantha* group and half from the *C. medica* group. Similarly, sour oranges had half their features from each of the *C. reticulata* and *C. maxima* groups. Sweet orange and clementine were admixtures of the *C. maxima* and *C. reticulata* groups.. Regular and 'Sweet' lemons and bergamot were admixtures of three groups: *C. maxima*, *C. reticulata*, and *C. medica*. Close to half of the genetic material in 'Volkamer' and 'Meyer' lemons was of the *C. medica* group, and half was of the *C. reticulata* group (Figure I.2, Annex II, Figure S.I.1 and Table S.I.8).

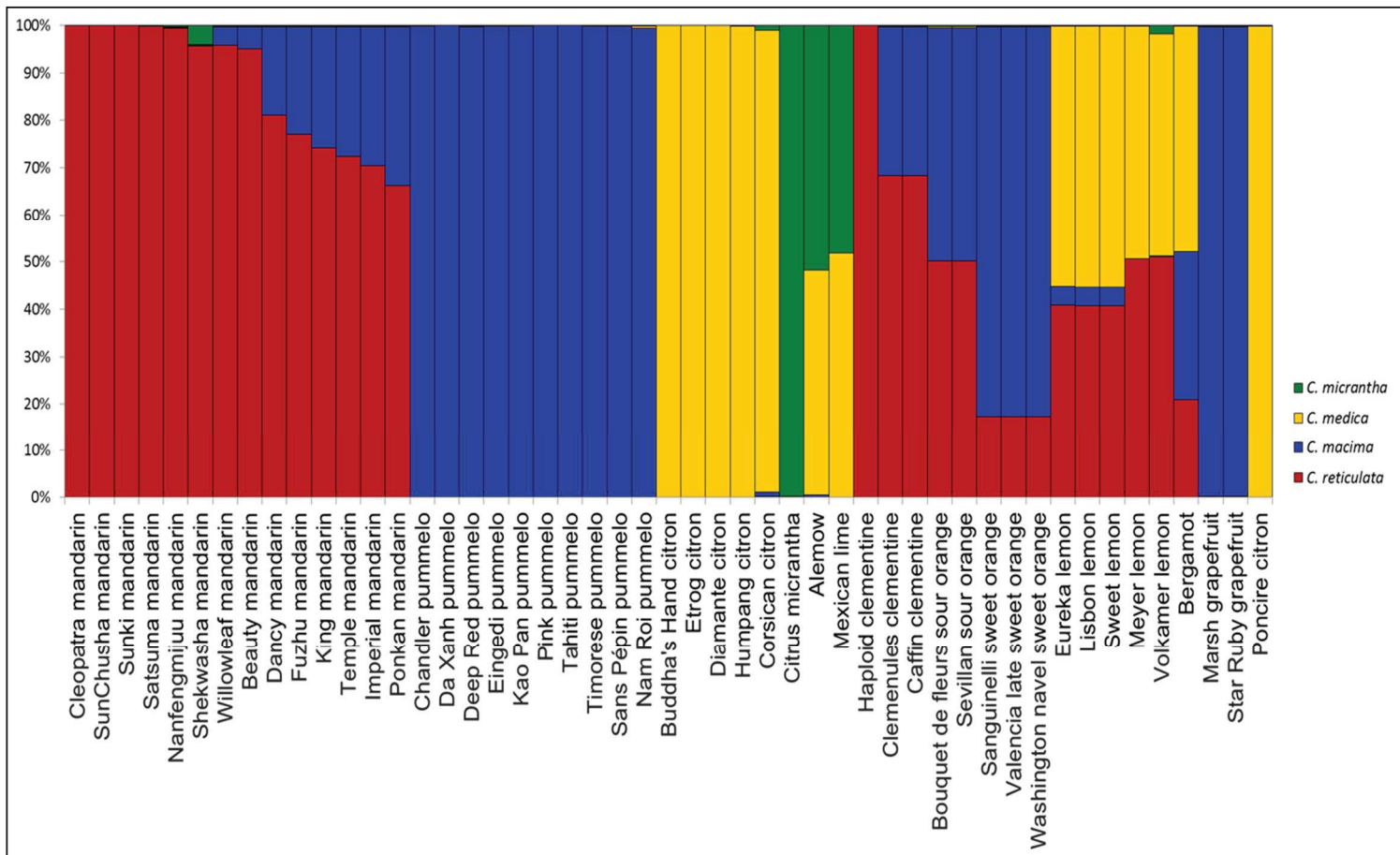


Figure I.2 - Estimated population structure representation based on the average values of ten Structure runs at K = 4.

Table I.5: Pairwise population F<sub>ST</sub> values

	Mandarin		Pummelo		Citron	
	SNP	F <sub>ST</sub>	SNP	F <sub>ST</sub>	SNP	F <sub>ST</sub>
Pummelo	172	0.502±0.061				
Citron	171	0.666±0.061	142	0.585±0.066		
<i>C. micrantha</i>	167	0.574±0.079	143	0.499±0.091	127	0.719±0.087

PCA analysis confirmed the organization of the whole diversity coming from the four ancestral varietal groups (Figure I.3). The three primary axes encompassed 56.3% of the total observed diversity. The first axis mainly separated citrons and *C. micrantha* from pummelos and mandarins. The second axis distinguished pummelos from other ancestral varietal groups. Finally, the third axis separated *C. micrantha* from other groups. ‘Alemow’ and ‘Mexican’ lime displayed intermediate positions between citrons and *C. micrantha*. Regular and ‘Sweet’ lemons and bergamot had intermediate positions between citrons and mandarins/sour oranges. Clementine lay within the mandarin cluster, while grapefruit was included in the pummelo cluster. Sweet orange and sour orange were located between the pummelo and mandarin clusters. The mandarin group displayed two noticeable subclusters. The subcluster that contained clementines and mandarins that were potentially introgressed by pummelo was displaced towards the pummelo cluster.

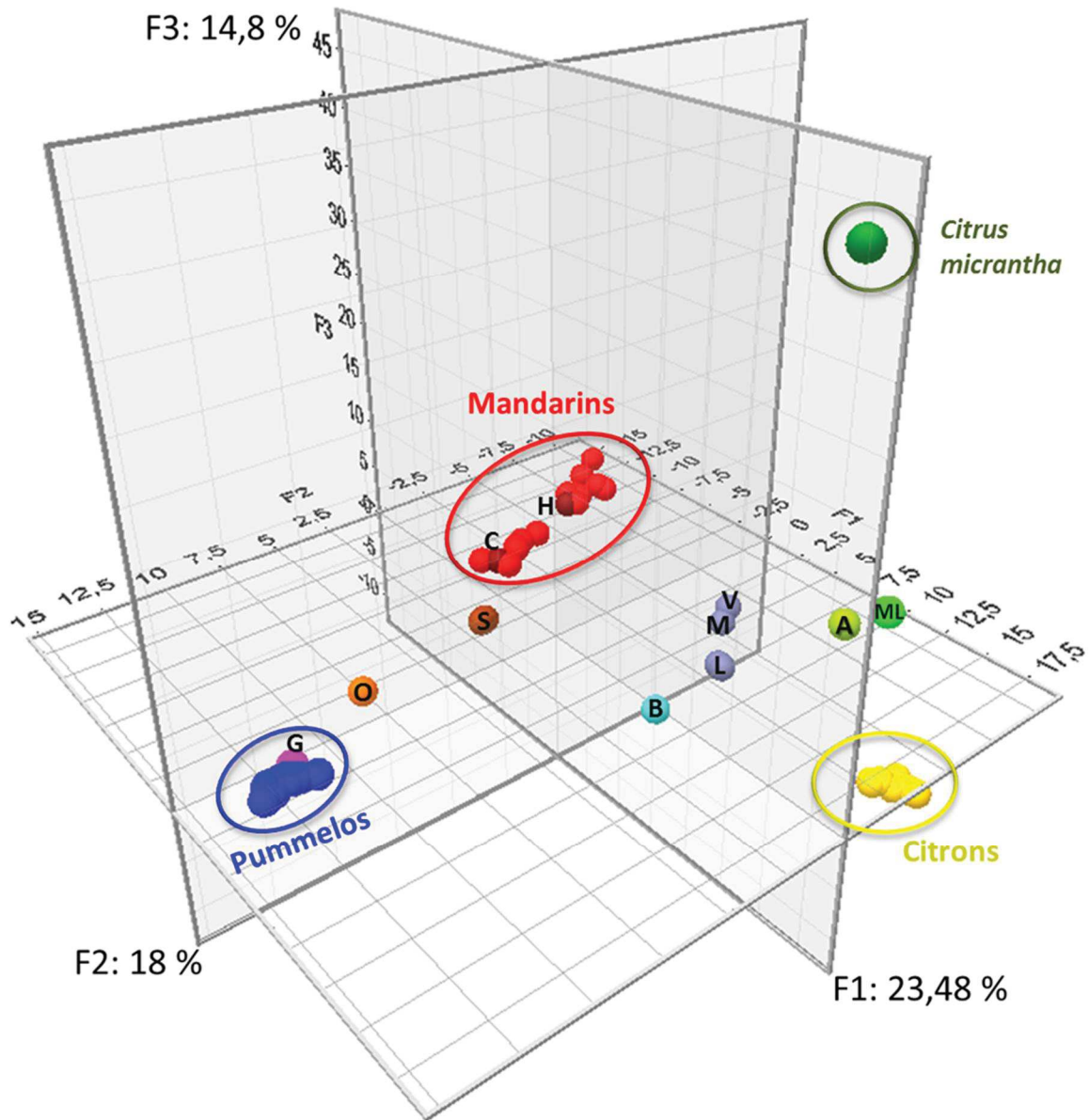


Figure I.3 - Organization of genotypic SNP diversity.

All varieties and all SNP data were analyzed by PCA. ML: 'Mexican' lime; A: 'Alemow'; V: 'Volkamer' lemon; M: 'Meyer' lemon; L: Regular and 'Sweet' lemons; B: Bergamot; H: Haploid clementine; C: Clementines; S: Sour oranges; O: Sweet oranges; G: Grapefruits.

Analysis of linkage disequilibrium (LD) between SNPs along the chromosome (Annex II, Figure S.I.2) also testified to a very high population genetic structure of the varietal sample. Significant LD values were observed across the whole chromosome, even for SNPs at distally opposing positions.

### Gene fragment haplotype inference and phylogeny

For each gene fragment, two haplotypes were inferred for each variety. NJA and maximum likelihood analysis of haplotypes was performed to determine phylogenetic relationships, and the two methods produced the same outcomes. For example, for the 2P35391362 gene fragment (Figures I.4), three, three, one, and two different haplotypes were identified in the *C. reticulata*, *C. maxima*, *C. micrantha*, and *C. medica* clusters, respectively.

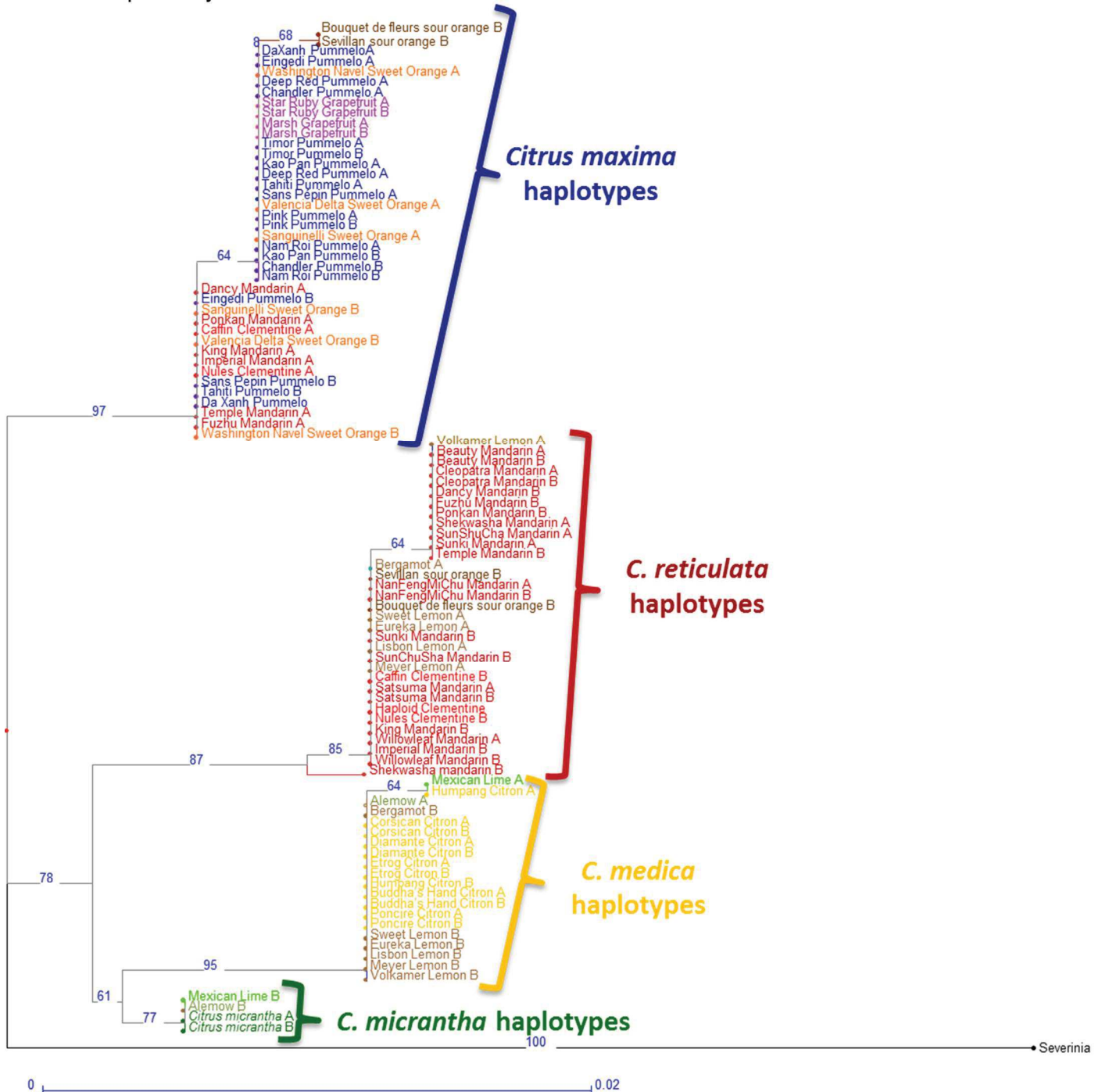


Figure I.4 - Neighbor-joining analysis (NJA) of the haplotypic data for the 2P35391362 gene fragment.

Multilocus haplotypic analysis also provided evidence of interspecific introgressions in varieties representative of one of the four supposed ancestral varietal groups. For this fragment, six mandarins shared one *C. maxima* haplotype with pummelos. Haplotypic analysis allowed clear inference of phylogenetic inheritance patterns for 2P35391362 in the secondary citrus species (Annex II, Figure S.I.3). For example, clementine clearly exhibited interspecific heterozygosity (*C. maxima/C. reticulata*): one haplotype was shared with sweet orange in the *C. maxima* cluster, and one was shared with 'Willowleaf' mandarin in the *C. reticulata* cluster. The second sweet orange haplotype was also in the *C. maxima* cluster and was shared with grapefruits that were homozygous for this haplotype. Evidence of interspecific inheritance was also found in sour orange (*C. maxima/C. reticulata*), bergamot (*C. medica/C. reticulata*), 'Eureka', 'Lisbon', 'Sweet', 'Volkamer', and 'Meyer' lemons (*C. medica/C. reticulata*), and 'Mexican' lime and 'Alemow' (*C. medica/C. micrantha*). NJA of genotypic information from the same data set (Figure I.5) provided a representation of two apparent *C. reticulata* clusters with unclear relationships. One of the clusters included accessions that exhibited interspecific inheritance when haplotype was assessed (several mandarins, sour oranges, and clementines). Similarly, lemons, limes, 'Alemow', and bergamot lay between *C. medica* and *C. micrantha*, clusters and branching did not provide definitive phylogenetic information.

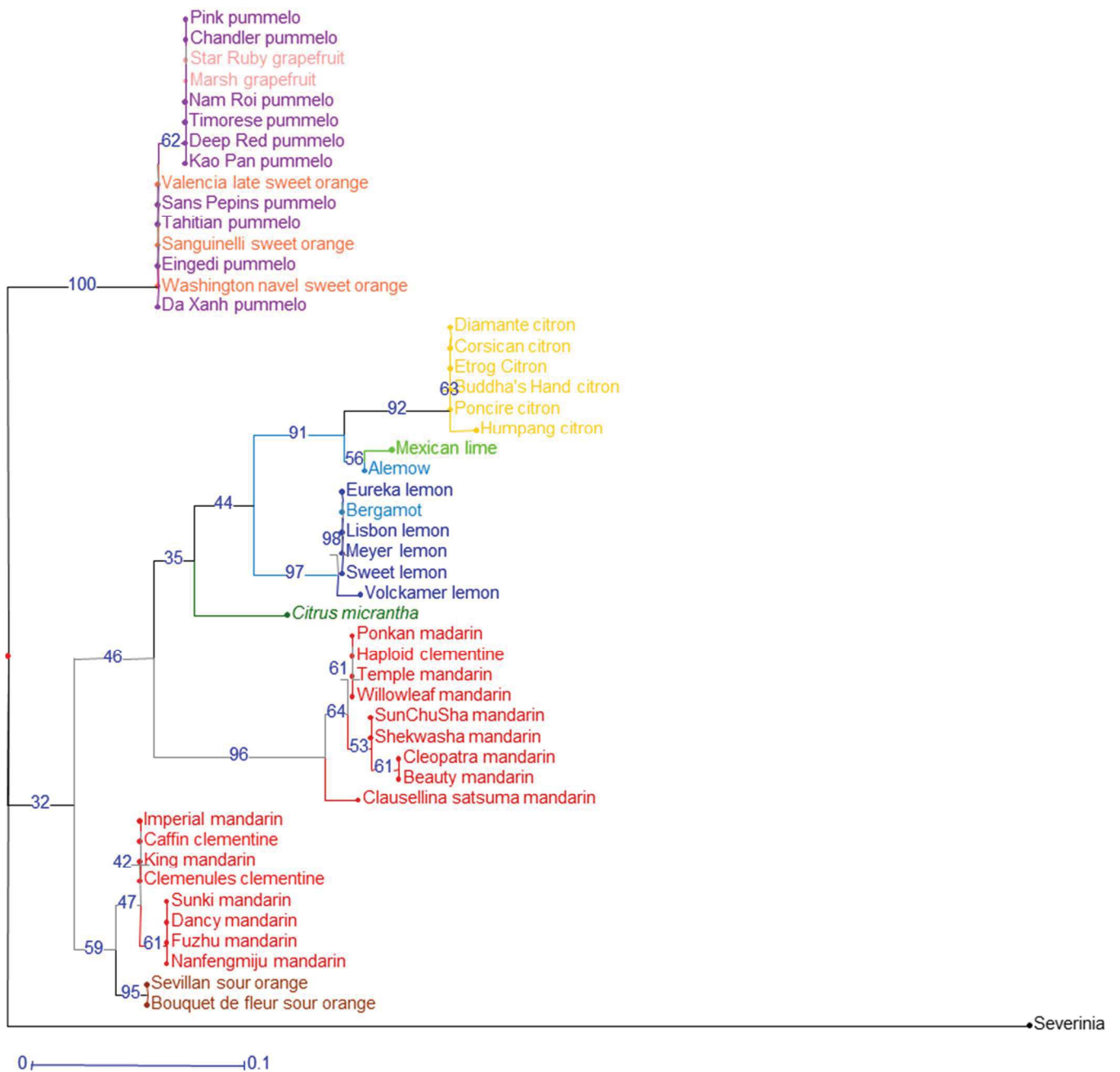


Figure I.5 - Neighbor-joining analysis (NJA) of the genotypic data for the 2P35391362 gene fragment.

A total of 210 haplotypes were identified through analysis of 16 gene fragments on chromosome 2 (Table I.6 and Annex II, Table S.I.9). From the phylogenetic analysis of each fragment, we considered 77, 58, 34, and 25 haplotypes to be representative of *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*, respectively. For 16 haplotypes, the organization of the genetic diversity of the corresponding fragment was insufficient to infer phylogenetic origin. The indeterminate haplotypes mostly concerned mandarin and pummelo.

Table I.6 - Number of haplotypes attributed to the four basic taxa or with indeterminate phylogenetic origin

Gene Fragment	<i>C. reticulata</i>	<i>C. maxima</i>	<i>C. medica</i>	<i>C. micrantha</i>	Indeterminate	Total
2P737170	4	6	2	2	0	14
2P3068140	4	2	1	1	1	9
2P4517048	5	3	2	1	0	11
2P8108334	10	7	3	2	2	24
2P11442721	8	5	2	1	0	16
2P13928427	3	2	2	2	2	11
2P21022460	1	2	2	1	2	8
2P25198627	5	1	3	2	1	12
2P26819388	8	2	2	1	2	15
2P29538734	7	6	4	2	0	19
2P30446231	6	7	3	2	1	19
2P32507721	3	1	1	2	4	11
2P33506778	2	1	1	1	1	6
2P33532337	2	2	1	1	0	6
2P35391362	3	3	2	1	0	9
2P36235952	6	8	3	3	0	20
<b>Total</b>	<b>77</b>	<b>58</b>	<b>34</b>	<b>25</b>	<b>16</b>	<b>210</b>

*C.: Citrus*

The haplotypic structure of each accession was used to schematize the phylogenetic origin of genome fragments along chromosome 2 (Figure I.6). In the absence of data regarding the phase between different haplotypes, this representation was made genotypically (homozygous for one ancestral taxon or heterozygous between two taxa). A single genotype was used to represent a varietal group when no polymorphisms were observed between varieties.



Chromosome 2

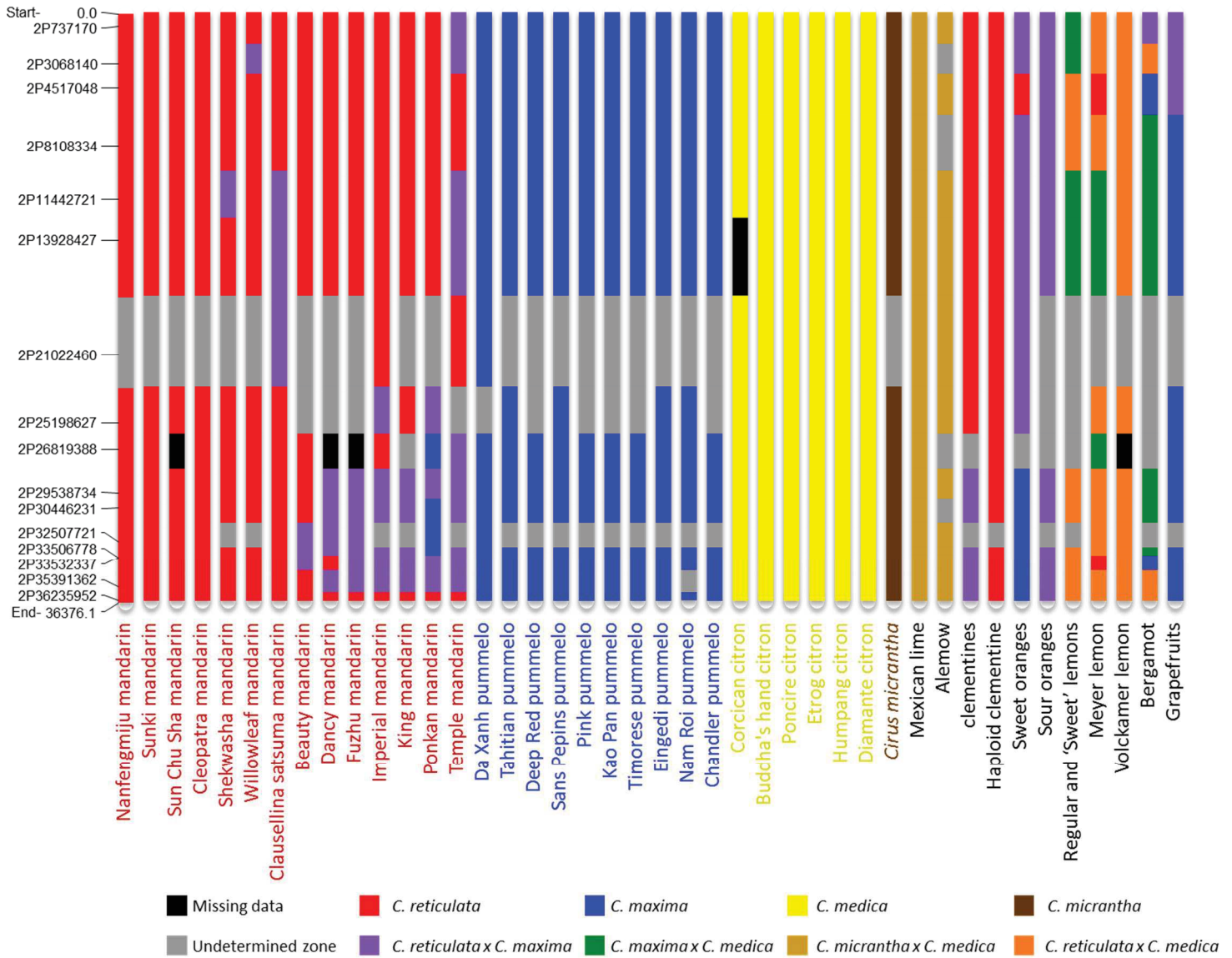


Figure I.6 - Genotypic structure of chromosome 2 in 48 Citrus varieties inferred from haplotypic data.

Ten of the fourteen mandarins were introgressed by *C. maxima*, mostly in heterozygosity. Two homozygous fragments for a *C. maxima* haplotype (ma1/ma1) and one fragment heterozygous for two *C. maxima* haplotypes (ma1/ma2) were found in ‘Ponkan’ mandarin. No evidence of interspecific introgression was observed for the representatives of the other three ancestral varietal groups. Completely heterozygous interspecific structures between *C. micrantha* and *C. medica* were observed for ‘Mexican’ lime and ‘Alemow’. Sour orange displayed complete heterozygosity between *C. reticulata* and *C. maxima*. Grapefruit appeared to have inherited mostly *C. maxima* haplotypes but displayed heterozygosity with *C. reticulata* at the start of the scaffold. Sweet orange was mostly heterozygous between *C. reticulata* and *C. maxima*, with a small fragment at the first part of the scaffold inherited solely from *C. reticulata*, and a genome area at the end of the scaffold inherited exclusively from *C. maxima*. Bergamot and regular, ‘Sweet’, and ‘Meyer’ lemons displayed similar structures that mainly comprised heterozygous regions of *C. medica*/*C. reticulata* and *C. medica*/*C. maxima*.

However, two small homozygous regions (2P4517048 and 2P33532337 gene fragments) were observed in 'Meyer' lemon (*C. reticulata* homozygosity re3/re3 and re2/re2) and bergamot (*C. maxima* homozygosity ma1/ma1 and ma2/ma2). No exploitable data were obtained for one gene fragment of 'Volkamer' lemon. For the other 15 gene fragments, 'Volkamer' lemon systematically displayed one haplotype corresponding with the *C. medica* cluster. The other haplotypes for 14 of these gene fragments were assigned to the *C. reticulata* cluster. The remaining haplotype was in a cluster of indeterminate phylogeny (Annex II, Table S.I.10).

### Revised genetic relationships between the four basic taxa after removal of introgressed genomic regions identified in mandarin from haplotypic analysis.

The identification, from haplotypic analysis, of introgressed pummelo fragments in mandarin genotypes prompted a revision of the relationships of the ancestral basic taxa (*C. maxima*, *C. reticulata*, *C. medica*, and *C. micrantha*) relative to the varietal groups deriving from these taxa (pummelos, mandarins, citrons and micrantha). The average SNP density within *C. reticulata* (Table I.7) was lower (2.85 SNP/kb) than in mandarin (3.93 SNP/kb) (Table I.4). Conversely, the *C. maxima*/*C. reticulata* average differentiation was 11.15 SNP/kb (10.41 SNP/kb between mandarins and pummelos). The differentiation values of *C. reticulata* with *C. micrantha* and *C. medica* were similar to those of mandarin with micrantha and citron, respectively.

Table I.7 - Intra- and interspecies group dissimilarity (average number of SNP/kb between two varieties) after elimination of introgressed haplotypes

	<i>C. reticulata</i>	<i>C. maxima</i>	<i>C. medica</i>
<i>C. reticulata</i>	2.85*		
<i>C. maxima</i>	11.15	1.86*	
<i>C. medica</i>	14.80	11.21	1.24*
<i>C. micrantha</i>	13.82	10.61	12.19

\*average number of SNP/kb at intra-specific level

For each SNP,  $G_{ST}$  values were estimated for each basic species relative to all other species. This allowed estimation of the value of each considered SNP to confirm that the surrounding genome fragment was inherited from the given species (SNP specific-diagnostic points). Corrections from the introgression information increased the number of diagnostic markers for *C. reticulata* and *C. maxima* relative to the initial data for mandarin and pummelo (Annex II, Figure S.I.4). The number of SNP loci with an average  $G_{ST}$  value >0.8 increased from 14 and 6 for mandarins and pummelos to 27 and 10 for *C. reticulata* and *C. maxima*, respectively. The highest number of totally discriminant SNPs ( $G_{ST} = 1$ ) was observed for *C. medica* (27) followed by *C. reticulata* (22), *C. micrantha* (21), and *C. maxima* (8) (Annex II, Table S.I.11).

## Discussion

### **Genotype and haplotype information from 454 parallel sequencing of 400–600 bp amplicons can identify admixture structures and infer the evolutionary history of species with reticulate evolution**

Three hundred heighten SNPs were found in 16 gene fragments from chromosome 2. The SNPs/kb rate within introns (53.6) was highly similar to the rate previously determined for the *Citrus* genus (51.5) by Garcia-Lor *et al.* (2013). The SNPs/kb rate within exons was slightly higher in this study (38.0) than in the previous study (29). Taken together, and including the small 3' UTR regions, 48.3 SNPs/kb were identified. This rate varied between gene fragments (range: 11.2–79.7).

The observed higher heterozygosity in secondary species than in the basic taxa, as well as the higher diversity in mandarin and pummelo compared to citron, was in agreement with previous studies (Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013; Ollitrault *et al.*, 2012b). Moreover, the high structuration of the diversity around *C. maxima*, *C. medica*, *C. reticulata*, and *C. micrantha* revealed by Structure and PCA agreed with previous molecular (Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2013; Herrero *et al.*, 1996; Nicolosi *et al.*, 2000) and numerical taxonomy (Barrett and Rhodes, 1976) studies, which recognizes the four basic taxa as the ancestors of the cultivated *Citrus* species. The important ancestral taxon differentiation and the limited number of reticulations and further interspecific hybridizations also resulted in the generalized LD observed in this study. LD was maintained even for fragments on opposing telomeres, also noted in previous studies for markers on different chromosomes (Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2012b).

The relative levels of differentiation between *C. maxima*, *C. medica*, *C. reticulata*, and *C. micrantha* varied (10.61–14.8 SNPs/kb), and was on average 6.7 times higher than the within-taxon diversity (from 1.24 in *C. medica* to 2.85 in *C. reticulata*). This diversity pattern allowed inferring haplotype phylogenetic origin for 12 of the 16 genes examined on chromosome 2. Differentiation was low for the four genes in the central part of the chromosome, and this resulted in clusters of indeterminate phylogenetic origin. The indeterminate haplotypes mainly concerned mandarins, pummelos, and their secondary species haplotypes.

Haplotype analysis demonstrated *C. maxima* introgressions in genotypes generally considered to be true mandarins. After removal of these haplotypes from the analysis of the supposed ancestral taxa, higher monolocus differentiation was observed between *C. reticulata* and *C. maxima*. This also allowed more precise estimations of *C. reticulata* intraspecific polymorphism. The identification of introgressed areas from haplotypic analysis, therefore, provided better species tree reconstruction. As recommended by Ramagudu *et al.* (2013), species trees can be improved by using loci that generate gene trees that are more clearly resolved. Haplotypic analysis has potential in this regard, and will allow the deselection of regions with incomplete lineage sorting or interspecific introgressions.

In the present study, 454 amplicon sequencing was successfully used to determine haplotypes in heterozygous genotypes and to analyze admixtures resulting from reticulate evolution. The broader utility of this method for identifying polymorphisms and inferring haplotype phylogenetic origins in other plants will depend on polymorphism rates within and between subspecies or species.

## Determination of the phylogenetic structure of chromosome 2 in several *Citrus* species and varieties provided insights into the origins of modern cultivated citrus

Haplotype NJA analysis of each gene fragment allowed the phylogenetic inheritance of genome fragments along chromosome 2 to be inferred for the 48 analyzed genotypes. Although a small number of haplotypes remained of indeterminate phylogenetic origin, the results provided an invaluable overview of the phylogenetic structure of chromosome 2 and the origin of modern *Citrus*.

The representative genotypes of the pummelo and citron horticultural groups appeared to be pure *C. maxima* and *C. medica*, respectively, and no interspecific introgressions were identified. Similarly, no evidence of introgression was found in *C. micrantha*. Conversely, evidence of introgression by *C. maxima* was found in 10 of the 14 mandarins studied. This corresponds with recent research (Wu *et al.*, 2014) in which WGS analysis of 'Willowleaf' and 'Ponkan' mandarins demonstrated introgression in these varieties considered to be true mandarins by citrus taxonomists. Three of the four mandarin varieties lacking evidence for introgression ('Cleopatra', 'Sunki', and 'Sun Chu Sha') are used mostly as rootstock and do not share the edible mandarin mitotype revealed by Froelicher *et al.* (2011). This particular mandarin clade should, therefore, probably not be considered as ancestral to modern cultivated mandarins. The fourth mandarin ('Nanfengmiju') without evidence for introgression shares the cytoplasm of edible mandarins.

The parentage hypothesis of some important commercial species and cultivars suspected to have arisen from reticulate evolution was checked by analyzing the haplotype phylogeny for each gene fragment (Annex II, Figure S.I.3).

*Citrus sinensis* (sweet oranges) and *Citrus aurantium* (sour oranges): phenotypic data (Barrett and Rhodes, 1976) and molecular marker studies (Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b; Uzun *et al.*, 2011; Uzun *et al.*, 2009) suggested that these two species derived from hybridizations between the *C. maxima* and *C. reticulata* gene pools. Both species have *C. maxima* maternal phylogeny as determined by chloroplast (Bayer *et al.*, 2009) and mitochondrial genome analysis (Froelicher *et al.*, 2011). In the present haplotype analysis within chromosome 2, sour orange displayed *C. maxima/C. reticulata* heterozygosity for each gene fragment. Sweet orange displayed *C. reticulata/C. reticulata* and *C. maxima/C. maxima* genome regions in addition to *C. maxima/C. reticulata* heterozygosity. The presence of a *C. maxima/C. maxima* region at the end of chromosome 2 disproves the hypothesis of a (*C. maxima* × *C. reticulata*) × *C. reticulata* ancestry proposed by Roose *et al.* (2009) from SSR data, and Xu *et al.* (2013) from WGS data. This was also determined by examination of two genes by Garcia-Lor *et al.* (2013) and confirmed by whole genome resequencing data from the ICGC (Wu *et al.*, 2014). These results suggest a possible direct F1 interspecific origin for sour orange and a more complex origin for sweet orange that would involve two parents each with *C. reticulata* and *C. maxima* admixture. These conclusions are in agreement with those proposed by the ICGC (Wu *et al.*, 2014). Considering that many mandarin cultivars are introgressed by *C. maxima*, a backcross model of (pummelo × mandarin) × mandarin rather than (*C. maxima* × *C. reticulata*) × *C. reticulata* would reconcile the Wu *et al.* (2014) and Xu *et al.* (2013) hypotheses. For 8 of the 16 gene fragments analyzed in the present study, both sweet

orange and sour orange were heterozygous but did not share haplotypes, therefore discarding the hypothesis of a direct relationship between them.

*Clementine*: It is generally agreed that, a little more than one century ago in Algeria, Father Clement selected clementine as a chance seedling from a 'Mediterranean' mandarin ('Willowleaf'). Previous molecular studies suggested that clementine was a mandarin × sweet orange hybrid (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2010; Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b), and this was recently confirmed by WGS analysis (Wu *et al.*, 2014). From the haplotype data, the larger part of chromosome 2 in clementine appears to be inherited from *C. reticulata*, with *C. maxima*/*C. reticulata* heterozygosity at the end of the orientated chromosome (phytozome.net (Department of Energy's Joint Genome Institute and Center for Integrative Genomics, 2014)) in agreement with WGS data (Wu *et al.*, 2014). The haplotype alleles of clementine, sweet orange, and 'Willowleaf' mandarin are in complete agreement with the hypothesis of a 'Willowleaf' × sweet orange origin.

*C. paradisi* (grapefruits): The origin of grapefruit is attributed to a natural hybridization between pummelo (*C. maxima*) and sweet orange (*C. sinensis*) in the Caribbean after the discovery of the New World by Christopher Columbus (de Moraes *et al.*, 2007; Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2012b; Scora *et al.*, 1982; Webber, 1943). The haplotype analyses agree with this hypothesis, showing coherent haplotypes for most of the gene fragments. In grapefruit, only one fragment (2P32507721) displayed a haplotype observed neither in sweet orange nor in the pummelo accessions (nor in other basic species clusters). However, this gene fragment displayed insufficient differentiation to allow full phylogenetic assignment, and the unassigned grapefruit haplotype may have been inherited from a pummelo not included in our limited samples. Chromosome 2 of grapefruit is mainly inherited from *C. maxima* and displays a small region of *C. maxima*/*C. reticulata* heterozygosity at the start of the scaffold.

*Citrus limon* (lemons): Based on RFLP, RAPD, and CAPS data, Nicolosi *et al.* (2000) proposed that "regular lemons" arose from hybridization between *C. aurantium* and *C. medica*. This hypothesis was supported by nuclear SSR (Garcia-Lor *et al.*, 2012) and SNP (Ollitrault *et al.*, 2012b) analyses. Moreover, the maternal *C. aurantium* parentage was confirmed by study of mitochondrial indels (Froelicher *et al.*, 2011). In the present study, 'Eureka', 'Lisbon', and 'Sweet' lemon varieties were highly heterozygous and identical. These lemons are very likely somatic mutants of the same hybrid ancestor. The three lemons display successive genome regions with *C. reticulata*/*C. medica* or *C. maxima*/*C. medica* heterozygosities. The haplotype allele analysis completely concurs with the sour orange × citron hypothesis. Indeed, systematic haplotype sharing between lemon and sour orange and the location of the second haplotypes within *C. medica* clusters were observed. 'Meyer' lemon also appeared to be of tri-specific hybrid origin (Garcia-Lor *et al.*, 2012) and displayed *C. maxima*/*C. medica* and *C. reticulata*/*C. medica* heterozygosity, as well as two gene fragments homozygous for a *C. reticulata* haplotype. Even if the 'Meyer' lemon were found to have a sweet orange-like mitotype (Froelicher *et al.*, 2011), as there were only two shared haplotypes between sweet orange and Meyer lemon over the 16 gene fragments, the haplotype analysis disproved the hypothesis that sweet orange was the female parent. 'Volkamer' lemon fragment gene haplotypes suggest that this genotype was a direct hybrid of *C. reticulata* and *C. medica*.

*Citrus aurantifolia* ('Mexican' lime, 'Alemow', and bergamot): These three citrus types were considered to be distinct species, namely, *C. aurantifolia*, *C. macrophylla*, and *C. bergamia* respectively, by Tanaka (1961). 'Mexican' lime and 'Alemow' displayed interspecific heterozygosity between haplotypes of the *C. medica* and the *C. micrantha* clusters. For 'Mexican' lime, exact haplotype sharing with the analyzed *C. micrantha* sample was found for 15 of the 16 gene fragments. This is in agreement with the hypothesis proposed by Nicolosi *et al.* (2000) that suggests 'Mexican' lime is a *C. micrantha* × *C. medica* hybrid. Maternal phylogeny was recently confirmed by mitochondrial marker analysis (Froelicher *et al.*, 2011). Similar results were observed for 'Alemow'. However, exact haplotype correspondence with the analyzed *C. micrantha* sample was found only for 12 gene fragments. This suggests that the maternal parent of 'Alemow' was closely related to the analyzed *C. micrantha*, which is in agreement with the Swingle and Reece (1967) description of 'Alemow' as a possible hybrid of *Citrus celebica* Koord (a papeda distinct from *C. micrantha*) or some other species of the subgenus *Papeda*, with a species of the subgenus *Citrus*. In 1811, Galesio (1811) proposed that bergamot was a hybrid between lemon and sour orange. However, alternative hypotheses were proposed based on molecular studies. Chen *et al.* (1991) suggested that bergamot could be a hybrid between citron and lime, Herrero *et al.* (1996) and Federici *et al.* (2000) proposed hybridization between sour orange and sweet lime, and hybridization between sour orange and citron was proposed by Nicolosi *et al.* (2000) and Li *et al.* (2010). The present haplotypic analysis disproved the hypotheses of hybridization between sour orange and citron, and between lemon and 'Mexican' lime, because bergamot displayed haplotypes not found in any of these theoretical parents.

### Implications for secondary species breeding

Some secondary apomictic species such as *C. aurantium* (*C. maxima* × *C. reticulata*) and *C. aurantifolia* (*C. micrantha* × *C. medica*), or genotypes such 'Volkamer' lemon (*C. reticulata* × *C. medica*), displayed interspecific heterozygosity for each gene fragment. They may have resulted directly from reticulation without further sexual recombination. For such secondary species, , innovative "like species" cultivars should be searched by direct hybridisation between the ancestral corresponding parental taxa, focusing on germplasm providing the suitable tolerance or resistance traits.

Conversely, other secondary species such as *C. sinensis* and *C. limon* ("Regular lemon" types) displayed more complex chromosome structures that testified to further interspecific recombination after the first reticulation events. For example, lemons ('Eureka', 'Lisbon', and 'Sweet' cultivars) systematically had one of their haplotypes within the *C. medica* cluster and the other in either the *C. maxima* or the *C. reticulata* cluster. Under our hypothesis of a sour orange × citron origin, the changes between *C. reticulata*/*C. medica* and *C. maxima*/*C. medica* heterozygosities along the chromosome suggest that at least three interspecific crossing over events occurred to produce the sour orange gamete that generated the lemon prototype. Previous studies (de Moraes *et al.*, 2007; Li *et al.*, 2010) and the present work demonstrated that grapefruit resulted from hybridization between pummelo and sweet orange.. For these three important citrus horticultural groups, it will be necessary to have a complete view of the nine chromosome admixture organizations to be able to rebuild similar genomic admixture structures from germplasm. Of these, "regular lemons" should be the simplest to assess despite the three-taxa structure, as it likely resulted from a

relatively straightforward sequence of interspecific hybridizations (*C. maxima* × *C. reticulata*) × *C. medica*). Genomic-assisted selection within progenies resulting from these crossing schemes should allow selection of very close interspecific mosaic structures. Such crossing will, however, be more complex for sweet orange and grapefruit because the two parents of sweet orange were themselves of interspecific origin. However, adequate pre-breeding at the parental level and genomic selection schemes over two or three generations should allow the reconstruction of similar interspecific mosaic genome structures from *C. maxima* and *C. reticulata* germplasm alongside desired resistance traits.

## Conclusion

Sixteen gene fragments on chromosome 2 were sequenced in 48 genotypes using 454 amplicon sequencing. The length of the reads and the level of differentiation between the ancestral taxa of modern citrus allowed efficient haplotype phylogenetic assignments for most gene fragments. The analysis of admixture genomic structures of modern species and cultivars revealed *C. maxima* introgressions in most modern mandarin cultivars. The haplotype results corresponded with previous hypotheses regarding the origin of many secondary citrus species, and provided a novel interpretation for the evolution of chromosome 2. Haplotyping of well-dispersed genome fragments should prove to be widely applicable, particularly for the analysis of evolutionary patterns within gene pools that experienced reticulate evolution. It is clear that this and other NGS methods will dramatically change methods of phylogenetic analysis. Regarding citrus breeding, the interspecific mosaic structure of all nine chromosome should be pursued, as this will provide the opportunity to rebuild the secondary species genomes from ancestral taxa bearing desirable traits.

## Competing interests

The authors declare that they have no competing interests.

## Author's contributions

FC and GA performed target genomic fragment selection, primer design, amplicon library preparation, genetic analysis of the SNP data, molecular genetic studies, and drafted the manuscript. AGL, FL, XP, and JPJC participated in target genomic fragment selection, and molecular genetic studies. LN participated in the design of the study and its coordination. PO conceived the study, and participated in its design and coordination, and drafted the manuscript. All authors read and approved the final manuscript.

## Acknowledgments

This work was supported by a grant (AGL2011-26490) from the Ministry of 'Economía y Competitividad'– 'Fondo Europeo de Desarrollo Regional' (FEDER) and a grant (Prometeo II/2013/008) from the Generalitat Valenciana, Spain.

We gratefully acknowledge David Karp for his help reviewing the manuscript.

---

## Chapitre II

---



# **Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern Citrus varieties**

Franck Curk<sup>1,2</sup>, Gema Ancillo<sup>2</sup>, Frederique Ollitrault<sup>2</sup>, Xavier Perrier<sup>3</sup>, Jean-Pierre Jacquemoud-Collet<sup>3</sup>, Andres Garcia-Lor<sup>2</sup>, Luis Navarro<sup>2</sup>, Patrick Ollitrault<sup>2,3</sup>

<sup>1</sup>UMR AGAP, Institut National de la Recherche Agronomique (Inra), Centre Inra de Corse, F-20230 San Giuliano, France

<sup>2</sup>Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada (Valencia), Spain

<sup>3</sup>UMR AGAP, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), TA A-108/02, 34398 Montpellier, Cedex 5, France

PLOS ONE, accepted.

## Titre et Résumé en français

# Les SNP nucléaires diagnostiques d'espèces développés à partir de séquençage 454 d'amplicons révèlent les structures génomiques interspécifiques des agrumes cultivés.

La majorité des agrumes du genre *Citrus* sont issus d'hybridations interspécifiques impliquant quatre taxons ancestraux (*C. reticulata*, *C. maxima*, *C. medica*, et *C. micrantha*) suivies de recombinaisons interspécifiques limitées par la multiplication végétative. Il en résulte des génomes présentant des hétérozygoties interspécifiques fréquentes. De plus, une grande partie de la diversité phénotypique des cultivars provient de la différenciation initiale des taxons de base. Déchiffrer la structure phylogénomique des ressources génétiques du genre *Citrus* est donc essentiel pour exploiter efficacement la biodiversité des agrumes dans les programmes d'amélioration variétale.

L'objectif de cette étude était (i) de développer un lot de marqueurs SNP diagnostiques des quatre taxons de base couvrant les 9 chromosomes et (ii) de les utiliser pour identifier la structure phylogénomique des espèces secondaires et des cultivars modernes. Les SNP diagnostiques d'espèces ont été sélectionnés à partir de données de séquençage parallèle (méthodologie 454) d'amplicons de 57 gènes pour 26 génotypes des 4 taxons de base. Parmi les 1053 SNP identifiés dans les 28 507 kb séquencées, 273 se sont révélés fortement diagnostiques pour l'un des taxons de base. 105 marqueurs SNP diagnostiques d'espèces (méthodologie Kaspar) ont été utilisés pour analyser la structure génétique des variétés et porte-greffe d'agrumes.

Ils ont révélés des introgressions de *C. maxima* pour presque tous les mandariniers et toutes les récentes sélections de petits agrumes. Ceci suggère que la réticulation *C. reticulata* x *C. maxima* et les processus d'introgression ont été particulièrement importants dans la domestication des mandariniers. Le large éventail de constitutions phylogénomiques entre *C. reticulata* et *C. maxima* révélé chez les mandariniers, tangelos, tangors, orangers, bigaradiers, pomelos et orangelos est favorable à la mise en place d'études de génétique d'associations basées sur les structures phylogénomiques. L'identification des structures génomiques est cohérente avec les hypothèses antérieures sur l'origine des diverses espèces secondaires et a permis d'identifier les origines probables d'un certain nombre de variétés d'agrumes acides (citronniers et limettiers). Les marqueurs SNP diagnostiques développés seront particulièrement utiles pour une évaluation systématique de la constitution interspécifique des ressources génétiques du genre *Citrus* et d'autres études génétiques.

## Abstract

Most cultivated *Citrus* species originated from interspecific hybridisation between four ancestral taxa (*C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*) with limited further interspecific recombination due to vegetative propagation. This evolution resulted in admixture genomes with frequent interspecific heterozygosity. Moreover, a major part of the phenotypic diversity of edible citrus results from the initial differentiation between these taxa. Deciphering the phylogenomic structure of citrus germplasm is therefore essential for an efficient utilization of citrus biodiversity in breeding schemes.

The objective of this work was to develop a set of species-diagnostic single nucleotide polymorphism (SNP) markers for the four *Citrus* ancestral taxa covering the nine chromosomes, and to use these markers to infer the phylogenomic structure of secondary species and modern cultivars. Species-diagnostic SNPs were mined from 454 amplicon sequencing of 57 gene fragments from 26 genotypes of the four basic taxa. Of the 1,053 SNPs mined from 28,507 kb sequence, 273 were found to be highly diagnostic for a single basic taxon. Species-diagnostic SNP markers (105) were used to analyse the admixture structure of varieties and rootstocks.

This revealed *C. maxima* introgressions in most of the old and in all recent selections of mandarins, and suggested that *C. reticulata* × *C. maxima* reticulation and introgression processes were important in edible mandarin domestication. The large range of phylogenomic constitutions between *C. reticulata* and *C. maxima* revealed in mandarins, tangelo, tangors, sweet orange, sour orange, grapefruit, and orangelo is favourable for genetic association studies based on phylogenomic structures of the germplasm. Inferred admixture structures were in agreement with previous hypotheses regarding the origin of several secondary species and also revealed the probable origin of several acid citrus varieties. The developed species-diagnostic SNP marker set will be useful for systematic estimation of admixture structure of citrus germplasm and for diverse genetic studies.

## Introduction

Citrus and its relatives are native in Southern to Eastern Asia, Malaysia, New Caledonia, and Australia (Swingle and Reece, 1967). The genus *Citrus* L. includes commercially important cultivars grown in tropical to temperate parts of the world over several thousands of years. Two major systems are widely used to classify *Citrus* species: the Swingle and Reece classification (1967), which considers 16 species, and the Tanaka classification (1961), which identifies 156 species. More recently, Mabberley (1997) proposed a new classification of edible citrus recognising three species and four hybrid groups.

In this paper, we will refer to the Swingle and Reece (1967) classification system widely used in the citrus scientist community. Despite the difficulties involved in establishing a consensual classification of edible citrus, molecular analyses provided decisive information for the comprehension of domestication and the relations between various cultivated species of *Citrus* (Barkley *et al.*, 2006; Federici *et al.*, 1998; Garcia-Lor *et al.*, 2013a; Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013b; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b). These studies identified four ancestral taxa [*C. medica* L. (citron), *C. reticulata* Blanco (mandarin), *C. maxima* (Burm.) Merr. (pummelo), and *C. micrantha* Wester (papeda)] as the ancestors of all cultivated *Citrus*. The differentiation between these sexually compatible taxa may be explained by the foundation effect in different geographic zones and initial allopatric evolution. *Citrus maxima* originated in the Malay Archipelago and Indonesia, *C. medica* evolved in north-eastern India and the nearby region of Myanmar and China, *C. reticulata* diversification occurred over a region including Vietnam, southern China, and Japan (Scora, 1975; Webber *et al.*, 1967) and *C. micrantha* seems to be originated from southern Philippian archipelago (Swingle and Reece, 1967). Moreover, diversity studies of morphological characteristics (Barrett and Rhodes, 1976; Ollitrault *et al.*, 2003), primary metabolites (Luro *et al.*, 2011), and secondary metabolites (Fanciullino *et al.*, 2006) proved that a major part of the phenotypic diversity of edible citrus resulted from differentiation between the basic taxa. Secondary species [*C. sinensis* (L.) Osb. (sweet orange), *C. aurantium* L. (sour orange), *C. paradisi* Macf. (grapefruit), *C. limon* (L.) Burm. (lemon), and *C. aurantifolia* (Christm.) Swing. (lime)] arose from hybridisations between the four basic taxa (Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013b; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b). The partial apomixis of most of the secondary species has been an essential element limiting the number of further interspecific meiosis events. Therefore, most of the genomes of cultivated *Citrus* are mosaics of large chromosome fragments from the basic taxa in frequent interspecific heterozygosity. Another consequence of apomixis and horticultural vegetative propagation practices is that most citrus horticultural groups (sweet oranges, limes, lemons, grapefruits, clementines and satsumas) have minimal intragroup genetic diversity resulting from clonal variation/selection (Ollitrault and Navarro, 2012). These horticultural groups are therefore particularly susceptible to emerging diseases. Moreover, conventional breeding of these varietal groups is hampered by the complex genetic structures that determine their specific phenotypes. However, useful natural phenotypic variability exists in the citrus gene pool, and traits are present for resistance to biotic and abiotic constraints (Krueger and Navarro, 2007). The efficient utilization of this biodiversity in innovative breeding schemes will require prior insight into the phylogenetic origin and genomic structures of secondary species and modern cultivars. Recent whole genome sequencing projects (Wu *et al.*, 2014; Xu *et al.*, 2013) confirmed that *C. aurantium*, *C. sinensis*, and *C. clementina* (clementine) resulted from

reticulation events between the *C. reticulata* and *C. maxima* gene pools and enabled to decipher the phylogenetic origin of genomic fragments over the whole genome. However, the genomic structures of other secondary species and most modern varieties resulting from sexual crosses remain to be studied. For such objective it is essential to identify diagnostic molecular polymorphisms of the four citrus basic species throughout the genomes, and to develop molecular markers for routine phylogenetic genotyping of large germplasm collections. Moreover, diagnostic markers for ancestral taxa will aid the management of interspecific introgression in sexual breeding schemes and, more widely, will enable studies of sexual recombination at the diploid or polyploid levels such as analysis of 2n gamete formation mechanisms for genotypes of interspecific origin.

Simple sequence repeat markers (SSRs) were widely developed in citrus during the last 15 years (Biswas *et al.*, 2012; Kijas *et al.*, 1997; Liu *et al.*, 2013; Luro *et al.*, 2008; Ollitrault *et al.*, 2010) SSRs are (Barkley *et al.*, 2006; Gulsen and Roose, 2001a; Gulsen and Roose, 2001b) advantageous because they are highly polymorphic, codominant, generally locus-specific, and randomly dispersed throughout the genome. However, Barkley *et al.* (2009) showed that homoplasmy may limit the usefulness of SSR markers in identifying the phylogenetic origin of DNA fragments in citrus. Garcia-Lor *et al.* (2012) and Ollitrault *et al.* (2014) showed that insertion/deletion polymorphisms (indels) and single nucleotide polymorphisms (SNPs) were more suitable as efficient specific diagnostic markers. Recent nuclear phylogenetic studies based on amplicon Sanger sequencing (Garcia-Lor *et al.*, 2013b; Ramadugu *et al.*, 2013) revealed SNPs differentiating the four basic taxa. These SNPs were used for successful development of a diagnostic SNP KASPar assay (Garcia-Lor *et al.*, 2013a). Several diagnostic SNPs for differentiation of *C. maxima* and *C. reticulata* were also identified from SNP markers developed from clementine Bac-end sequencing (Ollitrault *et al.*, 2012b). In addition, Wu *et al.* (2014) revealed a huge number of diagnostic SNPs differentiating *C. maxima* and *C. reticulata* through analysis of whole genome sequencing (WGS) data from several mandarins and pummelos. However, the currently available validated specific diagnostic markers are still low in number, genome coverage is patchy, and markers are particularly scarce for *C. medica* and *C. micrantha*.

The objective of the present work was (i) to develop a set of species-diagnostic markers for the four *Citrus* ancestral taxa with coverage of the nine chromosomes of the citrus haploid genome, and (ii) to estimate the interspecific admixture genomic structure of the secondary cultivated species and several modern cultivars from 20<sup>th</sup> century breeding programs. Specific diagnostic markers were mined from 454 amplicon sequencing of 57 gene fragments from 26 representative genotypes of the four basic taxa (eleven mandarins, nine pummelos, five citrons, and one papeda). Eighty-five SNP marker analyses based on competitive allele-specific PCR were developed. Effectiveness of marker development and transferability to related genera of the Aurantioideae subfamily is described in addition to the identification of ancestral alleles and validation of specific mutation occurrences in the four phylogenetic branches. Admixture analysis was performed using 73 markers successfully developed from the 454 SNP mining and 32 specific diagnostic SNP markers from previous research (Garcia-Lor *et al.*, 2013a; Ollitrault *et al.*, 2012b).

## Material and Methods

### Plant material

Leaves from 86 accessions of the *Citrus* genus and related genera were collected from pathogen-free plants from the IVIA Citrus Germplasm Bank (Valencia, Spain; accessions with IVIA identification numbers) and the Inra/Cirad citrus collection of San Giuliano (Corsica, France; accessions with SRA identification number, Annex III, Table S.II.1). The Swingle and Reece (1967) botanical classification was used for scientific names.

Twenty-six accessions representative of the four basic taxa (eleven mandarins [nine *C. reticulata* and two *C. tachibana*], nine pummelos [*C. maxima*], five citrons [*C. medica*], and one papeda [*C. micrantha*]) were used for SNP mining by 454 amplicon sequencing.

The study of admixture genomic structure of modern varieties using KASPar SNP markers was based on 70 accessions, including 24 of the 26 accessions noted above. For this study, the four ancestral taxa of the *Citrus* genus were represented by 33 accessions: seventeen mandarins, eight pummelos, six citrons, and two *C. micrantha* (Annex III, Table S.II.1).

Representatives of secondary citrus species included ten limes and lemons (four *C. aurantifolia*, six *C. limon*), three sour oranges (*C. aurantium*), three sweet oranges (*C. sinensis*), three grapefruits (*C. paradisi*), one ‘Combava’ (*C. hystrix*), and one ‘Nasnaran’ mandarin (*C. amblycarpa*). Sixteen recent hybrid varieties from international breeding programs or supposed natural interspecific hybridisation were also used (four mandarin hybrids, eight tangors, three tangelos, and one orangelo, Annex III, Table S.II.1).

Transferability of the KASPar markers across the Aurantioideae subfamily was studied by the analysis of 14 accessions representative of the two tribes of the Aurantioideae (*Clausenae* and *Citreae*). In *Clausenae*, the subtribe *Clauseniae* was represented by two accessions (*Clausena excavata* Burm. f. and *Murraya koenigii* (L.) Spreng). Within the *Citreae*, two subtribes were represented: *Triphasilineae* (*Triphasia trifolia* (Burm. F.) P.Wils.; one accession), and *Citrinae* (twelve accessions representing nine genera including the *Citrus* genus). Analysis of the *Citrinae* was conducted according to the subdivision into three groups proposed by Swingle and Reece (1967): one accession of the “primitive citrus fruit” group (*Severinia buxifolia*), two accessions of two genera of the “near citrus fruit” group (*Atalantia ceylanica* (Arn.) Oliv and *Citropsis gilletiana* Swingle & M. Kellerm), and nine accessions of the “true citrus fruit trees” group that included six genera (two *Fortunella*, two *Poncirus*, one *Eremocitrus*, one *Microcitrus*, one *Clymenia*, and one presumed intergeneric hybrid) in addition to the *Citrus* species (Annex III, Table S.II.1).

### DNA extraction

High molecular weight genomic DNA was extracted from leaf samples using the DNeasy Plant Mini Kit (Qiagen S.A.; Madrid, Spain) according to the manufacturer’s instructions.

### Target genomic fragment selection

The reference citrus whole genome sequence, released in Phytozome (Department of Energy’s Joint Genome Institute and Center for Integrative Genomics, 2014) by the International Citrus Genome Consortium, was used to select gene fragments. Annotated genes were acquired from:

<ftp://ftp.jgi-psf.org/pub/compugen/phytozome/v9.0/Cclementina/annotation/> at the Phytozome webpage (“Cclementina\_182\_gene.gff3” file).

Duplicated and overlapping genes were discarded. Then, from a specific annotation of the whole sequence for SSR (up to tetranucleotidic motifs and at least 11 bp sequences), all genes presenting microsatellite motifs were eliminated. For this study, 57 gene fragments covering the nine chromosomes were selected for SNP mining in genomic areas complementary to previously identified SNP marker sets (Garcia-Lor *et al.*, 2013a; Ollitrault *et al.*, 2012b).

### **Amplicon library preparation**

The 454 sequencing technique requires amplicon primers containing a directional GS FLX Titanium primer sequence (which includes a four base library “key” sequence) at the 5’ portion of the oligonucleotide in addition to the gene-specific sequence at the 3’ end. Multiplex Genotype Identifier (MID) sequences defined by Roche (Life Sciences Corp., 2009b) (Annex III, Table S.II.2) were added between the primer A (or B) and gene-specific sequences to allow for automated software identification of samples after pooling and sequencing (Life Sciences Corp., 2009a).

For the 57 selected gene fragments (Annex III, Table S.II.3), 57 primer pairs were designed according to the Access Array™ System for 454 Sequencing Platform User Guide (Fluidigm Corp., 2014) and loaded on the Fluidigm Access Array. The PCR products generated on the 48.48 Access Array IFC (Fluidigm 48.770 Digital PCR Workflow Quick Reference Card) were first analysed using an Agilent 2100 Bioanalyzer (Agilent DNA 1000 Kit Guide) to check the quality of the PCR products. Next, the PCR products were pooled in equal volume to create one PCR product library. The PCR product library was purified using AMPure beads. After purification, the PCR product library was quantified before proceeding to emulsion PCR. The PCR product library was quantified using the Quant-iT PicoGreen fluorimetry system (Quant-iT™ PicoGreen® User Guide).

### **Sequencing and sequence data analysis for SNP calling**

Raw reads obtained from 454 pyrosequencing were pre-processed by removing low quality reads and adapter/primer sequences using PRINSEQ (2012). High quality reads were automatically identified and sorted by MID and specific gene primers using SFF Tool commands of Newbler software (Life Sciences Corp., 2010).

For each sequenced gene of each variety, 454 pyrosequencing reads were aligned using SeqMan NGen software version 7.0 (DNASTAR Inc., 2014). Consensus genomic sequences were generated from alignments.

### **KASPar genotyping**

SNP genotyping was performed using KASPar technology (KBioscience®; <http://www.kbioscience.co.uk/>). The KASPar™ Genotyping System is a competitive, allele-specific dual Förster resonance energy transfer (FRET)-based assay for SNP genotyping. Primers were designed by LGC Genomics® based on the SNP locus-flanking sequence (approximately 50 nucleotides either side of the SNP). Two 40-mer allele-specific oligonucleotides and one common 20-mer oligonucleotide were defined for each locus. Detailed information for all SNP markers can be found in Annex III, Table S.II.4. Additional details about this genotyping method can be found in Cuppen (2007).

The fluorescence signals of PCR products were measured with Fluostar Omega (BMG) and genotype calling was made with KlusterCaller software (LGC Genomics).

### Genetic analysis of the SNP data

SNP numbers and locations were identified from sequence data using SniPlay online software (Dereeper *et al.*, 2011; Dereeper *et al.*, 2013).

Unbiased expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), and F Stat parameters ( $F_w$  (Wright, 1978) and  $F_{ST}$ ) were calculated using GENETIX v. 4.03 software (Belkhir *et al.*, 1996-2004).

The search for SNPs diagnostic for each taxa was based on  $G_{ST}$  parameter (Nei, 1973) estimations for the concerned taxa considering two subpopulations: (1) the concerned taxa ( $T_i$ ), and (2) a theoretical population of the three other basic taxa ( $T-i$ ). Analysis was performed from the estimated allele frequency of each taxa considering the same population size for each taxa to estimate the frequency of the two subpopulations ( $T_i$  and  $T-i$ ) and the whole population ( $Tot$ ) frequency.  $G_{ST}$  estimations were computed using Excel software:

$$G_{ST\ Taxai} = (H_{eTot} - (H_{eTi} + H_{eT-i})/2) / H_{eTot}$$

where  $H_e$  is the expected proportion of heterozygous loci per individual ( $H_e = 1 - \sum p_i^2$ , where  $p_i$  is the frequency of a given allele in the considered population or subpopulation). Values of  $G_{ST}$  range from 0 to 1. Low values indicate that little variation is proportioned among subpopulations, while high values denote that a large amount of variation is found among subpopulations. In this study,  $G_{ST\ Taxai}=1$  indicated that the taxa  $i$  was totally differentiated from the three other basic taxa and probably fixed for a mutant allele that most likely occurred in the taxa  $i$  after its separation.

Factorial Analyses from fragment sequences were performed using DARwin software (Perrier and Jacquemoud-Collet, 2006).

Principal Component Analysis (PCA) was performed using XLSTAT software.

Genotypic genetic relationships were studied by Neighbour-joining analysis (NJA) based on the SNP data using DARwin software (Perrier and Jacquemoud-Collet, 2006) with the simple matching dissimilarity index. This simple matching dissimilarity index was also used to infer the intra- and inter-taxa average differentiation.

Population structure was inferred using the Structure program version 2.3.4 (Pritchard Lab, 2014), which implemented a model-based clustering method using genotype data (Falush *et al.*, 2003; Pritchard *et al.*, 2000). No *a priori* population structure was defined. The linkage model option was used, with allele frequencies correlated and computed probability of the data for  $K$  estimating. Analyses were made with  $K$  value (number of subpopulations) varying from 1–10. The statistics used to select the correct  $K$  value were as in Evanno *et al.* (2005): the mean likelihood,  $L(K)$ ; the mean difference between successive likelihood values of  $K$ ,  $L'(K)$ ; the absolute value of this difference,  $|L''(K)|$ ; and  $\Delta K$ , which is the mean of the absolute values of  $L''(K)$  divided by the standard deviation of  $L(K)$ . The likelihood distribution  $L(K)$  and  $\Delta K$  were the main values used to choose the optimal  $K$  value of the population. Ten runs of Structure were performed with 50,000 steps of burning followed by 50,000 Monte Carlo Markov Chain (MCMC) repetitions. For the better  $K$  value, the ten independent Structure run cluster outputs were permuted and aligned, and average frequency and



standard error of the contribution of each basic population was estimated. The relative genetic distance between successive markers was directly obtained from the reference clementine genetic map (Ollitrault *et al.*, 2012a) or inferred from the physical position of the markers using the curves relating genetic map position to physical location (Wu *et al.*, 2014).

## Results

### 454 SNP mining and genotype calling

Fluidigm amplification followed by 454 sequencing produced 295,169 useful reads. The reads were classified according to their MID, and then Titanium sequences and MID sequences were removed using 454 software tools. All reads were attributed to one of the 1,482 ( $57 \times 26$ ) amplicons according to the fragment gene sequence. Forty-six gene fragment/variety amplicons did not have a corresponding read (3.1%), and eight had insufficient read numbers for genotype calling. The average of reads per amplicon was 205.55 (for 1,436 amplicons with reads). However, the distribution of the number of reads per amplicon was highly heterogeneous (Figure II.1), and 414 amplicons (27.94% of total gene fragments/varieties) had fewer than 50 reads.

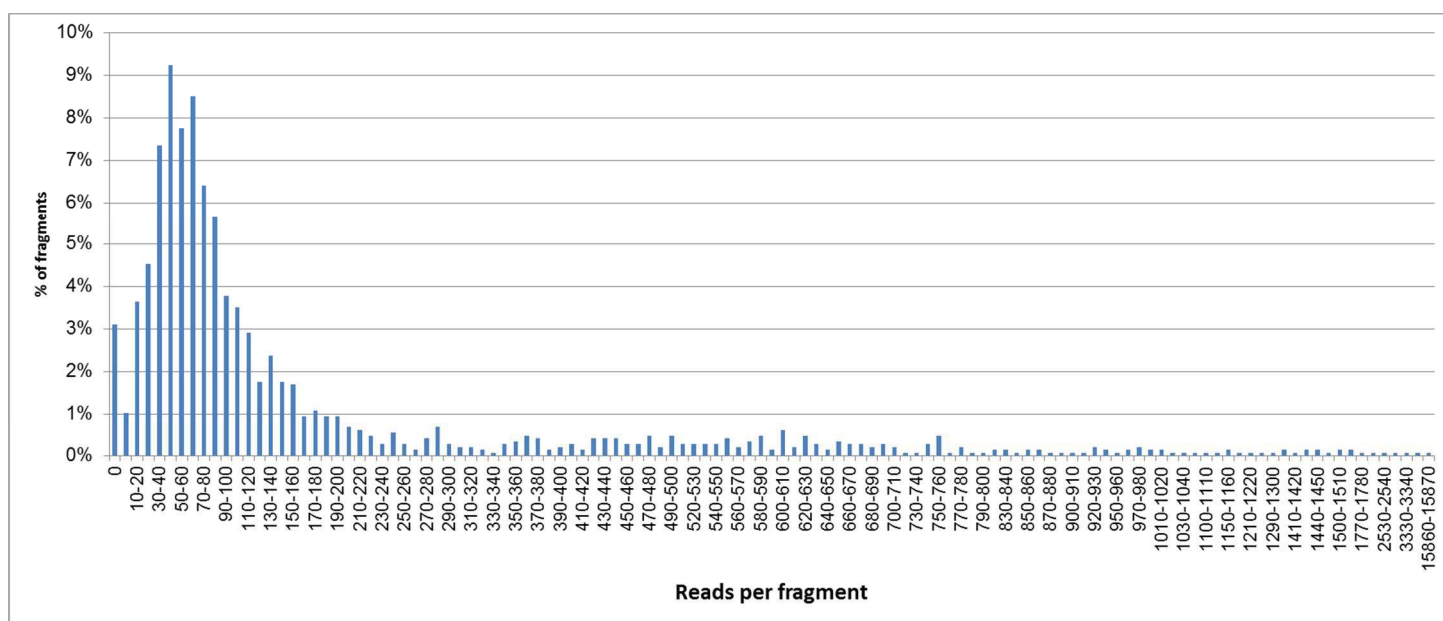


Figure II.1. Distribution of read numbers per fragment.

For *C. micrantha* (represented by a single accession), gene fragments with missing data were Sanger sequenced to complete the data set.

A total of 1,053 SNPs were identified over 28,507 bp readable sequence for 57 gene fragments (Table II.1). SNP genetic diversity parameters (Table II.2 and Annex III, Table S.II.5) were calculated for each SNP position. The SNP/Kb rate varied between gene fragments (Annex III, Table S.II.6, range: 11.19 [fragment C2P25] to 93.31 [fragment C7P8]) and chromosomes (Table II.1; range: 30.48–50.10 SNP/kb in chromosomes 1 and 7, respectively). No significant differences were found between chromosomes for Ho, He, and Fw, and respective average parameter values were 0.064, 0.230, and 0.544.

Table II.1. Average SNP diversity by chromosome

	Sequence size (bp)	Number of fragments per chromosome	SNP / chromosome	SNP / kb	Ho	He	Fw
Chr 1	3,084	6	94	30.48	0.049 ± 0.021	0.229 ± 0.034	0.582 ± 0.093
Chr 2	3,516	7	111	31.57	0.088 ± 0.026	0.231 ± 0.030	0.455 ± 0.080
Chr 3	2,543	5	79	31.06	0.055 ± 0.020	0.249 ± 0.040	0.594 ± 0.093
Chr 4	3,839	8	152	39.59	0.044 ± 0.013	0.232 ± 0.028	0.588 ± 0.072
Chr 5	2,475	5	71	28.69	0.059 ± 0.017	0.215 ± 0.037	0.501 ± 0.107
Chr 6	3,408	7	105	30.81	0.052 ± 0.015	0.221 ± 0.031	0.577 ± 0.087
Chr 7	2,475	5	124	50.10	0.073 ± 0.024	0.211 ± 0.027	0.420 ± 0.082
Chr 8	4,142	8	187	45.15	0.071 ± 0.015	0.218 ± 0.022	0.546 ± 0.062
Chr 9	3,025	6	130	42.98	0.074 ± 0.023	0.265 ± 0.031	0.623 ± 0.066
Total	28,507	57	1,053	36.94	0.064 ± 0.009	0.230 ± 0.010	0.544 ± 0.027

The values of the fixation index (Fw; 0.54) and average  $F_{ST}$  value (0.45) over the whole population suggest an important structuration of the analysed varietal sample, as confirmed by the NJA (Figure II.2).

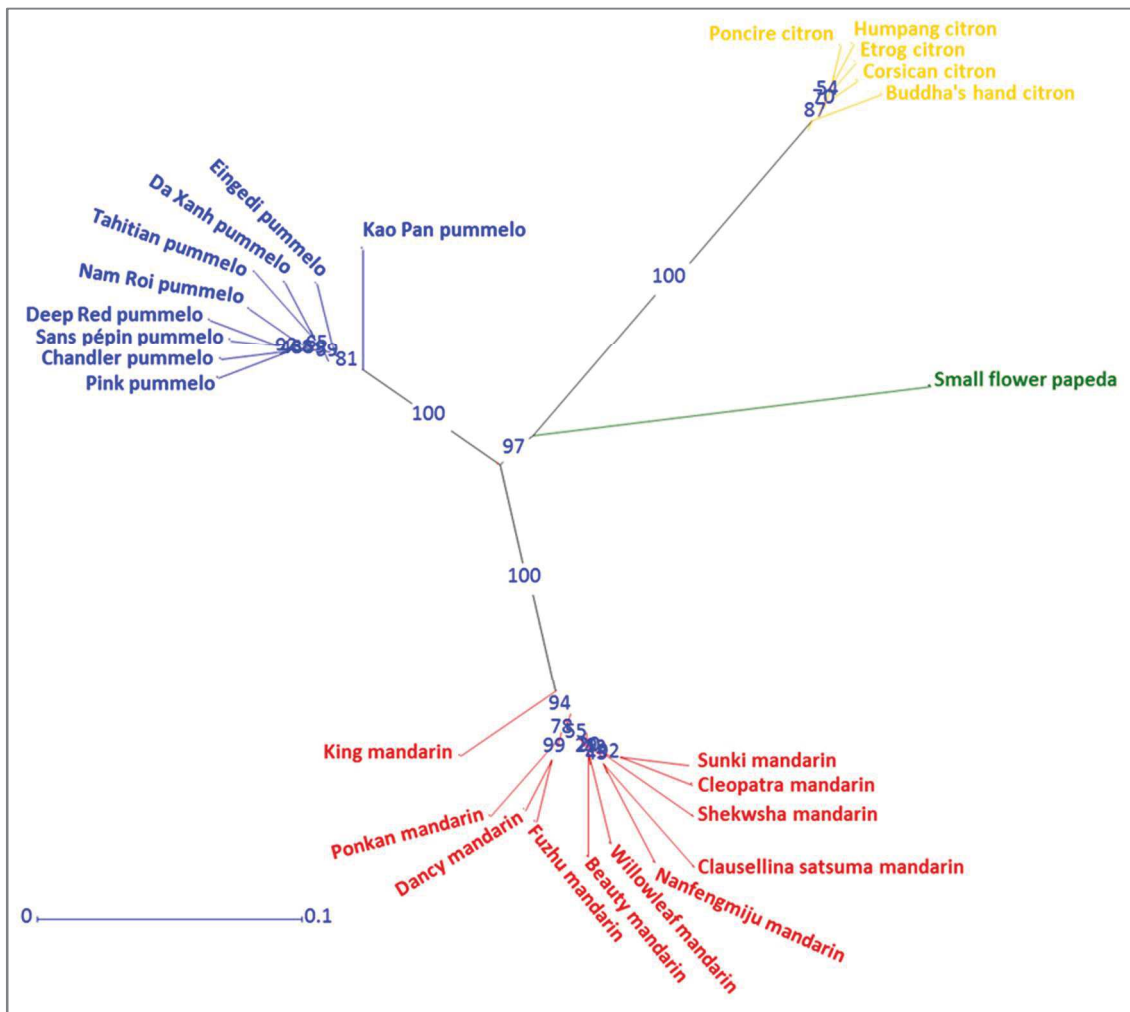


Figure II.2. Neighbour-joining analysis (NJA) of the 1,053 SNPs in all gene fragments.

Intra-horticultural group  $F_w$  values ranged from 0.128 in mandarins to 0.171 in pummelos. Mandarin displayed the higher intragroup diversity (15.61 SNP/Kb and  $H_e = 0.102 \pm 0.006$ ). Citron displayed low heterozygosity ( $H_o = 0.04 \pm 0.02$ ) and polymorphism ( $H_e = 0.04 \pm 0.003$ ) compared with mandarins and pummelos. Only one representative of *C. micrantha* was available, and its observed heterozygosity value (0.07) lay between the pummelo and mandarin values (Table II.2).

Table II.2. Diversity of basic citrus taxa

	SNPs	SNPs / kb	$H_o$	$H_e$	$F_w$	$F_{st}$
Pummelos (pop = 9)	297	10.42	$0.056 \pm 0.008$	$0.072 \pm 0.004$	$0.171 \pm 0.052$	
Citrons (pop = 5)	132	4.63	$0.038 \pm 0.018$	$0.043 \pm 0.003$	$0.146 \pm 0.025$	
<i>C. micrantha</i> (pop = 1)	71	2.49	0.068	$0.034 \pm 0.008$	-1.000	
Mandarins (pop = 11)	445	15.61	$0.082 \pm 0.011$	$0.102 \pm 0.006$	$0.128 \pm 0.012$	
Total	1,053	36.94	$0.064 \pm 0.009$	$0.230 \pm 0.010$	$0.544 \pm 0.027$	$0.445 \pm 0.020$

The analysis of the average number of SNPs/kb between two varieties within and between the four supposed basic taxa revealed values of 1.62–3.49 SNPs/kb within groups and 9.96–13.19 SNPs/kb at the inter-group level (Table II.3).

Table II.3. Intra- and inter-horticultural group dissimilarities (average SNP/kb between two varieties)

	Mandarins	Pummelos	Citrons
Mandarins	3.49*		
Pummelos	9.96	2.61*	
Citrons	13.19	11.54	1.62*
<i>C. micrantha</i>	11.82	9.96	12.49

\*Average number of SNP/kb at the intra-horticultural group level.

### Determination of species-diagnostic SNPs

An initial analysis of taxa differentiation at each SNP position was performed directly from genotype calls from the representative of each of the four basic taxa. Diagnostic SNP assignment was based on the  $G_{ST}$  parameter as described in the Materials and Methods. The distribution of the highest  $G_{ST}$  value of each SNP for the four basic taxa (Figure II.3; and Annex III, Table S.II.5) showed that 47% of the highest SNP  $G_{ST}$  values were  $>0.5$ . Moreover, *C. medica* (112 diagnostic SNPs) and *C. micrantha* (91) displayed many more highly diagnostic SNPs ( $G_{ST} > 0.9$ ) than *C. reticulata* (42) or *C. maxima* (26).

For  $G_{ST}$  values  $<0.5$ , a SNP was considered as specific for a taxon when it was polymorphic within the taxon but fixed for the same allele in the other three taxa. For example, the 55 *C. micrantha* SNPs with  $G_{ST}$  values of 0.3–0.4 corresponded with heterozygous SNP in *C. micrantha* fixed for a same allele in mandarin, pummelo, and citron. No specific allele was observed for 92 SNPs with the highest taxa  $G_{ST}$  value  $\leq 0.5$  (similar patterns of diversity in at least two basic taxa, denoted “N.D.” in Figure II.3).

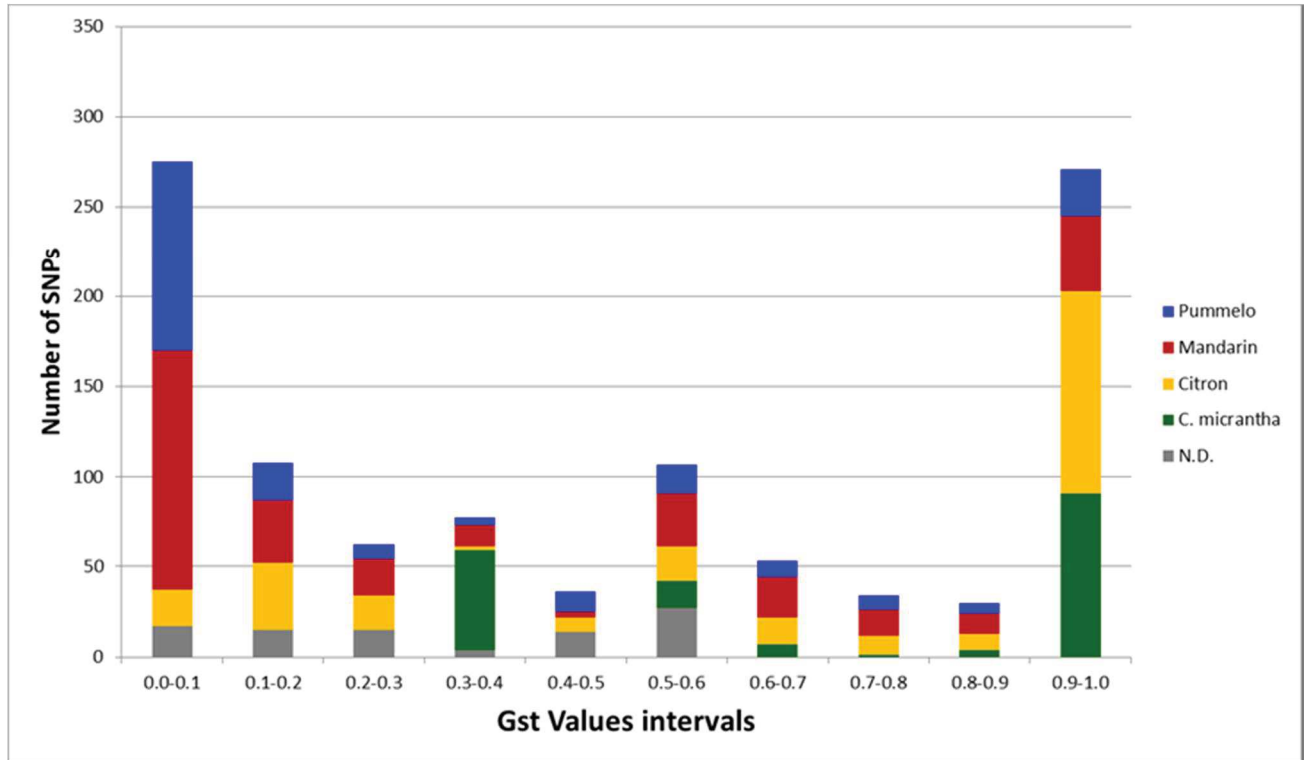


Figure II.3. Distribution of the highest SNP  $G_{ST}$  values for the four basic horticultural groups (454 data from cultivars); N.D., non-diagnostic.

Previous WGS (Wu *et al.*, 2014) and amplicon sequencing (Curk *et al.*, accepted; Garcia-Lor *et al.*, 2013b) studies showed that several modern varieties generally considered as representative of citrus basic taxa were introgressed by other species. This was particularly notable for *C. maxima* in mandarins and *C. reticulata* in pummelos. This observation should explain the low proportion of strong diagnostic SNPs for pummelos and mandarins compared with citrons and *C. micrantha*. Therefore, given that our objective was to identify diagnostic SNPs in the basic taxa, it was essential to identify such introgressions in the modern varieties of mandarin, pummelo, citron, and *C. micrantha* in order to better estimate the allelic frequencies in the basic taxa. For this purpose, information provided by Factorial Analysis (from dissimilarity values between each pair of accessions) and the estimation of heterozygosity of each genotype was combined for each gene fragment. Genotypes having interspecific phylogeny for the considered fragment were expected to be in intermediate positions between basic taxa clusters and to display much higher heterozygosity than genotypes without interspecific heterozygosity. The C2P27 fragment is provided as an example in Figure II.4. For this fragment, four mandarins (‘King’, ‘Dancy’, ‘Ponkan’, and ‘Fuzhu’) displayed clear interspecific heterozygosity. Indeed, while clusters of the basic taxa displayed low heterozygosity ( $<0.03$ ), these four mandarins had high

heterozygosity (average = 0.53), and their intermediate positions between the *C. reticulata* and *C. maxima* clusters resulted from heterozygosity for SNPs differentially fixed in the *C. maxima* and *C. reticulata* clusters.

It should be noted that *C. micrantha*, which had a relatively central position in the 1/2 plan, had very low heterozygosity and was totally differentiated from the three other basic taxa in the third axis.

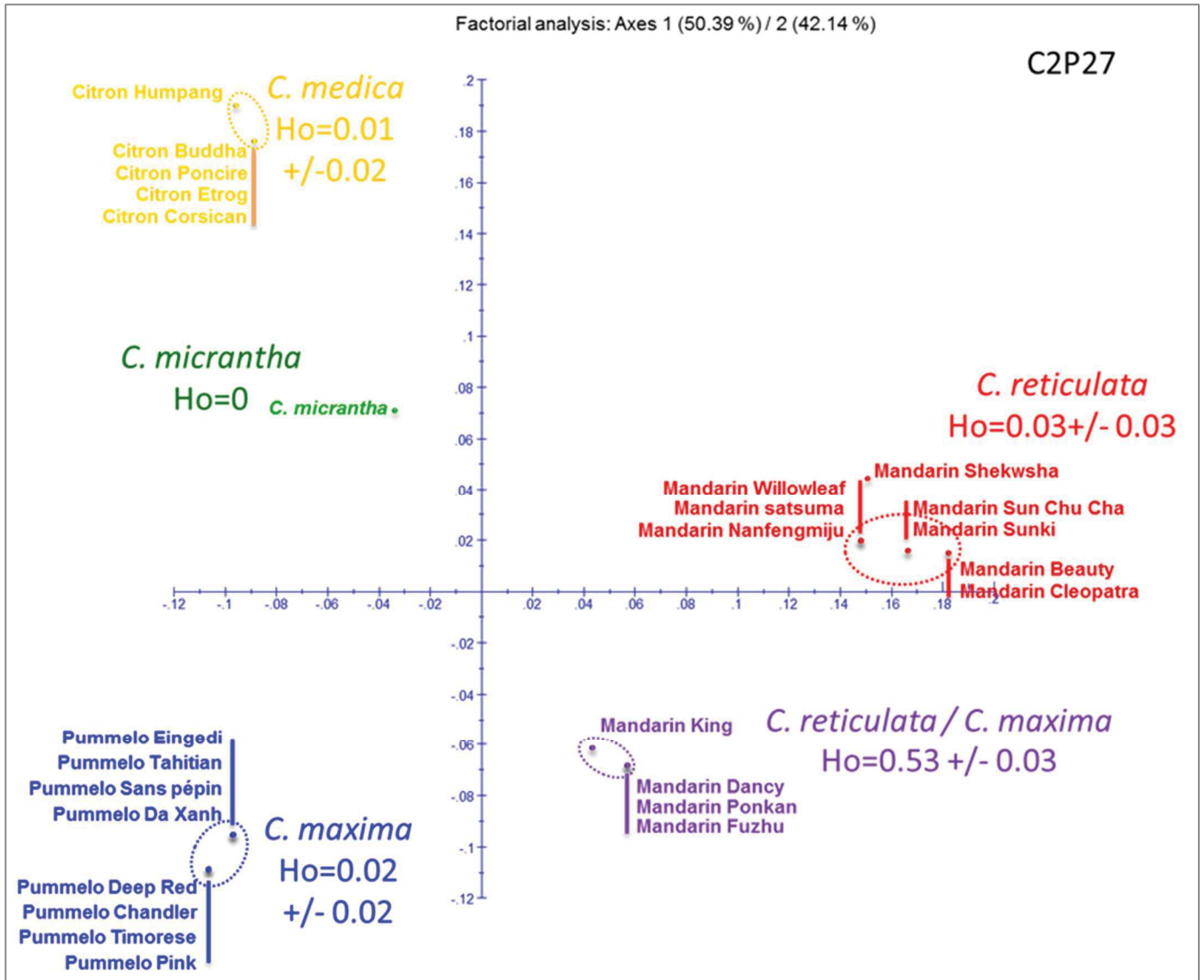


Figure II.4. *C. reticulata*/*C. maxima* heterozygosity of four mandarins for the C2P27 gene fragment.

Basic taxa allelic frequencies, population diversity parameters (Table II.4) and  $G_{ST}$  parameters were then re-estimated, with only gene fragments and varieties considered that did not exhibit interspecific heterozygosity.

Table II.4. Diversity of citrus basic taxa after removal of genotypes with interspecific heterozygosity at the gene fragment level

		SNP/kb	Ho	He	Fw	Fst
<i>C. maxima</i>	211	7.40	0.051 ± 0.009	0.056 ± 0.003	0.048 ± 0.006	
<i>C. medica</i>	132	4.63	0.039 ± 0.018	0.045 ± 0.003	0.146 ± 0.025	
<i>C. micrantha</i>	71	2.49	0.071	0.035 ± 0.002	-1.000	
<i>C. reticulata</i>	282	9.89	0.059 ± 0.018	0.075 ± 0.005	0.147 ± 0.017	
<b>Total</b>	<b>1,011</b>	<b>35.46</b>	<b>0.053 ± 0.006</b>	<b>0.238 ± 0.010</b>	<b>0.588 ± 0.028</b>	<b>0.903 ± 0.009</b>

The new whole population parameter values (diversity index (He) = 0.24, and fixation index (Fw) = 0.59) were similar to the previous one, suggesting that the global structuration of the analysed varietal samples was conserved. Mandarin polymorphism decreased from 0.102 ± 0.006 to 0.075 ± 0.005. Pummelo polymorphism decreased from 0.072 ± 0.004 to 0.056 ± 0.003.

The distribution of the highest  $G_{ST}$  value of each SNP for the four basic taxa (Figure II.5) showed that 423 of highest SNP  $G_{ST}$  values were >0.6. *C. medica* (113 SNPs) and *C. micrantha* (92 SNPs) continued to have a larger number of diagnostic SNPs ( $G_{ST}$  >0.9) than *C. reticulata* or *C. maxima*, but the number of strong diagnostic markers for the latter two taxa improved, respectively reaching 72 for *C. reticulata* and 43 for *C. maxima* (Figure II.5). Ninety-five SNPs with a  $G_{ST}$  value ≤0.5 did not exhibit any specificity for one of the basic taxa (N.D., Figure II.5) and did not meet our diagnostic objectives.

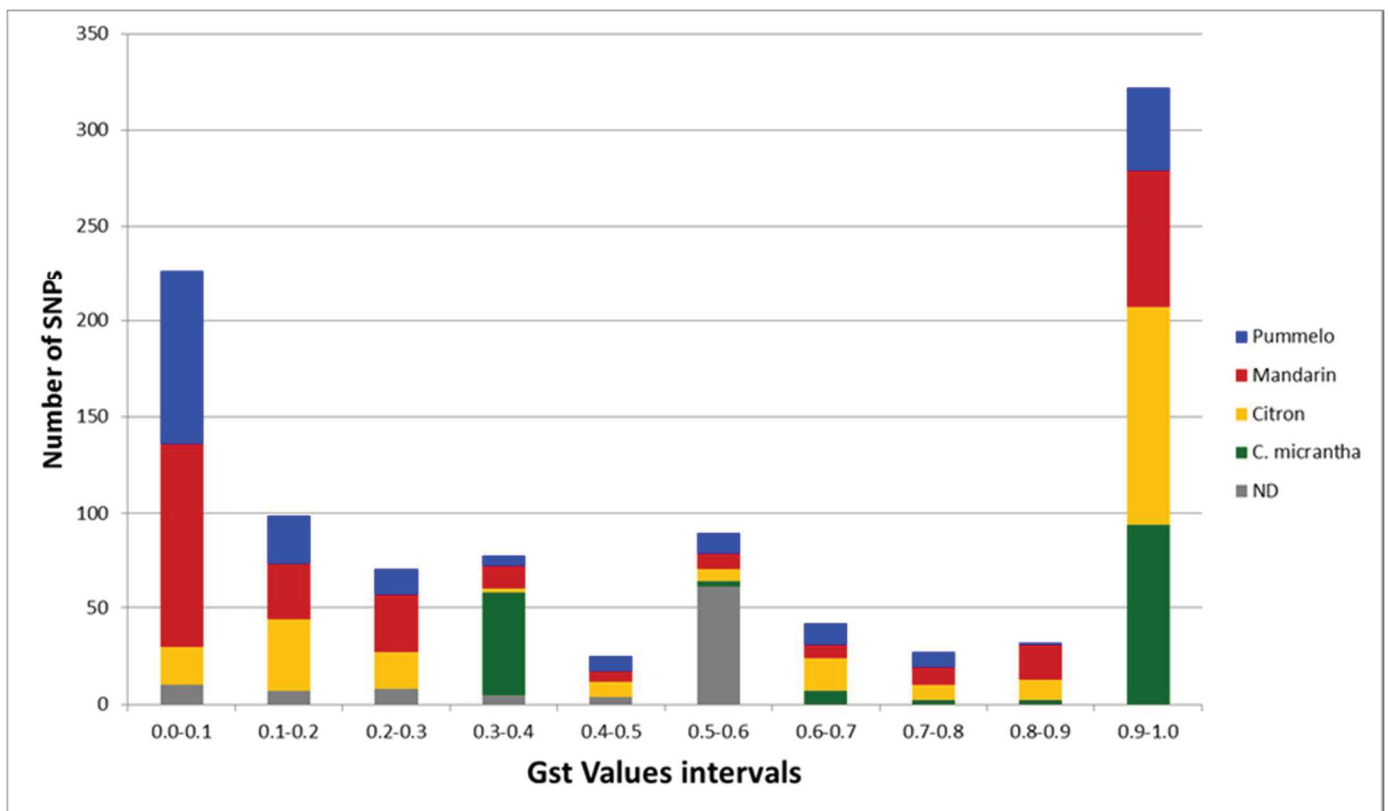


Figure II.5. Distribution of SNP  $G_{ST}$  values for the four basic taxa from estimated ancestral taxa allelic frequencies after removal of genotypes with interspecific heterozygosity at the gene fragment level.

## KASPar marker development

### *SNP marker development*

SNPs were selected from those with high differentiation values between basic taxa and those displaying intraspecific variability within each basic taxon (SNPs with fixed allele in three taxa and displaying polymorphism in the fourth taxon). To limit the risk of PCR drift between alleles, SNPs were rejected for KASPar marker development if further SNPs or indels were in close proximity. Eighty-five SNP markers were developed from the 454 SNP mining data. Amplification failed for four of these SNP markers, and eight further SNP regions produced inconsistent results: these twelve markers were thus discarded. The remaining 73 markers were successfully amplified, and according to the 454 genotyping data, 63 of these SNPs were diagnostic for one of the four taxa (12, 9, 20, and 22 markers for *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*, respectively). Ten markers characterised intraspecific polymorphisms (two, five, and three markers for *C. reticulata*, *C. maxima*, and *C. medica*, respectively).

### *Conformity between 454 and KASPar genotype calling*

Twenty four varieties that had been examined by 454 sequencing were genotyped using the 73 KASPar markers (Annex III, Table S.II.4). Of the 73 × 24 genotyping points, 3.6% displayed genotype calling discrepancies between the two methods. The highest discrepancy rate (27%, Figure II.6) was found for the 8p2427684 marker corresponding to the C8P4 gene fragment; however, this region was notable for poor 454 sequencing quality, and data for nine varieties were missing. KASPar data were validated using two technical replicates, and the discrepancies between the two methods were therefore more likely due to 454 genotype calling errors.

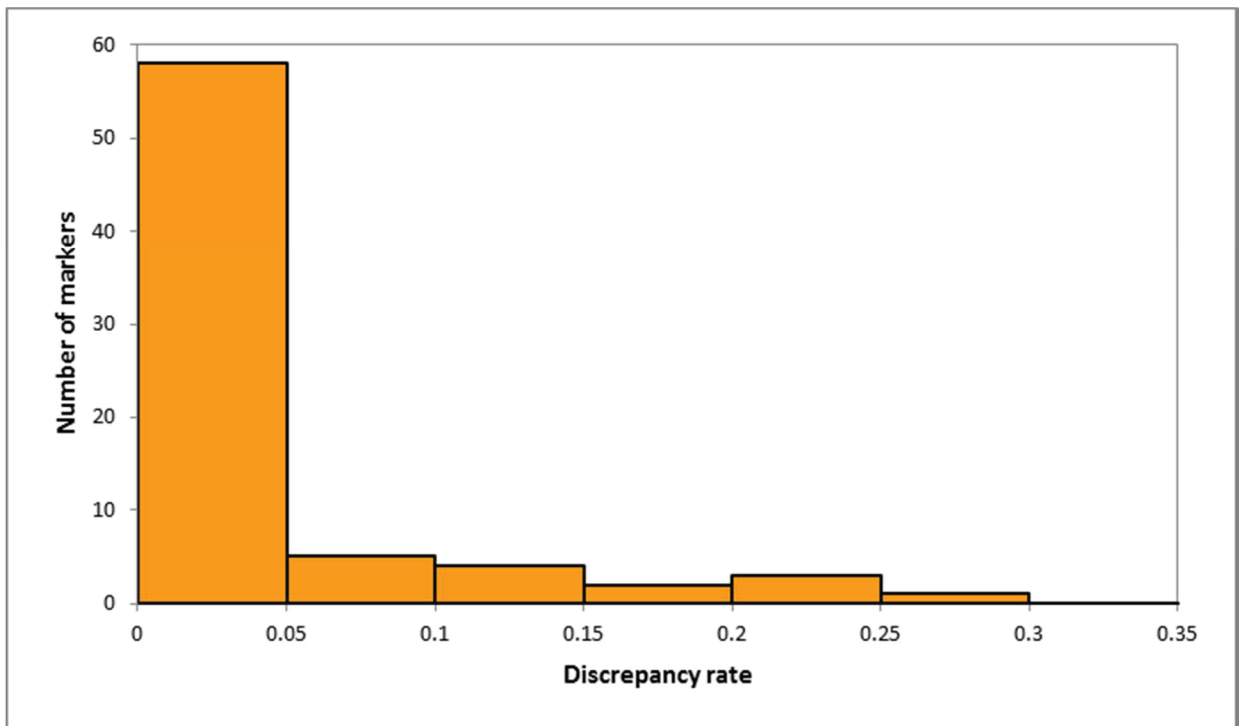


Figure II.6. Distribution of the markers displaying discrepancies between 454 and KASPar genotype calling (x axis; discrepancy rate; y axis number of markers).

### *Transferability of KASPar markers to related genera in the Aurantioideae subfamily*

Utility of the KASPar markers across the Aurantioideae subfamily was assessed using 14 accessions grouped in increasingly distant taxonomic entities according to Swingle and Reece (1967). Missing data and polymorphism information for these different groups are provided in Table II.5.

Table II.5. SNP marker screening in different citrus species and subtribes of the Aurantioideae subfamily

	N	MD	PL	Ho
<i>Citrus</i> genus	70	1.0	73	0.05
True citrus *	8	4.3	18	0.02
Near Citrus	2	9.6	7	0.05
Primitive Citrus	1	12.3	2	0.03
Triphasilineae	1	24.7	0	0.00
Clauseniae	2	38.4	3	0.03

N: sample size; MD: missing data (%); PL: number of polymorphic loci; Ho: observed heterozygosity; \*True citrus without *Citrus* genus.

The missing data rate was very low in *Citrus* (1.0%) and low in the “true citrus fruit trees” group (4.3%, excluding the *Citrus* genus). The missing data rate increased, respectively, to 9.6% and 12.3% in the “near citrus” and “primitive citrus” groups of the Citrinae subtribe. The missing data rate reached a level of 24.7% for the other subtribe of the Citreae tribe (*Triphasilineae*) and 38.4% for the two representatives of the *Clauseniae* tribe. These results indicate an increasing loss of transferability with increasing taxonomic distance. As expected due to the discovery panel, the *Citrus* genus was the most polymorphic (73 polymorphic loci), followed by the “true citrus fruit trees” group without the *Citrus* genus (18 polymorphic loci). Diversity within the other taxa was considerably lower. For some markers (46/73), all accessions except those in the *Citrus* genus displayed the same homozygous genotypes for the non-diagnostic allele (allele found to be diagnostic for one of the basic taxa using 454 data). For these 46 markers, the alternative allele fixed in all citrus and relative taxa, with the exception of one of the four basic citrus taxa, was clearly the ancestral allele. The alternative allele therefore resulted from a mutation that occurred during the separation of the differentiated basic taxa.

### **Admixture structure of modern citrus varieties**

In addition to the 73 new KASPar SNP markers, we selected 32 of our previously developed SNP markers (Garcia-Lor *et al.*, 2013a; Ollitrault *et al.*, 2012b) to complete the analysis of admixture structure of modern citrus varieties. This selection was made based on the available differentiation parameter data between the basic taxa. Respectively, 30, 23, 30, and 22 of the 105 markers were selected as diagnostic or to represent specific alleles of *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*.



The distribution and allele specificity of the 105 KASPar SNP markers used for Citrus admixture analysis is provided in Figure II.7. Specific markers for all basic taxa were present on each chromosome. However, some large lacunas without markers were still present, particularly in chromosomes 3 and 5.

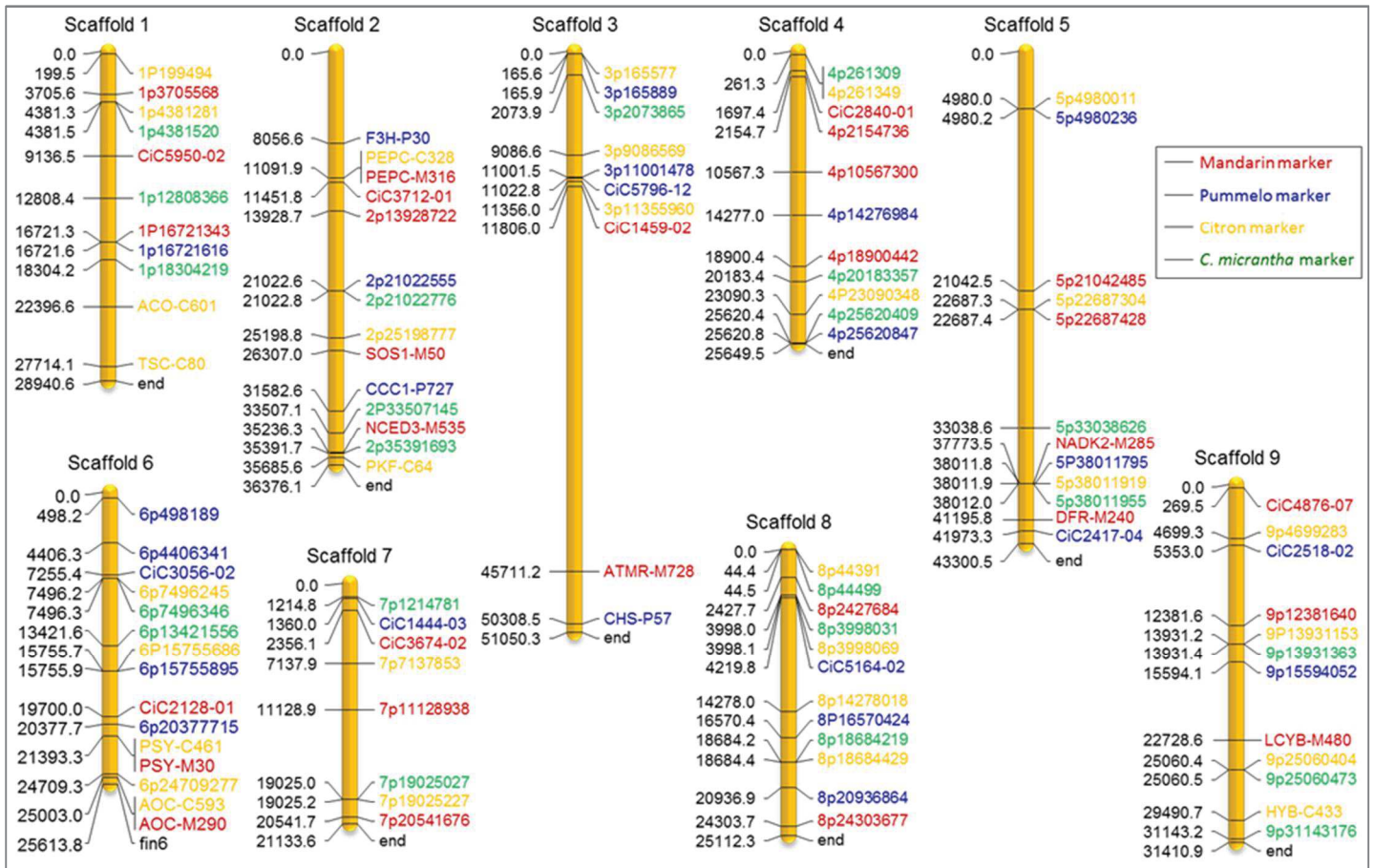


Figure II.7. Location of all KASPar SNP markers used (105 markers; position in kb).

Diversity and structuration parameters were calculated using 33 accessions representative of the four basic horticultural groups: seventeen mandarins, eight pummelos, six citrons, and two *C. micrantha*. With the exception of *C. micrantha*, which was totally homozygous and displayed no polymorphisms between the two accessions, the intraspecific parameters  $H_o$  and  $H_e$  were not significantly different between the whole set of SNPs identified by 454 and the results of the KASPar analysis (Table II.6).

Table II.6. Diversity of the four horticultural groups for 105 species-diagnostic SNP markers

	N	$H_o$	$H_e$	$F_w$	$F_{ST}$
Pummelos	8	0.035 ± 0.010	0.034 ± 0.020	-0.052 ± 0.146	
Citrons	6	0.014 ± 0.008	0.016 ± 0.015	0.058 ± 0.177	
<i>C. micrantha</i>	2	0.000	0.000	N	
Mandarins	17	0.073 ± 0.033	0.068 ± 0.020	-0.047 ± 0.060	
Total	33	0.049 ± 0.020	0.326 ± 0.027	0.872 ± 0.034	0.903 ± 0.028

N: sample size;  $H_o$ : observed heterozygosity;  $H_e$ : Nei diversity index;  $F_w$ : Wright fixation index;  $F_{ST}$  inter-basis taxa differentiation index.

However, the structuration of the population was higher with the selected KASPar markers, with average  $F_w$  and  $F_{ST}$  values of  $0.87 \pm 0.03$  and  $0.90 \pm 0.03$ , respectively. Moreover, no significant deviation to panmixis ( $H_o$  and  $H_e$  with similar values and confidence interval of  $F_w$  including the null value) was observed within the three polymorphic horticultural groups (mandarin, pummelo, and citron). Diversity parameter ( $H_o$  and  $H_e$ ) and structuration parameter ( $F_w$ ,  $F_{ST}$ , and specific  $G_{ST}$ ) values for each marker are in Annex III, Table S.II.7. The distribution of the  $G_{ST}$  parameters (Figure II.8) confirmed the efficiency of the applied selection with 18, 14, 27, and 22 markers with a  $G_{ST} > 0.8$  for the *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha* diagnostic markers, respectively. However, two markers (7p11128938 and CiC2518-02) were found to share similar polymorphisms in mandarins and pummelo and were therefore discarded from the structure and specific allele homozygosity/heterozygosity analyses. We therefore concluded that 29, 22, 30, and 22 of the selected markers displayed specific alleles for *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*, respectively. Nine of these markers, with lower specific  $G_{ST}$  values ( $< 0.6$ ), were fixed for the same allele in three horticultural groups but displayed an intraspecific polymorphism in the fourth group (one, five, and three markers with variant alleles only in mandarin, pummelo, and citron, respectively).

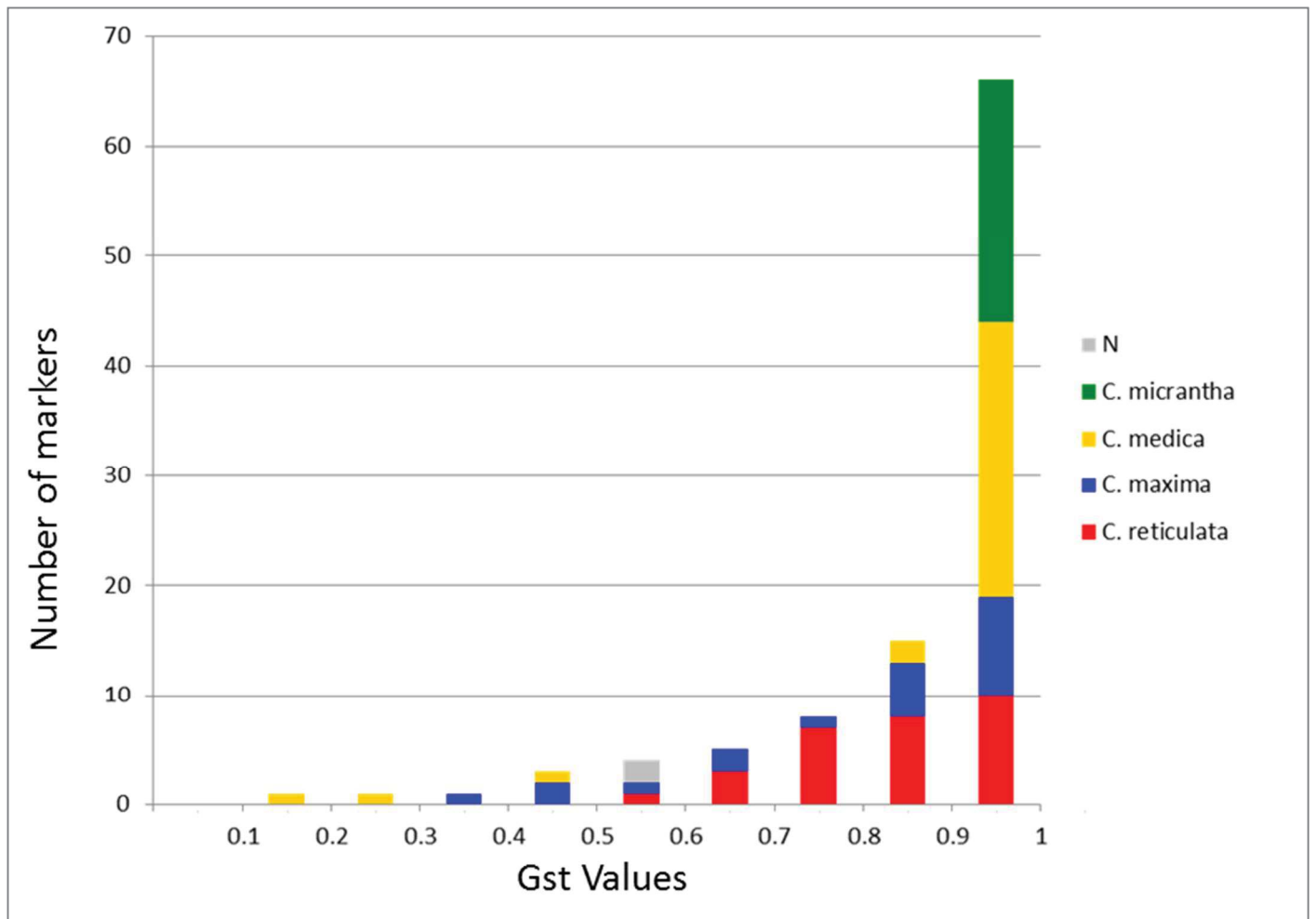


Figure II.8. Distribution of individual diagnostic marker  $G_{ST}$  values for the four basic taxa through consideration of representatives of the four basic horticultural groups as subpopulations.

When secondary species accessions were considered (Table II.7), no intraspecific polymorphisms were found within *C. aurantium* (three accessions), *C. sinensis* (three accessions), or *C. paradisi*, while four different genotypes were observed in *C. aurantifolia* (four accessions) and five in *C. limon* (six accessions). All secondary species displayed high heterozygosity values ( $H_o$ ) compared with the four horticultural groups.  $H_o$  ranged from 0.35 in *C. paradisi* to 0.48 in *C. limon*. The mandarin hybrid, tangor, and tangelo groups displayed increasing average  $H_o$  values (0.18, 0.21, and 0.25, respectively). However, substantial heterozygosity variations were observed between varieties within the tangor group (0.13–0.30), and the differences between the groups were thus not statistically significant.

Table II.7. Diversity of secondary species and modern hybrid varieties assessed using 105 SNP markers

	N	$H_o$	NSG
<i>C. aurantium</i>	3	0.424 ± 0	1
<i>C. aurantifolia</i>	4	0.457 ± 0.020	4
<i>C. limon</i>	6	0.485 ± 0.045	5
<i>C. paradisi</i>	3	0.346 ± 0	1
<i>C. sinensis</i>	3	0.371 ± 0	1
<i>C. amblycarpa</i>	1	0.456	1
<i>C. hystrix</i>	1	0.168	1
Mandarin hybrids	4	0.180 ± 0.021	4
Tangors	8	0.214 ± 0.046	7
Tangelos	3	0.248 ± 0.011	3
Orangelo	1	0.205	1

N: number of accessions by taxa;  $H_o$ : observed heterozygosity; NSG: number of single genotypes.

PCA was performed with the data from the 70 *Citrus* cultivars genotyped with the 105 KASPar SNPs. The population displayed a very strong structuration, with 85.6% of the total diversity encompassed by the three first axes (40.6%, 23.8%, and 21.3%, respectively). The fourth axis supported only 1.7% of the diversity. The first axis mainly differentiated citrons from other species (Figure II.9).

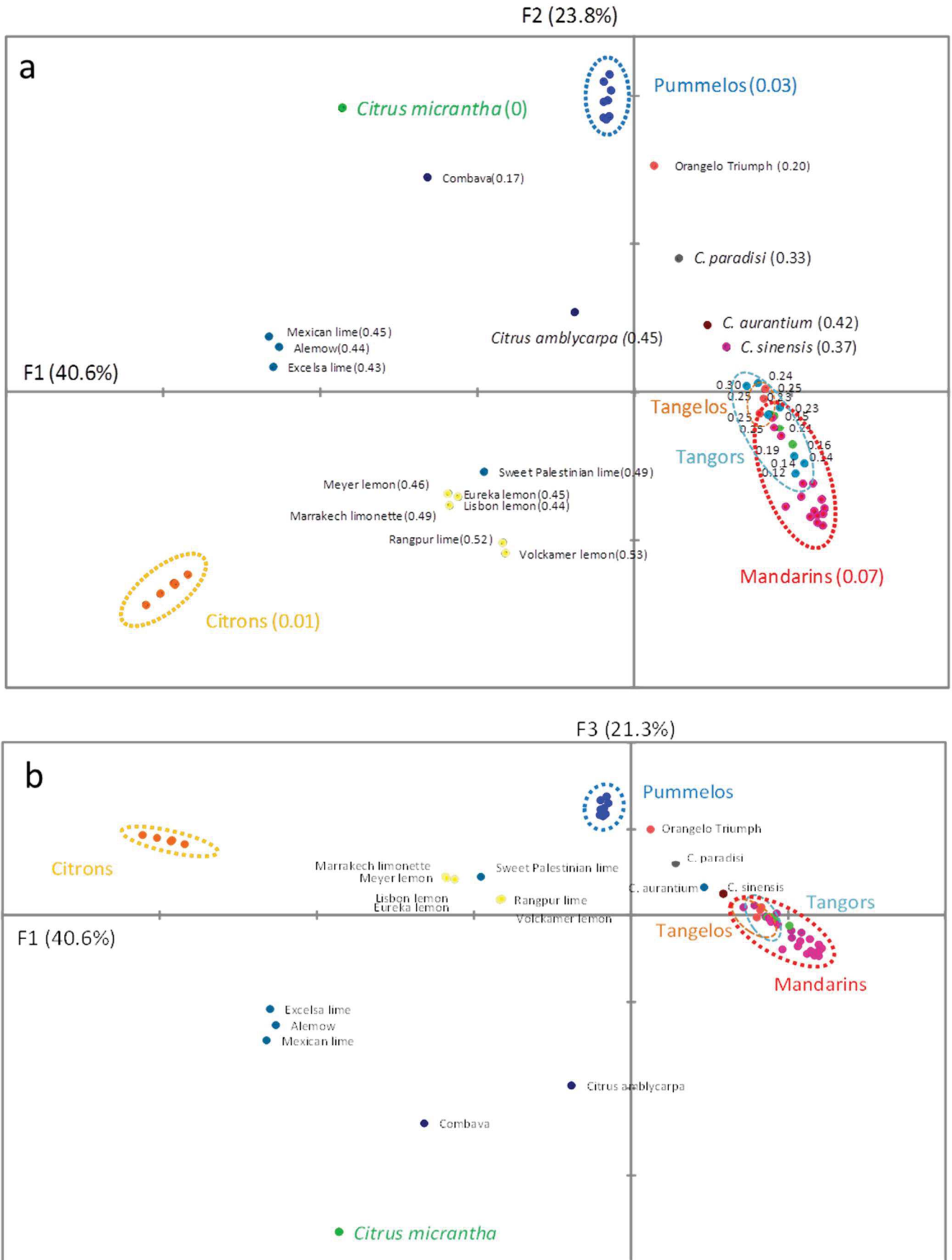


Figure II.9. PCA distribution of 70 *Citrus* cultivars from 105 SNP marker genotyping (first three axes). Individual heterozygosity is provided in parentheses after the variety name in the F1/F2 plan. a. F1/F2; b. F1/F3.

The second axis distinguished *C. maxima* and *C. micrantha* from *C. reticulata* and *C. medica*, while the third axis separated *C. micrantha* for all the other basic taxa. Interestingly, mandarin hybrids, tangors, tangelos, sweet oranges, sour oranges, and grapefruits were distributed along a line between the pummelo and mandarin clusters. Varietal heterozygosity increased from the distal parts of the segment defined by the pummelo cluster (average  $H_o = 0.03$ ) and the mandarin cluster (average  $H_o = 0.07$ ) to the central region ( $H_o = 0.42$  for the three accessions of the *C. aurantium* sour orange group). ‘Mexican’ lime, ‘Alemow’, and ‘Excelsa’ lime were clustered in intermediary positions between the citron cluster and *C. micrantha*, with heterozygosity values of 0.43–0.45. ‘Rangur’ lime and ‘Volkamer’ lemon also displayed high heterozygosity values (0.52 and 0.53, respectively) and were located between the citron and the mandarin clusters. The ‘Palestinian’ sweet lime, ‘Meyer’ lemon, and ‘Eureka’ lemon were slightly displaced, having higher values for the F2 axis and heterozygosity values of 0.44–0.49. *C. amblycarpa* had a similar heterozygosity level but was located between *C. micrantha* and the mandarin cluster.

The relative contributions of the ancestral taxa to the modern cultivars suggested by PCA and average heterozygosity levels were confirmed by admixture analysis using Structure software (Pritchard Lab, 2014) and a deeper analysis of cultivar homozygosity and heterozygosity for the four sets of diagnostic markers (markers with specific alleles for *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*).

Structure analysis with the linkage option was made without *a priori* defined populations, and ten replicate runs were performed for varied K values of 1–10. Analysis of delta K indicated that optimal results were obtained with  $K = 4$  (Annex III, Table S.II.8). Minimal variability in estimated frequencies was observed between the ten runs for  $K = 4$ . The average values of the ten runs (Annex III, Table S.II.9) are shown in Figure II.10.

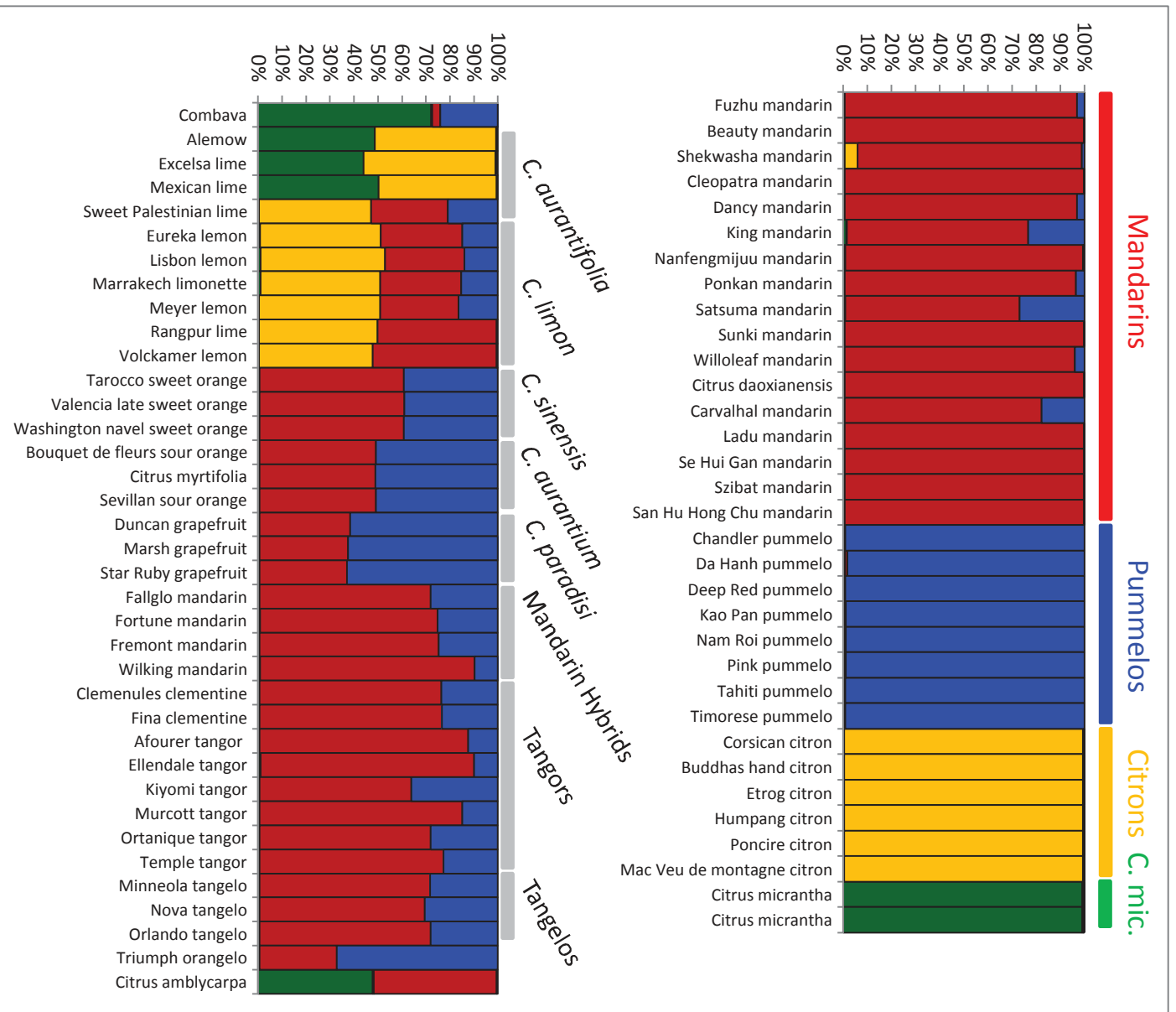


Figure II.10. Structure analysis of 70 Citrus cultivars from genotyping data and 103 SNP markers. Red, blue, yellow, and green correspond to the inferred contributions of *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*, respectively.

Among the four basic horticultural groups, no evidence of interspecific introgression was found for the analysed accessions of pummelos, citrons, and *C. micrantha* (Figure II.10 and II.11). Interspecific introgressions of *C. maxima* were observed in several mandarins, with higher values noted in Satsuma (27%), 'King' (23%), and 'Carvalho' mandarins (18%). Both structure and diagnostic marker analyses revealed a *C. medica* introgression in the 'Shekwasha' mandarin. No significant introgression was found for nine mandarins ('Beauty', 'Cleopatra', 'Nanfengmiju', 'Sunki', *C. daoxianensis*, 'Ladu', 'Se Hui Gan', 'Szibat', and 'San Hu Hong Chu') with structure analysis. However, the diagnostic SNP analysis showed that five of these were heterozygous for one of the *C. maxima* diagnostic markers ('Beauty', 'Nanfengmiju', 'Ladu', 'Se Hui Gan', and 'San Hu Hong Chu').

Three of the four *C. aurantifolia* accessions ('Alemow', 'Excelsa', and 'Mexican' limes) displayed very similar phylogenomic patterns, with close to 50% contributions from *C. medica* and *C. micrantha* (Figure II.10) as confirmed by the heterozygosity of *C. micrantha* and *C. medica* diagnostic markers. The fourth *C. aurantifolia* accession ('Palestinian' sweet lime) had a distinct pattern and appeared to have a similar phylogenetic structure to four of the six accessions of *C. limon* ('Meyer', 'Eureka', and 'Lisbon' lemons, and 'Marrakech' limonette). These contained close to 50% *C. medica* and displayed a three ancestral taxa admixture pattern (*C. medica*, *C. reticulata*, and *C. maxima*) with a greater contribution from *C. reticulata* (32–34%) than *C. maxima* (15–20% in 'Palestinian' sweet lime). Diagnostic alleles of *C. medica*, *C. maxima*, and *C. reticulata* were found systematically in heterozygosity in these five accessions. The two last *C. limon* accessions ('Rangpur' lime and 'Volkamer' lemon) had a very similar phylogenetic structure with around 50% from both *C. reticulata* and *C. medica* (Figure II.10). These accessions displayed a complete heterozygosity for their *C. reticulata* and *C. medica* diagnostic alleles and had no diagnostic alleles for the other taxa (Figure II.11).

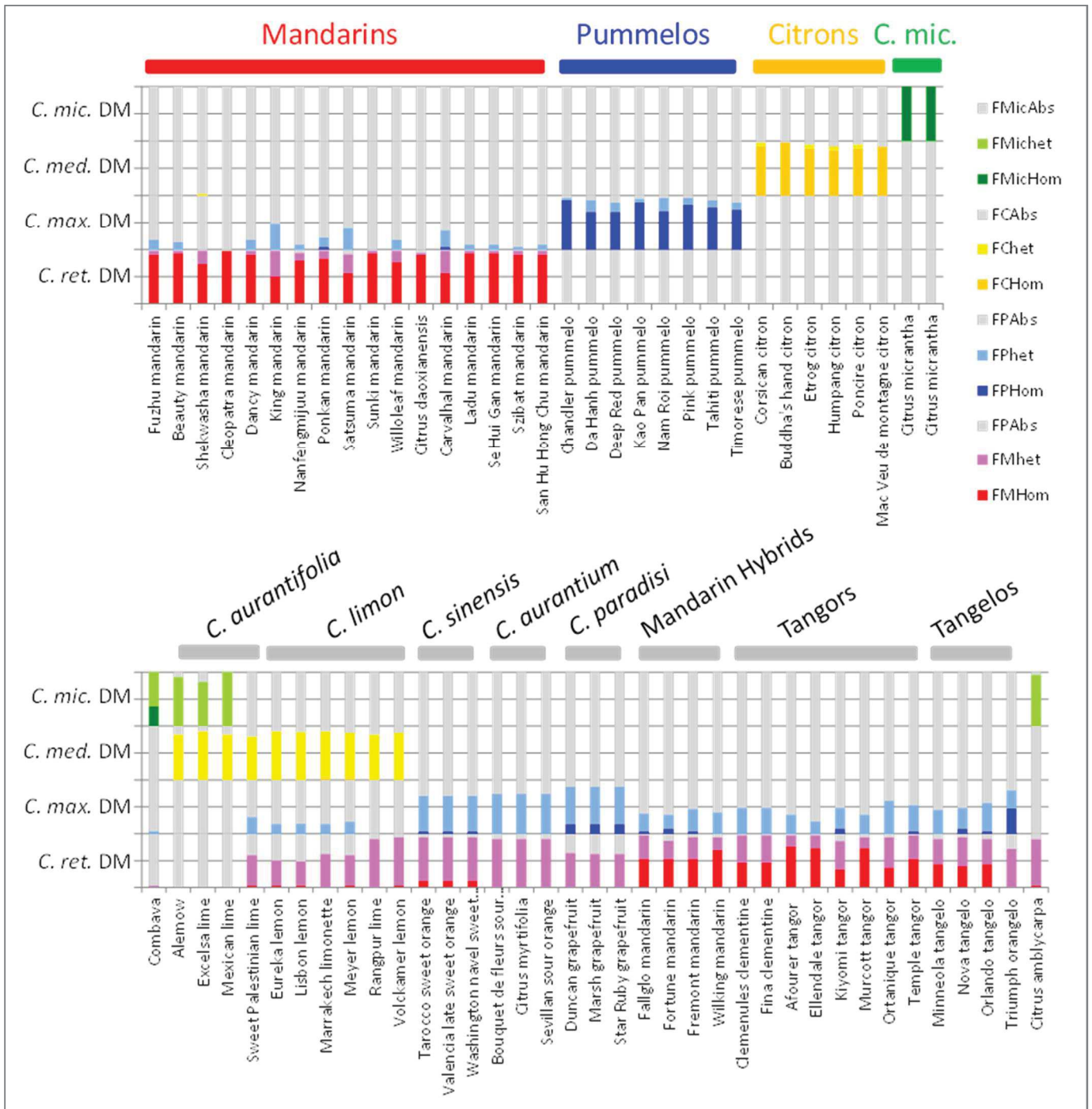


Figure II.11. Frequency of specific alleles of the four basic taxa in homozygosity and heterozygosity for 70 *Citrus* accessions analysed with 103 diagnostic markers. *Taxa DM*: diagnostic markers for the considered taxa; FMHom, FMHet, FMAbs, FPHom, FPHet, FPAbs, FCHom, FCHet, FCABs, FMicHom, FMicHet, and FMicAbs are, respectively, the frequency of homozygous, heterozygous, and absent specific alleles for *C. reticulata*, *C. maxima*, *C. medica*, and *C. micrantha*.



All accessions of *C. sinensis*, *C. aurantium*, *C. paradisi*, mandarin hybrids, tangors, tangelos, and orangelo had a two ancestral taxa admixture structure (*C. reticulata* and *C. maxima*) with variable contributions. The three *C. aurantium* accessions displayed around 50% of the contribution to each taxon and appeared totally heterozygous for their *C. reticulata* and *C. maxima* diagnostic alleles. With 60% and 40% of *C. reticulata* and *C. maxima* contributions, respectively, *C. sinensis* had a more complex genomic structure and part of the diagnostic alleles for *C. reticulata* and for *C. maxima*, were in homozygosity. *C. paradisi* accessions had higher contributions of *C. maxima* (63%) than *C. reticulata* (37%) and displayed homozygous diagnostic alleles of *C. maxima* in addition to heterozygous alleles from *C. maxima* and *C. reticulata*. All mandarin hybrids, tangors, and tangelos had a greater contribution (from 63% for 'Kiyomi' tangor to 89% for 'Wilking' mandarin) from *C. reticulata* than *C. maxima*. All displayed homozygous diagnostic alleles of mandarin in relatively high proportion (34% in 'Kiyomi' to 77% in "Nadorcott") as well as heterozygous diagnostic alleles of *C. reticulata* and *C. maxima*. In addition, seven accessions presented one or two *C. maxima* diagnostic alleles in homozygosity. The 'Triumph' orangelo displayed a similar pattern to *C. paradisi* but with a higher proportion of homozygosity for *C. maxima* diagnostic alleles (29% of the total number of *C. maxima* diagnostic alleles).

*C. amblycarpa* displayed around 50% contributions from *C. reticulata* and *C. micrantha* and appeared totally heterozygous for its *C. reticulata* and *C. maxima* diagnostic alleles.

The 'Combava' (*C. hystrix*) genome structure remained unclear. The structure analysis for  $K = 4$  revealed relative contributions of 72%, 25%, and 3% for *C. micrantha*, *C. maxima*, and *C. reticulata*, but the direct analysis of basic diagnostic SNPs only testified to *C. micrantha* specific alleles in homozygosity and heterozygosity without any specific alleles for the three other basic taxa.

## Discussion

### **Amplicon 454 sequencing was an efficient approach for species-diagnostic SNP mining in *Citrus* and for competitive allele-specific PCR marker development**

SNPs have become the most abundant and powerful polymorphic codominant markers that can be identified and characterised across whole genomes (Edwards and Batley, 2010). SNPs allow the development of very dense genetic linkage maps in animals and plants (Chen *et al.*, 2008; Troggio *et al.*, 2007; van Os *et al.*, 2006). Moreover, SNPs are generally considered to have a high identity by descent rate and thus are useful for phylogenetic and genetic association studies (de Bakker *et al.*, 2006; McCouch *et al.*, 2010). The primary limitation of SNP markers for gene-pool diversity analysis is that the revealed genetic organisation of the genotyped germplasm is strongly dependent on the discovery panel and the selection strategy used to develop a set of markers from all identified SNP positions (Albrechtsen *et al.*, 2010; Bradbury *et al.*, 2011; Clark *et al.*, 2005; Negrini *et al.*, 2010; Rosenblum and Novembre, 2007; Trask *et al.*, 2011). This ascertainment bias is particularly notable when SNPs are selected from only one sequenced heterozygous genotype. Examples include SNP characterisation in *Vitis vitifera* L., which used the whole genome sequence of the cultivar 'Pinot Noir' (Vezzulli *et al.*, 2008), and in *Citrus*, which used BAC-end sequencing data from the 'Clemenules' clementine (Ollitrault *et al.*, 2012b). Moreover, unexpected alleles may exist at any polymorphism. These unknown or 'null' alleles can interfere with accurate genotyping of the expected alleles and potentially impact genetic studies in a negative manner (Carlson *et al.*, 2006). The frequencies of these null alleles are likely to be higher when working with wider genetic distances between the genotyped samples and the discovery panel, due to additional polymorphisms in the genome area targeted by the marker PCR primers.

Next-generation technologies such as 454 amplicon sequencing present affordable opportunities to reduce genome complexity to well-dispersed gene fragments and provide information in approximately 500 pb read sequences. This allows extending greatly the discovery panel compared with a WGS approach. With respect to our objective of developing species-diagnostic PCR markers, the relatively long length of reads presented two advantageous features. First, multilocus sequences were produced, which were more powerful for inference of phylogenetic origins than single-locus sequences. Thus, in this study, combining the comparative average heterozygosity of the amplicons between the different accessions with PCA allowed the identification of mandarin varieties with interspecific heterozygosity for the considered fragment. After removing relevant fragment data for the introgressed accessions for the estimation of ancestral taxa allelic frequency, a better monolocus differentiation was observed for *C. reticulata* and *C. maxima* as testified by  $G_{ST}$ , and the identification of diagnostic alleles for these taxa was improved. Second, a decisive advantage is that it allows selecting diagnostic SNP position without close additional polymorphism that should affect the competitive allele-specific PCR of the developed KASPar marker.

The *Citrus* genus had favourable structuration for the identification of diagnostic markers from 500 pb amplicon sequences. The global SNP average rate was 36.7 SNP/kb (1,053 SNPs identified from 28,507 kb), and the average number of SNPs between varieties of two horticultural groups was 10.0–13.19 SNP/kb. As might be expected, this global SNP rate was lower than the 52.9 SNP/kb, reported by Garcia-Lor *et al.* (2013b), achieved by

Sanger sequencing of 27 gene fragments in four related genera (*Poncirus*, *Fortunella*, *Microcitrus*, and *Eremocitrus*) and *Citrus*. However, the authors noted values similar to those reported here for the average number of SNPs between varieties of pummelos, citrons, and mandarins. At the intraspecific level, the higher diversity in *C. reticulata* and *C. maxima* than in *C. medica* is in agreement with previous studies (Garcia-Lor *et al.*, 2012; Garcia-Lor *et al.*, 2013b; Ollitrault *et al.*, 2012b). Moreover, the high structuration of the diversity around *C. maxima*, *C. medica*, *C. reticulata*, and *C. micrantha*, confirmed by NJA and fixation index analysis ( $F_W$  and  $F_{ST}$ ), agreed with previous molecular studies (Barkley *et al.*, 2006; Federici *et al.*, 1998; Garcia-Lor *et al.*, 2013b; Herrero *et al.*, 1996; Nicolosi *et al.*, 2000) and numerical taxonomies based on phenotypic traits (Barrett and Rhodes, 1976) that recognised these taxa as the ancestors of the cultivated *Citrus* species. As a consequence of the high structuration resulting from the allopatric evolution of the four basic taxa, 271 of the 1,053 SNPs were found to be highly diagnostic for one of the four basic taxa (specific  $G_{ST} > 0.9$ ). A set of 85 SNPs was selected on the basis of their species- $G_{ST}$  value, their genome location, and the absence of close additional polymorphisms, and was used to develop novel KASPar markers. Seventy three provided consistent results. These 73 markers were complemented by 32 previously developed markers (Garcia-Lor *et al.*, 2013a; Ollitrault *et al.*, 2012b), and the full set was used to analyse the admixture genome structure of modern varieties.

### **The SNP marker set revealed the phylogenetic origins and admixture genomic structures of modern citrus cultivars and rootstocks**

#### ***Raw quantifications of the contributions of four ancestral taxa to modern varieties were coherent with previously estimated values from WGS data in seven genotypes.***

Recently, Wu *et al.* (2014) analysed the phylogenomic structure of several citrus varieties from WGS data. This analysis revealed *C. maxima* introgression in two mandarin varieties ('Ponkan' and 'Willowleaf') that were generally considered as true representatives of *C. reticulata*. In addition, the proportion of the *C. maxima* genome was quantified in these two cultivars, one mandarin hybrid ('W. Murcott' = 'Nadorcott'), clementine, sweet orange, and sour orange. Interestingly, the values found in our study from structure analysis based on a limited set of markers (103) were well correlated with the previous observations from WGS data (Wu *et al.*, 2014). Indeed, Wu *et al.* (2014) found decreasing proportions of *C. maxima* from sour orange to 'Willowleaf' as follows: sour orange (0.49), sweet orange (0.44), clementine (0.21), 'W. Murcott' (0.15), 'Ponkan' (0.077), and 'Willowleaf' (0.045). Our structure analysis inferred values of 0.509, 0.391, 0.236, 0.125, 0.035, and 0.040, respectively, for the same varieties. The small introgression of *C. reticulata* in 'Chandler' pummelo (0.002) identified by Wu *et al.* (2014) was not identified at a significant level by structure analysis (*C. reticulata* value = 0.003, equivalent to artefact noise). For the seven cultivars shared by the two studies, the correlation coefficients for estimations of *C. maxima* and *C. reticulata* contributions in the two studies were very high (0.993). Moreover, the analysis of the proportions of the diagnostic alleles of *C. reticulata* and *C. maxima* in heterozygosity and homozygosity agreed with the Wu *et al.* (2014) analysis, with only heterozygous diagnostic alleles for sour orange, homozygous and heterozygous alleles of both taxa for sweet orange, and only specific *C. maxima* alleles in heterozygosity in 'Willowleaf', 'Nadorcott', and clementine. The high coherence of our structure analysis without *a priori* definition of ancestral populations, and our analysis based on heterozygosity/homozygosity of diagnostic alleles of the four basic taxa, also confirmed the validity of the selected markers as species-diagnostic markers. Therefore, we consider that

the estimations of interspecific genomic structure provided by the set of species-diagnostic markers constitute a first approximation of the true phylogenomic structures of the analysed genotypes. As a result of their selection process, these markers minimised intraspecific variability. The markers should therefore be combined with a set of SSR markers to retain intraspecific variability data when analysing germplasm diversity. Indeed, in citrus, comparative studies of SSRs, indels, and SNP markers (Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2014) showed that SSRs were the more powerful tool for analysis of intraspecific variability.

**Numerous old and modern mandarin, tangor, tangelos and orangelos varieties were introgressed by *C. maxima***

The representative genotypes of the pummelo and citron horticultural groups appeared, respectively, as pure *C. maxima* and *C. medica* without identified interspecific introgressions. Similarly, no evidence of introgression was found in *C. micrantha*. As mentioned above, based on WGS analysis, Wu *et al.* (2014) evidenced introgressions of *C. maxima* in two mandarin varieties considered as true representatives of *C. reticulata*. A 454 amplicon haplotype study for chromosome 2 (Curk *et al.*, accepted) also revealed introgression by *C. maxima* in nine of the thirteen mandarins studied. In the present work, structure analysis revealed such introgressions in eight of the seventeen mandarins analysed. In addition to these cultivars, species-diagnostic marker analysis displayed one or two specific *C. maxima* alleles in heterozygosity in six cultivars. Only three mandarins were found without any indication of interspecific introgression ('Cleopatra', 'Sunki', and *C. dao-xianensis*). *C. dao-xianensis* is considered as an ancestral mandarin (Nicolosi, 2007), and the two others are mostly used as rootstock. From cytoplasmic analysis with mitochondrial indels (our unpublished data), it appears that these three mandarins share the acidic mandarin mitotype defined previously (Froelicher *et al.*, 2011), while all mandarins found introgressed by *C. maxima* share the second sweet mandarin mitotype (Froelicher *et al.*, 2011) and are cultivated for fruit consumption. Therefore, the reticulation(s) event(s) between *C. reticulata* and *C. maxima* and further introgression processes appear to be important components of sweet mandarin domestication. Only deep genomic analysis will determine whether all sweet mandarins result from one or several reticulation events and how introgression produced the modern mandarin displaying only a limited part of the *C. maxima* genome. During the 20<sup>th</sup> century, mandarin breeding was based on sexual hybridisations between mandarins but also between mandarins and sweet oranges (tangors), and mandarins and grapefruits (tangelos). All these recent hybrids, as well as supposed natural tangors such as 'Ortanique', 'Murcott', 'Temple', 'Nadorcott', and clementine, displayed admixture structure genomes between *C. reticulata* and *C. maxima* with predominant contribution of *C. reticulata*. It appears that mandarins, tangelos, tangors, sweet oranges, sour oranges, grapefruits, and orangelos (grapefruit × sweet orange hybrid) provide a large range of phylogenomic constitutions between the *C. reticulata* and *C. maxima* clusters. This is favourable for genetic association studies based on phylogenomic structures of the germplasm.

**The phylogenetic origin of secondary species is confirmed or revealed**

The partial apomixis of most of the secondary species explained that, in agreement with previous molecular studies (Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012; Luro *et al.*, 2008; Ollitrault *et al.*, 2012b), no polymorphisms were found between the analysed accessions within *C. sinensis*, *C. aurantium*, and *C. paradisi*, although they were highly heterozygous (Ho of 0.371, 0.424, and 0.346, respectively, with the 105 KASPar SNP

markers). This confirmed that the intraspecific polymorphisms in these secondary species arose from punctual mutation, transposable element movement (Bretó *et al.*, 2001), or epigenetic variation. Conversely, intervarietal variability was found for most of the analysed *C. limon* and *C. aurantifolia* accessions with the exception of 'Eureka' and 'Lisbon' lemons.

*C. sinensis* and *C. aurantium* are believed to derive from hybridisations between the *C. maxima* and *C. reticulata* gene pools (Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2013b; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b; Uzun *et al.*, 2009; Wu *et al.*, 2014). Previous SSR marker studies (Garcia-Lor *et al.*, 2012; Roose *et al.*, 2009) and a SNP study using a narrow discovery panel (Ollitrault *et al.*, 2012b) suggest that predominant portions of their genomes arose from the *C. reticulata* gene pool. The present study concurred with the conclusions obtained from WGS data (Wu *et al.*, 2014). In PCA, sour orange displayed an intermediary position between the pummelos and the mandarin group. It was highly heterozygous for *C. maxima* and *C. reticulata* specific alleles and structure analysis inferred close to 50% contribution for each of the two species. This is in agreement with a direct hybridisation between *C. maxima* and *C. reticulata* as proposed by Garcia-Lor *et al.* (2013b) and Wu *et al.* (2014). Sweet orange appeared to have developed from a more complex combination between two parents already displaying admixture structure between *C. reticulata* and *C. maxima*, as testified by the presence of specific *C. maxima* and *C. reticulata* alleles in homozygosity.

*C. paradisi*: the origin of grapefruit is attributed to a natural hybridisation between pummelo (*C. maxima*) and sweet orange (*C. sinensis*). This hybridisation may have occurred in the Caribbean more than 200 years ago (Barrett and Rhodes, 1976; de Moraes *et al.*, 2007; Webber, 1943). In this study, grapefruit had an intermediary position between the sweet orange and pummelo gene pools in the PCA representation. *C. reticulata* specific alleles were displayed in heterozygosity and one *C. maxima* allele was displayed in heterozygosity and homozygosity. Our results are therefore consistent with the pummelo x sweet orange hybridisation hypothesis.

*C. aurantifolia*: Tanaka (1961) considered 'Mexican' lime, 'Excelsa' lime, 'Alemow', and 'Palestinian' sweet lime as four distinct species, namely, *C. aurantifolia*, *C. excelsa*, *C. macrophylla*, and *C. limettioïdes*, respectively. However, Swingle and Reece (1967) combined these within *C. aurantifolia*. In all of the analyses reported here, the three first genotypes displayed minimal differences, whereas 'Palestinian' sweet lime appeared much more related to several *C. limon* cultivars. In PCA, the 'Mexican' lime, 'Excelsa' lime, and 'Alemow' displayed an intermediary position between the citron cluster and *C. micrantha*. These were highly heterozygous for *C. micrantha* and *C. medica* specific alleles, and structure analysis inferred close to 50% contribution for each of the two species. Therefore, our results suggest a similar origin by direct hybridisation between *C. micrantha* and *C. medica* for these three varieties. For 'Mexican' lime, this agrees with the hypothesis proposed by Nicolosi *et al.* (2000). Froelicher *et al.* (2011) showed that 'Mexican' lime and 'Alemow' share the *C. micrantha* mitotype. Recent analysis with mitochondrial indels and chloroplastic SSRs (Curk *et al.*, in press) leads to the same conclusion for 'Excelsa' lime. Therefore, 'Mexican' lime, 'Excelsa' lime, and 'Alemow' clearly have similar papeda × *C. medica* origins. An enhanced study of papeda germplasm will be necessary to definitively conclude *C. micrantha* or another papeda as the female parent of these three varieties. 'Palestinian' sweet lime structure will be discussed with *C. limon*.

*C. limon*: 'Eureka'/'Lisbon' lemon, 'Marrakech' limonette, 'Meyer' lemon, 'Rangpur' lime, 'Volkamer' lemon were considered by Tanaka (1961) as four species, respectively, *C. limon*,

*C. limetta*, *C. meyeri*, and *C. limonia*. These four species were joined in *C. limon* by Swingle and Reece (1967). Our analysis clearly distinguished two main groups of admixture structure. The first was comprised of the *C. limon*, *C. limetta*, and *C. meyeri* species as defined by Tanaka. All displayed a three species admixture structure (*C. medica*, *C. reticulata*, and *C. maxima*) with specific alleles of these three taxa in heterozygosity. The 'Palestinian' sweet lime (*C. limettioïdes*) displayed a very similar pattern. The contribution of *C. medica*, as revealed by structure analysis, was close to 50% for all these varieties. Therefore, they are probably direct hybrids between *C. medica* and varieties with admixture structure between *C. maxima* and *C. reticulata*. Based on RFLP, RAPD, and CAPS data, Nicolosi *et al.* (2000) were the first to propose that "yellow lemons" arose from a hybridisation between *C. aurantium* and *C. medica*. This hypothesis was supported by nuclear SSR (Curk *et al.*, in press; Garcia-Lor *et al.*, 2012) and SNP (Ollitrault *et al.*, 2012b) analyses as well as mitochondrial research (Froelicher *et al.*, 2011), and is coherent with the present results for 'Eureka' and 'Lisbon' lemon. The two *C. limonia* accessions ('Volkamer' lemon and 'Rangpur' lime) shared similar characteristics that differed from the previous lemon and lime patterns. In PCA, the *C. limonia* accessions displayed an intermediary position between *C. medica* and the mandarin group. The accessions were heterozygous for most *C. medica* and *C. reticulata* specific alleles, and structure analysis inferred close to 50% contributions from each of the two species. Therefore, *C. limonia* accessions results probably from direct hybridizations between *C. reticulata* and *C. medica*. Previous mitochondrial marker analyses (Froelicher *et al.*, 2011) showed that 'Volkamer' lemon and 'Rangpur' lime shared the cytoplasm of acid mandarins that would be expected for the maternal parents of the two *C. limonia* accessions.

*C. amblycarpa* is native to Indonesia, where it is called Djerook leemo (United States Department of Agriculture Agricultural Research Service, Beltsville Area, 2013). It is generally considered to be a mandarin hybrid, and its common English name is 'Nasnaran' mandarin. However, Froelicher *et al.* (2011) showed that it has a papeda mitotype, identical to *C. micrantha* and *C. hystrix*. In PCA, *C. amblycarpa* displayed an intermediary position between *C. micrantha* and the mandarin group. It was highly heterozygous for *C. micrantha* and *C. reticulata* specific alleles, and structure analysis inferred close to 50% contribution from each of the two species. Therefore, the hypothesis of papeda × acidic mandarin proposed for *C. amblycarpa* by Ollitrault *et al.* (2012b) was confirmed.

## Conclusions

The 454 sequencing of 57 gene fragments covering the nine chromosomes of the haploid citrus set for 26 genotypes revealed that the length of 454 reads and the level of differentiation between the ancestral taxa of modern citrus allowed efficient selection of ancestral species-diagnostic markers. A large number (271) of the 1,053 SNPs mined from the 28,507 kb of amplicon sequence displayed specific  $G_{ST}$  values  $>0.9$  for one of the basic taxa. Seventy-three KASPar markers were successfully developed and used with 32 previously developed SNP markers for analysis of the admixture structure of actual varieties and rootstock. Good correlations were observed between the contribution of the four basic taxa inferred with the set of species-diagnostic markers and recent published data from WGS of seven citrus varieties. The analysis of admixture genomic structures of cultivated citrus species and cultivars with 105 species-SNP diagnostic markers revealed *C. maxima* introgressions in most modern mandarin cultivars and in all recent selections of small citrus issued from 20<sup>th</sup> century breeding programs. This suggests that *C. reticulata* × *C. maxima* reticulation events and introgression processes were important elements of sweet mandarin domestication. The large range of phylogenomic constitutions between *C. reticulata* and *C. maxima* revealed in modern mandarins, tangelo, tangors, sweet orange, sour orange, grapefruit, and orangelo germplasm appears to be favourable for genetic association studies based on phylogenomic structures of the germplasm. Inferred admixture structures of several secondary citrus species were in agreement with previous hypotheses regarding their origin. Admixture structures also revealed the genomic structure and probable origin of several acid citrus varieties ('Excelsa' lime, 'Rangpur' lime, 'Alemow', 'Marrakech' Limonette, 'Palestinian' sweet lime, and 'Volkamer' lemon) and the incorrect assignation of 'Palestinian' sweet lime to *C. aurantifolia* by Swingle and Reece (1967). The developed species-diagnostic SNP marker set will be very useful for systematic estimation of admixture structure of the citrus germplasm. In addition, the marker set will find many applications in citrus genetics for genetic mapping of secondary species, analysis of meiotic mechanisms (disomic/tetrasomic inheritance) in double-diploid secondary species, and study of the origin of 2n gametes in interspecific admixture genotypes, as well as for more routine activities such as nucellar/zygotic discrimination.

## Funding

This work was supported by a grant [AGL2011-26490] from the Ministry of 'Economía y Competitividad', 'Fondo Europeo de Desarrollo Regional' (FEDER) and a grant [Prometeo II/2013/008] from the Generalitat Valenciana, Spain.

---

## Chapitre III

---



# **Nuclear, mitochondrial and chloroplastic SSR, InDel and SNP markers reveal the diversity and the phylogenetic origin of limes and lemons.**

Franck Curk<sup>12</sup>, Frédérique Ollitrault<sup>2</sup>, Andres Garcia-Lor<sup>2</sup>, François Luro<sup>1</sup>, Luis Navarro<sup>2</sup> and Patrick Ollitrault<sup>23</sup>

<sup>1</sup>UMR AGAP, Institut National de la Recherche Agronomique (Inra), Centre Inra de Corse, F-20230 San Giuliano, France

<sup>2</sup>Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada (Valencia), Spain

<sup>3</sup>UMR AGAP, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), TA A-108/02, 34398 Montpellier, Cedex 5, France

Manuscript prepared to be submitted to *Annals of Botany*

## Titre et Résumé en français

### **La diversité et les origines phylogénétiques des limettiers et des citronniers sont révélées par l'utilisation conjointe de marqueurs SSR, indel et SNP, nucléaires, mitochondriaux et chloroplastiques**

Il existe de nombreuses déclarations contradictoires sur l'origine taxinomique des limettiers et des citronniers. Des études biochimiques, la taxinomie numérique et des études moléculaires plus récentes ont suggérées que toutes les espèces d'agrumes cultivées dérivait d'hybridation interspécifique entre quatre taxons de base (*C. reticulata*, les mandariniers; *C. maxima*, les pamplemoussiers; *C. medica*, les cédratiers et *C. micrantha* un papeda). De nombreux auteurs s'accordent sur l'origine des citronniers à fruits jaunes de type Lisbonne ou Eureka et des limettiers de type Mexicain. Cependant, l'origine des autres citronniers et limettiers reste controversée et non encore confirmée. L'objectif de ce travail était d'effectuer une analyse approfondie de la diversité, des structures génétiques et des origines des limettiers et des citronniers. Il a été basé sur 106 variétés de limettiers et citronniers et 33 représentants des quatre taxons de base ainsi que 7 variétés des autres espèces secondaires. Ce travail combine l'étude de la phylogénie maternelle (3 Indels mitochondriaux et 5 SSR chloroplastiques), l'analyse des structures nucléaires (8 Indels, 19 SSRs et 96 SNPs), ainsi que l'évaluation de la ploïdie par cytométrie en flux. Le groupe horticole des limettiers et des citronniers est apparu très polymorphe avec des individus diploïdes, triploïdes et tétraploïdes, une phylogénie maternelle impliquant quatre types cytoplasmiques sur les six rencontrés dans le genre *Citrus*, et la contribution, au niveau nucléaire des quatre taxons de base. Toutes les variétés de limettiers et de citronniers présentent une hétérozygotie élevée avec des structures génomiques interspécifiques impliquant deux, trois et même les quatre taxons ancestraux pour certaines variétés. *C. medica* est apparu comme un élément commun aux variétés classées dans les limes et citrons et en est très certainement très fréquemment le parent mâle direct, en combinaison avec *C. micrantha*, *C. reticulata*, *C. aurantium* et des hybrides [*C. maxima* x *C. reticulata*]. Deux origines ont été identifiées pour les limes triploïdes. Les limes *C. latifolia* résulteraient de la fécondation d'un ovule haploïde de *C. limon* par un gamète diploïde de *C. aurantifolia*, alors que les variétés triploïdes *C. aurantifolia* seraient issues d'un rétro-croisement interspécifique (un ovule diploïde de *C. aurantifolia* pollinisé par un pollen haploïde de *C. medica*). Il est clair qu'un grand nombre d'événements de réticulation indépendants a généré les sous-groupes génétiques des limettiers et des citronniers et que les polymorphismes intra-groupe sont probablement dus à des mutations ponctuelles, des variations épigénétiques ou des délétions de fragments génomiques. Ce nouvel éclairage sur les structures génétiques interspécifiques et les origines phylogénétiques des principaux sous-groupes de limettiers et citronniers est essentiel pour une meilleure mobilisation de la biodiversité des agrumes dans les programmes de création variétale (porte-greffe et variétés d'agrumes acides), dans l'objectif de mieux répondre aux contraintes liées aux stress biotiques et abiotiques.

## Abstract

The origin of limes and lemons has been a source of conflicting taxonomic statements. Biochemical studies, numerical taxonomy and more recent molecular studies suggested that all cultivated *Citrus* species should result from interspecific hybridization between four basic taxa (*C. reticulata*, the mandarins; *C. maxima*, the pummelos; *C. medica* the citrons and *C. micrantha* a papeda). Several authors agreed on the origin of the yellow lemons and the Mexican lime types. However, the origin of the other lemons and limes is still controversial or not confirmed yet. The objective of the present work was to perform an extended analysis of the diversity, genetic structure and origin of lime and lemon germplasm. It was based on 106 lime and lemon accessions and 33 representatives of the four basic taxa as well as seven accessions of the other secondary species. It combined maternal phylogeny studies based on three mitochondrial Indels, five chloroplastic SSRs and nuclear structure analysis based on 123 nuclear markers (8 Indels, 19 SSRs and 96 SNPs) and ploidy level evaluation by flow cytometry. The lime and lemon horticultural group appeared highly polymorph with diploid, triploid and tetraploid varieties, with maternal phylogeny involving four cytoplasmic types over the six encountered in the *Citrus* genus and nuclear contribution of the four *Citrus* basic taxa. All lime and lemon varieties were highly heterozygous and displayed interspecific admixture involving two or three, but also the four ancestral taxa genomes for a few accessions. *C. medica* appeared as a common component and was very probably the direct male parent for the main sub-groups in combination with *C. micrantha*, *C. reticulata*, *C. aurantium* and *C. maxima* x *C. reticulata* hybrids. Two origins were identified for the triploid limes. *C. latifolia* varieties should result from the fertilization of a haploid ovule of *C. limon* by a diploid gamete of *C. aurantifolia* while the *C. aurantifolia* triploid varieties should results from an interspecific backcross (a diploid ovule of *C. aurantifolia* pollinated by *C. medica*). It appeared that the limes and lemons resulted of many independent reticulation events defining genetic sub-groups and that intra-subgroup polymorphisms could be due to punctual mutations, epigenetic variations or genomic fragment deletion. The new insight on the global interspecific admixture structures and phylogenetic origin of the main sub-groups of limes and lemons will be essential for a better utilization of the citrus biodiversity to create new rootstock and acid citrus varieties with improved adaptation and resistance to biotic and abiotic constraints.

## Introduction

Lemons and limes are cultivated over all Mediterranean, subtropical and inter-tropical areas all over the world. With 15 million tonnes, they are the third most important citrus horticultural group. (FAO, 2014). With nearly 1.6 million tonnes internationally traded, lemons and limes are at the 11th place of fruit trades. Lemon and lime market concern both fresh-fruit and processed products. Lemon market, after increasing until 2007 is now in stagnation and in direct competition with limes which consumption dramatically increases since the 80's (Duportal *et al.*, 2013). A few cultivars are extensively cultivated all over the world while an important diversity exists in this horticultural group.

Southern to Eastern Asia, Malaysia, New Caledonia and Australia are the places of origin of Citrus fruits and their relatives (Swingle and Reece, 1967). The genus *Citrus* L., which includes limes and lemons, are grown from tropical to temperate parts of the world over several thousands of years. Based on Swingle and Reece classification (1967) lemons and limes are classified into two species, respectively *Citrus limon* (L.) Burm. F. and *Citrus aurantifolia* (Christm.) Swingle. According to the Tanaka classification (1954), limes and lemons are classified into 37 species. These conflicting classifications result from the total sexual compatibility between *Citrus* species and the frequent occurrence of apomixis (due to nucellar polyembryony; Scora, 1975; Barrett and Rhodes, 1976), that leads many taxonomists to consider interspecific hybrids fixed by apomixis or vegetative propagation as new species.

Thus, the origin of the lemon has been a source of conflicting statements by both historians and taxonomists. Although Bonavia (1888) concluded that the lemon reached India relatively late, Tolkowsky (1938) believed it to be a fruit mentioned in early Sanskrit texts assigned to the period 800 B.C., Webber *et al.* (1967), suggested Southern China or possibly upper Burma as the native home of the lemon. Early taxonomists hypothesized that *C. limon* was a derivative or hybrid of *C. medica* L., the citron. Gallesio (1811) and Tolkowsky (1938) noted that during the middle ages the lemon was considered as a variety of the citron and de Candolle (1886) considered lemon to be closely related to the citron. Lemon, lime and citron were included in the species *C. medica* by Linnaeus (1753) and other early taxonomist. More recent authors such as Risso (1813), Swingle (1914; 1943), Tanaka (1954), and Bhattacharya and Dutta (1956) gave separate specific names to both citron and lemon. Swingle (1914) isolated the citron as *C. medica*, lime as *C. aurantifolia*, and lemon as *C. limonia* Osbeck, but in 1943 he reclassified lemon as *C. limon*. Swingle also considered lemon to be a probable relative species of citron and suggested that it might be of hybrid origin, between citron and limes. Hodgson (1955) put citron, lemon and lime into one group. Therefore early taxonomic works recognized parentage between citrons, limes and lemons, but failed to propose clear and definitive conclusions on the origin of these species. Biochemical studies (Malik *et al.*, 1974; Scora, 1975), numerical taxonomy (Barrett and Rhodes, 1976) and more recent molecular studies (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b) suggested that all cultivated *Citrus* species should result from interspecific hybridization between four basic taxa (the mandarins, *C. reticulata* Blanco; the pummelos, *C. maxima* (Burm.) Merr.; the citrons, *C. medica* L., and the papedas *C. micrantha* Wester). Regarding limes and lemons, it is now generally accepted that they are related to citron (*Citrus medica* L.) (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Bayer *et al.*, 2009) but opinions differ on the origin of specific varieties. Barrett and Rhodes, (1976) and Federici *et al.* (1998) argued, in agreement with Swingle, that the classical 'Eureka' or

'Lisbon' type yellow lemons resulted from direct hybridization of citron and lime, while more recently Nicolosi *et al.* (2000) proposed that the 'Lisbon' type yellow lemon was the product of a direct hybridization between sour orange (*C. aurantium* L.) and citron. Sour orange is itself a supposed hybrid between *C. maxima* and *C. reticulata* (Swingle and Reece, 1967; Scora, 1975; Green *et al.*, 1986; Scora, 1988; Yamamoto *et al.*, 1993; Nicolosi *et al.*, 2000; Garcia-Lor *et al.*, 2013b; Wu *et al.*, 2014). 'Mexican' lime was proposed as a direct hybrid between citron and *C. micrantha* (Scora, 1975; Nicolosi *et al.*, 2000). These hypotheses for 'Lisbon' type yellow lemons and 'Mexican' type limes were recently agreed by relatively large SNP analysis (Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b). However, the origin of the other lemons and limes is still controversial or not confirmed yet.

The objective of the present work was to perform an extended analysis of the diversity, genetic structure and origin of lime and lemon germplasm. It was based on lime and lemon accessions (106) available in the INRA-CIRAD (France) and IVIA (Spain) germplasm depositories and representatives of the four basic taxa (33) as well as the other secondary species (7). It combined maternal phylogeny studies based on mitochondrial indels (Froelicher *et al.*, 2011) and chloroplastic SSRs (Bryan *et al.*, 1999; Weising and Gardner, 1999) and nuclear diversity analysis. At nuclear level, Barkley *et al.* (2009) showed that homoplasmy may limit the usefulness of SSR markers in identifying the phylogenetic origin of DNA fragments in citrus while Garcia Lor *et al.* (2012), Ollitrault *et al.* (2014) and Curk *et al.* (submitted a) showed that Insertion/deletion polymorphisms (indels) and SNPs were more adapted to select efficient specific diagnostic markers. However, selected markers for specific alleles of the ancestral taxa largely miss intraspecific variation (Garcia-Lor *et al.*, 2013b; Ollitrault *et al.*, 2014). Therefore we combined 8 indels and 96 SNP markers to reveal the interspecific structure of the different accessions and 19 SSR markers for intraspecific polymorphism information. A ploidy analysis of all accessions by flow cytometry completed the nuclear genome characterization.

## Material and Methods

### Plant material

Leaves from 147 accessions of the *Citrus* genus were collected from the IVIA Citrus Germplasm Bank of pathogen-free plants (Valencia, Spain; accessions with IVIA identification number) and the Inra/Cirad CRB Citrus (*Centre de Ressources Biologiques Citrus*, NFS96-900) of San Giuliano (Corsica, France; accessions with SRA identification number) [Annex IV, Table S.III.1]. Tanaka (1961) and Swingle and Reece (1967) botanical classifications are indicated when the varieties have been classified. According to Swingle and Reece classification (1967) twenty seven and sixty four varieties are representative of *C. aurantifolia* (Christm.) Swing. and *C. limon* (L.) Burm, respectively. In addition, fifteen unclassified varieties phenotypically related to the lemon and lime groups, twenty six accessions representative of the four basic taxa (twelve mandarins [nine *C. reticulata* and three *C. tachibana*], eleven pummelos [*C. maxima*], 8 citrons [*C. medica*], and 2 papedas [*C. micrantha*]), and 8 other secondary species (2 sour oranges [*C. aurantium*], 3 grapefruits [*C. paradisi*], 1 clementine [*C. clementine*] and 2 sweet oranges [*C. sinensis*] were included.

### Ploidy evaluation

Ploidy level was determined by flow cytometry according to the methodology described by Aleza *et al.* (2009). Each sample consisted of a small piece of leaf (0.5 mm<sup>2</sup> collected from each varieties with a similar leaf piece taken from a tetraploid control plant (doubled-diploid 'Shamouti' sweet orange; Aleza *et al.*, 2011). Samples were chopped together using a razor blade in the presence of a nuclei isolation solution (High Resolution DNA Kit Type P, solution A; Partec, Munster, Germany). Nuclei were filtered through a 30µm nylon filter and stained with a DAPI (4,6-diamine-2-phenylindol; High Resolution DNA Kit Type P, solution B; Partec) solution. Following a 5-min incubation period, stained samples were run in a Ploidy Analyzer (Partec PA) flow cytometer equipped with a HBO 100-W high-pressure mercury bulb and both KG1 and BG38 filter sets. Histograms were analysed using the dpac v2.0 software (Partec), which determines peak position, coefficient of variation (CV) and the relative ploidy index of the samples. In the cases that only one peak was observed, an additional analysis was performed with a diploid internal control.

### DNA extraction

High molecular weight genomic DNA was extracted from leaf samples using the DNeasy Plant Mini Kit (Qiagen S.A.; Madrid, Spain) according to the manufacturer's instructions.

### SSRs and Indel marker genotyping

All varieties were genotyped with nineteen SSRs (Kijas *et al.*, 1997; Froelicher *et al.*, 2008; Luro *et al.*, 2008; Cuenca *et al.*, 2011) and 8 indel nuclear markers (Garcia-Lor *et al.*, 2012), 5 chloroplastic SSRs (Bryan *et al.*, 1999; Weising and Gardner, 1999) and 3 mitochondrial indels (Froelicher *et al.*, 2011) [Annex IV, Table S.III.2 and Figure S1]. PCR amplifications were performed using a thermocycler ep gradient S (Eppendorf®) in 10 mL final volume containing 0.8 U of Taq DNA polymerase (Fermentas®), 2 ng/mL of citrus DNA, 0.2 mM of wellRED (Sigma®) dye-labelled forward primer, 0.2 mM of non dye-labelled reverse primer, 0.2 mM of each dNTP, 10X PCR buffer and 1.5 mM MgCl<sub>2</sub>. The PCR protocol was as follows: denaturation at 94°C for 5 min followed by 40 repeats of 30 s at 94°C, 1 min at 50°C or 55°C, 45 s at 72°C; and a final elongation step of 4 min at 72°C.

Capillary electrophoresis was carried out using a CEQ™ 8000 Genetic Analysis System (Beckman Coulter Inc., Fullerton, CA, USA). Data collection and analysis were carried out with GenomeLab GeXP (Beckman Coulter Inc.) version 10.0 software. Allele dosage was calculated using the MAC-PR (microsatellite DNA allele counting-peak ratio) method (Esselink *et al.*, 2004), validated in citrus by Cuenca *et al.* (2011).

### **Kaspar genotyping**

Ninety six SNP markers (Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013a; Curk *et al.*, submitted a), [Annex IV, Table S.III.2] were used for genotyping using KASPar technology by KBioscience® (<http://www.kbioscience.co.uk/>). The KASPar™ Genotyping System is a competitive, allele-specific dual Förster resonance energy transfer (FRET)-based assay for SNP genotyping. Primers were designed by LGC Genomics®, based on the SNP locus-flanking sequence (approx. 50 nucleotides on each side of the SNP). Two 40-mer allele-specific oligonucleotides and one common 20-mer oligonucleotide were defined for each locus. Detailed information for all SNP markers can be found in Annex IV, Table S.III.2 and their distribution over the nine chromosomes is schematized in Annex IV, Figure S.III.1. Additional details about this genotyping method can be found in Cuppen (2007). The fluorescence signals of PCR products were measured with Fluostar Omega (BMG) and genotype calling was made with KlusterCaller software (LGC Genomics). For triploid and tetraploid limes the allele dose was estimated from relative allele signal according to the method described by Cuenca *et al.* (2013a).

### **Genetic analyses**

Expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ), the F Stat parameters ( $F_w$  [Wright, 1978] and  $F_{ST}$ ) were calculated using GENETIX v. 4.03 software (Belkhir *et al.*, 1996-2004). The estimation of the value of the markers to differentiate one of the four basic taxa from the three other ones was based on  $G_{ST}$  parameter (Nei, 1973) considering two subpopulations: (1) the concerned taxa and (2) a theoretical population of the 3 other basic taxa.  $G_{ST}$  estimations were computed with GENETIX v. 4.03 software (Belkhir *et al.*, 1996-2004).

Genetic relationships were studied by Neighbour-Joining analysis (NJA) and Factorial Analyses with DARwin software (Perrier and Jacquemoud-Collet, 2006), using the simple matching dissimilarity index. Cluster robustness was tested by running 1000 bootstraps.

Population structure was inferred with the Structure version 2.3.4 program (Pritchard Lab, 2014), which implements a model-based clustering method using genotype data (Pritchard *et al.*, 2000; Falush *et al.*, 2003). No a priori population structure was defined. The option used was the linkage model, with correlated allele frequencies and computed probability of the data for K estimating. Analyses were made with K value (number of subpopulation) varying between 1 and 10. The statistics used to select the correct K value were the ones followed by Evanno *et al.* (2005): the mean likelihood,  $L(K)$ ; the mean difference between successive likelihood values of K,  $L'(K)$ ; the absolute value of this difference,  $|L''(K)|$ ; and  $\Delta K$ , which is the mean of the absolute values of  $L''(K)$  divided by the standard deviation of  $L(K)$ . The likelihood distribution  $L(K)$  and  $\Delta K$  were the main values used to choose the optimal K value of the population. Ten runs of Structure were performed with 50,000 steps of burning followed by 50,000 Monte Carlo Markov Chain (MCMC) repetitions. For the better K value, the 10 independent Structure run clusters output have been permuted and aligned and average frequency and standard error of the contribution of each basic population estimated.

## Results

### Ploidy evaluation

Five triploid ('Ambilobe', 'Coppennhad', 'Madagascar', 'Mothasseb', 'Tanepao') and one tetraploid ('Giant Key') limes, were identified among the 14 accessions classified as *C. aurantifolia* by Tanaka (1961). The five accessions of *C. latifolia* ('Bears', 'El Kseur', 'IAC-5', 'Persian', and 'Tahiti') were triploid. All other analyzed varieties were found to be diploid [Annex IV, Table S.III.1].

### Cytoplasmic diversity

The three indel mitochondrial markers revealed 5 mitotypes within *Citrus* [Annex IV, Figure S.III.2 and Table S.III.3], in agreement with Froelicher *et al.* (2011). Mitotype 1 was characteristic of true citrons and was not found in any other varieties. Mytotype 2 included the 2 *C. micrantha*, 12 *C. aurantifolia*, *C. aurata*, the 2 *C. excelsa*, and *C. macrophylla* and *C. weberii* accessions. Mitotype 3 included the height edible mandarins and the clementine, but no limes or lemons. Mitotype 4 corresponded to the so called acid mandarins by Froelicher *et al.* (2011). It included 4 non edible mandarins, two *C. aurantifolia*, the 3 *C. jambhiri*, the two *C. Karna*, the 6 *C. limonia*, the two *C. meyeri* and 6 *Citrus sp.* accessions. Mitotype 5 previously identified as *C. maxima* mitotype (Froelicher *et al.*, 2011) included all *C. maxima* accessions, the 2 *C. aurantium*, *C. bergamia*, the 5 *C. latifolia*, the 3 *C. limetta*, the 41 *C. limon*, the five *C. lumia*, the 3 *C. limettioïdes*, the 2 *C. paradisi*, the 2 *C. sinensis*, *C. hassaku*, *C. meyeri* and *C. pyriformis* accessions as well as 9 *Citrus sp.* accessions.

Chloroplastic markers revealed 6 chlorotypes mostly in agreement with the mitotypes [Annex IV Figure S.III.2] but one difference: the mytotype 5 was subdivided in two groups. The first one (chlorotype 5) shared the sour orange chlorotype with *C. bergamia*, the 5 *C. latifolia*, the 3 *C. limetta*, the 39 *C. limon*, 1 *C. lumia* and 4 *C. Citrus Sp.* The second one (chlorotype 6) was found in the 11 *C. maxima*, four *C. lumia*, the 3 *C. limettiodes*, the 2 *C. paradisi*, the 2 *C. sinensis*, *C. hassaku*, 1 *C. meyeri* and the *C. pyriformis* accessions as well as 2 *C. limon* and 5 *Citrus sp.* Coupling mitochondrial and chloroplastic data we therefore identified 6 cytoplasmic types (CT) corresponding to the 6 chlorotypes (Figure III.1 and Annex IV, Figure S.III.2). Lime and lemon accessions were found in four of the six CTs; respectively, 53, 16, 20 and 17 in the sour orange, pummelo/grapefruit/sweet orange, non-edible mandarin and *C. micrantha* CTs.



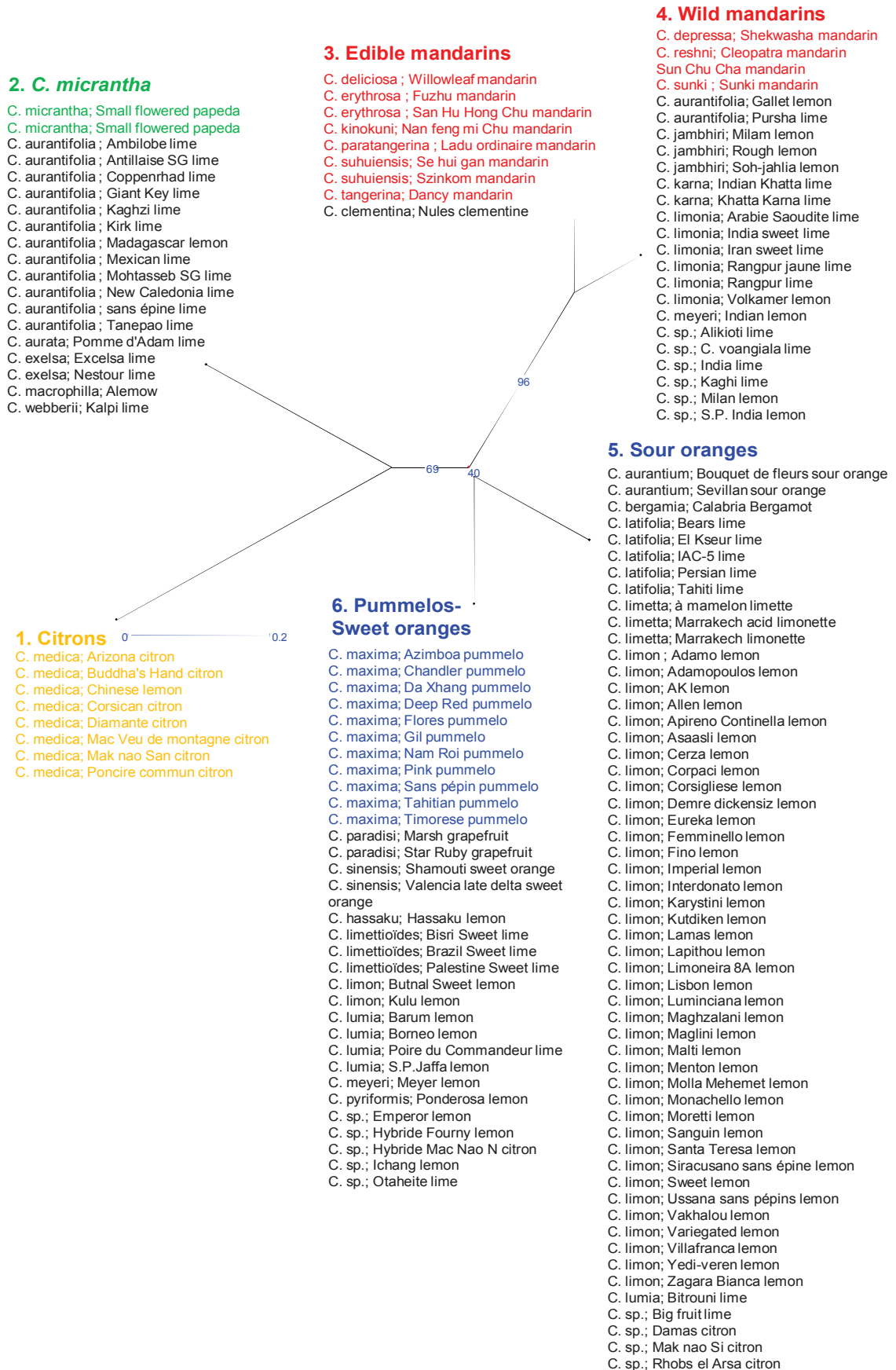


Figure III.1. Cytoplasmic type assignment of the 147 citrus varieties. NJ tree established from 3 mitochondrial indels and 5 chloroplastic SSRs.

## Nuclear diversity at diploid level

### Potential of the different kind of markers for ancestral taxa differentiation

SSRs were the more polymorphic markers with an average of 8.11 alleles/locus over the 26 accessions representative of the four basic taxa while the selected indels and SNPs displayed relatively 2.75 and 2 allele/locus (Table III.1). This higher SSR polymorphism was also observed within varieties (observed heterozygosity  $H_o = 0.42 \pm 0.02$  in average compared with  $0.10 \pm 0.07$  and  $0.07 \pm 0.02$  for Indels and SSRs respectively) and within each basic taxa ( $H_e$  ranged respectively from 0.21 to 0.55, 0 to 0.27 and 0.01 to 0.12 for SSRs, Indels and SNPs respectively). In returns SSRs displayed lower values for the global structuration parameters ( $F_w$  of  $0.47 \pm 0.09$ ,  $0.78 \pm 0.14$  and  $0.78 \pm 0.06$ ;  $F_{ST}$  considering the 4 basic taxa as subpopulations of  $0.51 \pm 0.09$ ,  $0.85 \pm 0.10$  and  $0.79 \pm 0.06$  for SSRs, Indels and SNPs respectively (Table III.1).

Table III.1. Basic horticultural group diversity (Ho/He/Fw)

		<i>C. reticulata</i> N=12	<i>C. maxima</i> N=11	<i>C. medica</i> N=8	<i>C. micrantha</i> N=2	Total N=33
19 SSRs	NA	4.47	3.26	1.89	1.44	8.11
	$H_o$	$0.610 \pm 0.112$	$0.454 \pm 0.145$	$0.100 \pm 0.064$	$0.406 \pm 0.240$	$0.422 \pm 0.0178$
	$H_e$	$0.551 \pm 0.098$	$0.474 \pm 0.114$	$0.211 \pm 0.109$	$0.211 \pm 0.122$	$0.771 \pm 0.044$
	$F_w$	$-0.110 \pm 0.066$	$0.045 \pm 0.223$	$0.459 \pm 0.260$	$-0.905 \pm 0.187$	$0.465 \pm 0.086$
	$F_{ST}$					$0.513 \pm 0.085$
	MLG	12	10	8	2	32
8 Indels	NA	1.63	1.36	1	1.13	2.75
	$H_o$	$0.208 \pm 0.174$	$0.056 \pm 0.066$	0	$0.125 \pm 0.244$	$0.102 \pm 0.067$
	$H_e$	$0.268 \pm 0.088$	$0.051 \pm 0.058$	0	$0.062 \pm 0.122$	$0.443 \pm 0.048$
	$F_w$	$-0.103 \pm 0.184$	$-0.084 \pm 0.044$	N	-1	$0.782 \pm 0.143$
	$F_{ST}$					$0.847 \pm 0.095$
	MLG	10	3	1	1	15
96 SNPs	NA	1.39	1.14	1.14	1.02	2
	$H_o$	$0.125 \pm 0.039$	$0.037 \pm 0.023$	$0.039 \pm 0.022$	$0.010 \pm 0.020$	$0.068 \pm 0.016$
	$H_e$	$0.118 \pm 0.035$	$0.044 \pm 0.025$	$0.038 \pm 0.022$	$0.005 \pm 0.010$	$0.327 \pm 0.026$
	$F_w$	$-0.061 \pm 0.082$	$0.111 \pm 0.213$	$-0.066 \pm 0.174$	-1	$0.775 \pm 0.057$
	$F_{ST}$					$0.787 \pm 0.055$
	MLG	10	10	7	1	28
123 markers	NA	1.89	1.48	1.24	1.05	2.99
	$H_o$	$0.205 \pm 0.047$	$0.102 \pm 0.039$	$0.045 \pm 0.020$	$0.069 \pm 0.044$	$0.125 \pm 0.028$
	$H_e$	$0.189 \pm 0.042$	$0.110 \pm 0.038$	$0.062 \pm 0.026$	$0.036 \pm 0.023$	$0.402 \pm 0.035$
	$F_w$	$-0.08 \pm 0.056$	$0.059 \pm 0.139$	$0.162 \pm 0.182$	$-0.92 \pm 0.145$	$0.727 \pm 0.050$
	$F_{ST}$					$0.748 \pm 0.059$
	MLG	12	11	8	2	33

N: number of analysed accessions; NA: mean number of allele/locus;  $H_o$ : observed heterozygosity;  $H_e$ : expected heterozygosity;  $F_w$ : Wright Fixation Index;  $F_{ST}$ : inter taxa structuration parameter; MLG: number of different multi-locus genotypes

The analysis of species  $G_{ST}$  confirmed the relatively low value of SSRs for specific differentiation (Figure III.2a) compared with selected indels and SNPs. Among the low  $G_{ST}$  values (<0.5), some SNP markers developed from clementine BAC end sequencing data (Ollitrault *et al.*, 2012b) revealed intraspecific variability for *C. reticulata* and *C. maxima* (Figure III.2B). To analyse the interspecific admixture of limes and lemons we selected the 73 markers with a  $G_{ST}$  max value over 0.5 and unambiguous identified specific allele (Figure III.2C).

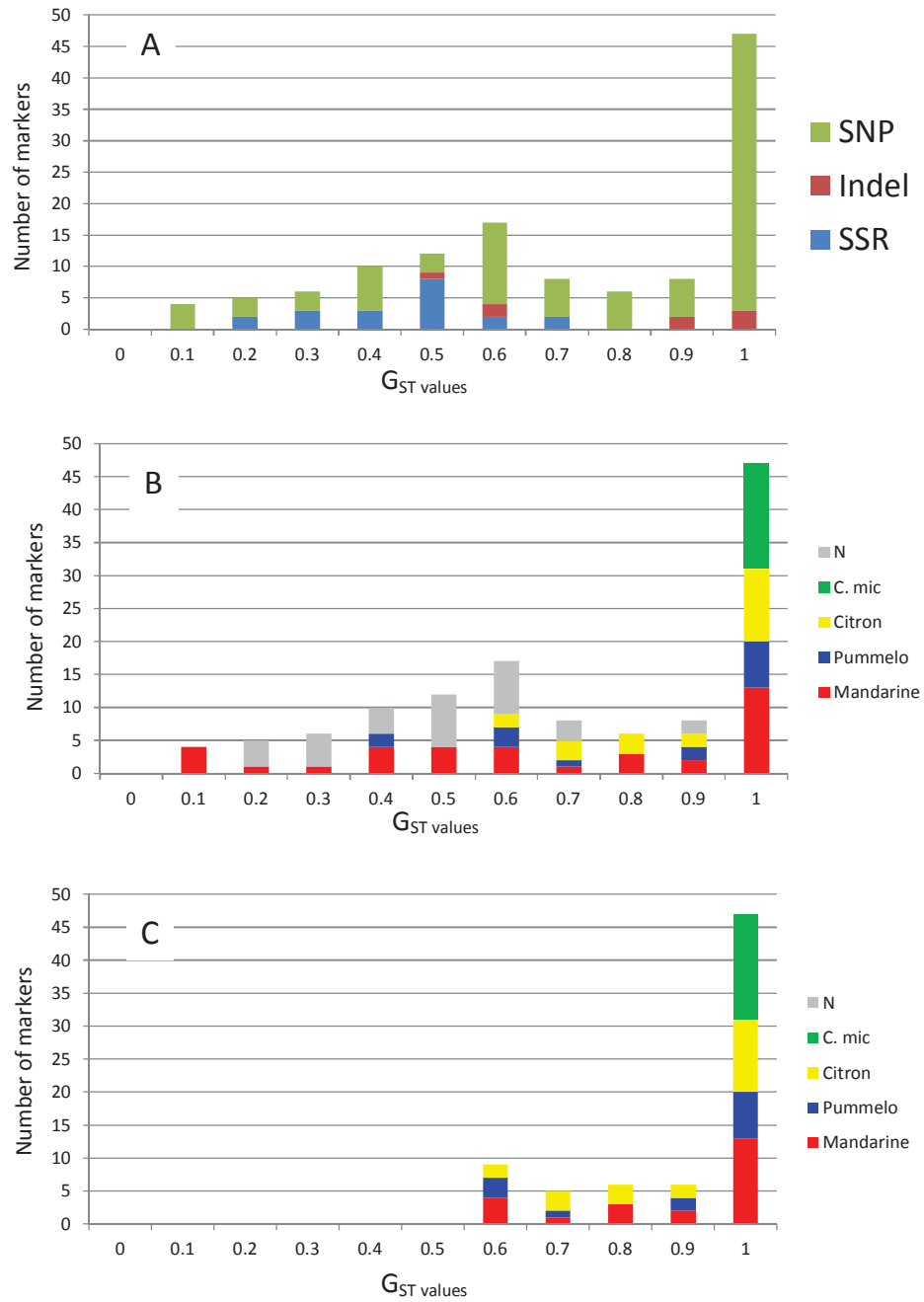


Figure III.2. Distribution of the higher specific  $G_{ST}$  values. A: comparative distribution of the three kinds of markers (SSRs, indels, and SNPs); B: diagnostic value for the ancestral taxa over the 123 markers; C: diagnostic value for the ancestral taxa over the 73 selected markers for Structure and specific allele Hom/het analysis.

### Nuclear structure of limes and lemons

Dissimilarities (simple matching index) between each pair of diploid varieties were calculated with DARwin software (Perrier and Jacquemoud-Collet, 2006). One hundred and three Multi Locus Genotypes (MLGs) were identified. Twenty seven yellow lemons (*C. limon*) displayed the same MLG (Annex IV, Table S.III.1 and Table III.2). Redundancy was also observed for 2 *C. lumia* ('Barum' and 'Borneo' lemons), 2 *C. limettioides* ('Bisri' and 'Brazil' sweet limes), 2 *C. limetta* ('A mamelon limette' and 'Marrakech acid limonette'), the two *C. sinensis*, the two *C. paradisi* and the two *C. aurantium*. In addition, the *C. webberii* accession ('Kalpi' lime) was found identical to one of the *C. excelsa* one ('Excelsa' lime). Genetic parameters were evaluated from the MLGs matrix.

Table III.2. Diversity of Swingle and Reece and Tanaka limes and lemons species based on MLG matrix and 123 molecular markers

Swingle and Reece	Tanaka	N	MLGs	Ho	He	Fw
<i>C. aurantifolia</i> (Christm.) Swing.	<i>C. aurantifolia</i>	8	8	0.46±0.06		
	<i>C. bergamia</i>	1	1	0.40±0.09		
	<i>C. excelsa</i>	2	2	0.46±0.09		
	<i>C. limettioides</i>	3	2	0.54±0.09		
	<i>C. macrophylla</i>	1	1	0.44±0.09		
	<i>C. webberii</i>	1	1	0.46±0.09		
	Total <i>C. aurantifolia</i>	16	14	0.47±0.06	0.352±0.036	-0.23±0.07
<i>C. limon</i> (L.) Burm.	<i>C. aurata</i>	1	1	0.44±0.09		
	<i>C. jambhiri</i>	3	3	0.55±0.07		
	<i>C. karna</i>	2	2	0.54±0.09		
	<i>C. limetta</i>	3	2	0.56±0.09		
	<i>C. limon</i>	41	15	0.51±0.08		
	<i>C. limonia</i>	6	6	0.58±0.08		
	<i>C. lumia</i>	5	4	0.35±0.05		
	<i>C. meyeri</i>	2	2	0.58±0.08		
	<i>C. pyriformis</i>	1	1	0.25±0.08		
	Total <i>C. limon</i>	64	36	0.51±0.06	0.36±0.04	-0.29±0.06
<i>C. paradisi</i> Macf.	<i>C. hassaku</i>	1	1	0.49		
<i>C. sp</i>	<i>C. sp.</i>	15	15	0.48±0.05	0.378±0.04	-0.23±0.05
All limes and lemons		95	66	0.49±0.05	0.38±0.04	-0.85±0.05

N: number of varieties; MLG: Multi-Locus Genotype; Ho: observed heterozygosity; He: expected heterozygosity; F<sub>w</sub>: Wright fixation index.

With an average of 0.49±0.05, all Tanaka (1961) and Swingle and Reece (1967) lime and lemon species displayed high heterozygosity (Table III.2) when compared with the four basic taxa (Table III.1). Similar average values (0.48±0.05) were observed for the unclassified limes and lemons like varieties). *C. pyriformis* displayed the lower Ho (0.25) while all other Tanaka species had values between 0.35 for *C. lumia* and 0.58 for *C. limonia*. Both *C. limon* and *C. aurantifolia* species of Swingle and Reece revealed similar intraspecific polymorphisms (He = 0.35±0.04 and 0.36±0.04 respectively) and an excess of heterozygosity (negative F<sub>w</sub> values) that should be explained by vegetative propagation of hybrid cultivars.

NJ analysis was performed on the 103 MLGs with the 123 molecular markers and revealed several clusters joining lime and lemon accessions with high bootstrap values (> 80%; Annex IV, Figure S.III.3). The varieties joined in a same cluster systematically shared the same CT. Therefore the presentation of nuclear results (clusters from NJ analysis, Structure analysis –Figure III.3– and study of the frequencies of diagnostic alleles of the four ancestral taxa in homozygosity and heterozygosity –Figure III.4–) is organized according to the CTs. STRUCTURE software analysis and homozygosity/heterozygosity

(Hom/Het) for the 4 species-diagnostic marker sets were based on the 73 markers having a specific  $G_{ST} > 0.5$  and unambiguous specific alleles. STRUCTURE analysis with the linkage option was made without *a priori* defined population, and ten replicate runs were performed for K values varying from 1 to 10. Analysis of delta K indicated that optimal results were obtained with K = 4 [Annex IV, Table S.III.4]. Low variability in estimated frequencies was observed between the ten runs for K = 4 [Annex IV, Table S.III.5]. The average values of the ten runs are shown in Figure III.3.

Nine of the ten MLGs sharing CT2 (*C. micrantha* CT) formed a main cluster (C) subdivided in three sub-clusters. Sub-cluster C1 included 3 *C. aurantifolia* accessions ('Antillaise', 'Mexican' and 'Sans épine' limes). Sub-cluster C2 joined the *C. aurata* accession ('Pomme d'Adam'), the two *C. excelsa* ('Nestour' and 'Excelsa' lime) and the *C. webberii* accession ('Kalpi' lime identical to 'Excelsa' lime). Two *C. aurantifolia* varieties were joined in C3 ('Kagzhi' and 'New Caledonia' limes). The *C. macrophylla* accession was also part of cluster C. These nine MLGs displayed very similar Structure patterns (Figure III.3) with close to half *C. medica* and *C. micrantha* contributions. The accession of sub-clusters C1 and C2 and *C. macrophylla* displayed also very similar Hom/Het patterns with most of the *C. medica* and *C. micrantha* specific alleles in heterozygosity and no other contribution (Figure III.4). All these genotypes should result from direct hybridization between *C. medica* and *C. micrantha* (or closely related) gene pools. Several SSR markers displayed heterozygous genotypes for the sub-clusters C1, C2 and *C. macrophylla* with different alleles between them. Therefore they probably resulted from 3 independent reticulation events. The analysis of the proportion of markers (over the 123) for which a *C. micrantha* x *C. medica* hybridization (based on our limited sample) should produce the genotypes of these secondary species [Annex IV, Table S.III.6] revealed higher congruency for 'Mexican' lime (95.1%) than for 'Excelsa' lime (88%) and 'Alemow' (89.3%). In addition to the heterozygous patterns observed in C1, C2 and *C. macrophylla*, the two accessions of the C3 cluster displayed small proportions of *C. medica* and *C. micrantha* specific alleles in apparent homozygosity (Figure III.3). Their origin should be therefore more complex and should result of backcross (BC) or F2 hybridization events or eventually the loss of chromosome fragments. From the 123 markers, direct *C. micrantha* x *C. medica* model displayed only 86% of congruency. The better sexual hybridization model tested was the BC one, 'Mexican' lime x Citron with 93.4% congruency that however cannot explain the homozygous *C. micrantha* alleles. For all C sub-clusters, the within sub-cluster variability was very limited and corresponded mostly to heterozygous/homozygous variations (fourteen Het/hom polymorphisms, only five Het/het and 0 Hom/Hom were observed). It may correspond to punctual mutations (and eventually genotyping errors for due to PCR drift - allele competition- for the Het/Hom variations).

The 'Kirk' lime sharing the CT2 was totally different to the other varieties at nuclear level. Indeed it combined significant contributions of the four ancestral taxa (Figure III.3) in heterozygosity (Figure III.4). It displayed respectively close to 50%, 50%, 75% and 25% *C. reticulata*, *C. maxima*, *C. medica* and *C. micrantha* specific alleles in heterozygosity. It should therefore results from the hybridization of two complementary direct interspecific hybrids. For example considering its *C. micrantha* CT, (*C. micrantha* x *C. medica*) x (*C. maxima* x *C. reticulata*) hybridization is a possibility. We tested the congruency of 'Mexican' lime x sour orange and 'Mexican' Lime x sweet orange models over the 123 markers [Annex IV, Table S.III.6] but they provided relatively low congruency (82.8 and 89.7 % respectively). However considering the diversity of the *C. maxima* and *C. reticulata* the 'Mexican' lime x (*C. maxima*/*C. reticulata*) model remain possible (93.1% congruency).

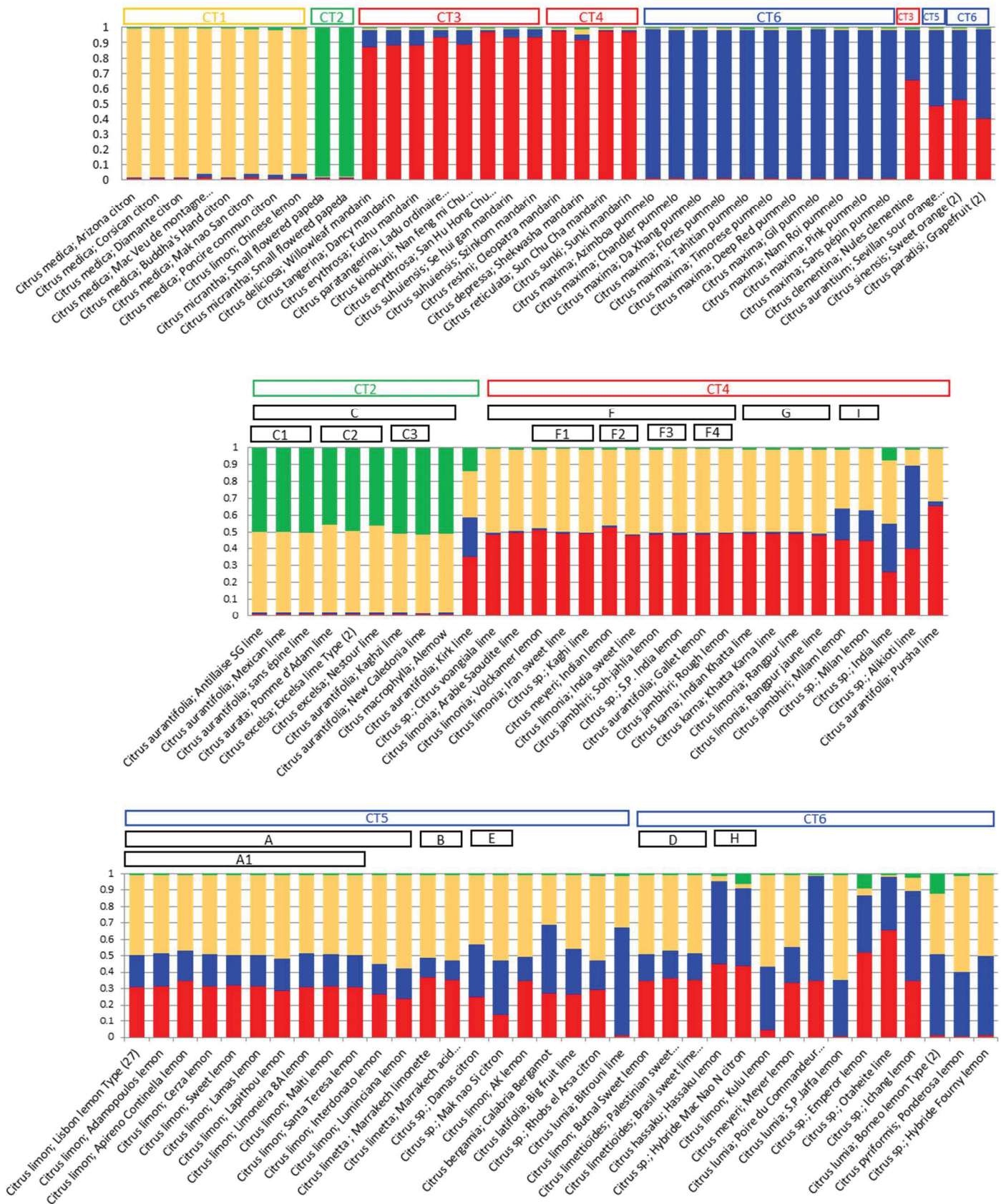


Figure III.3. Contribution (%) of the four ancestral taxa to the 103 MLGs. STRUCRURE software analyses from 73 Indel and SNP markers (average values for 10 runs with K=4). (blue: *C. maxima*; red: *C. reticulata*; green: *C. micrantha*; yellow: *C. medica*)

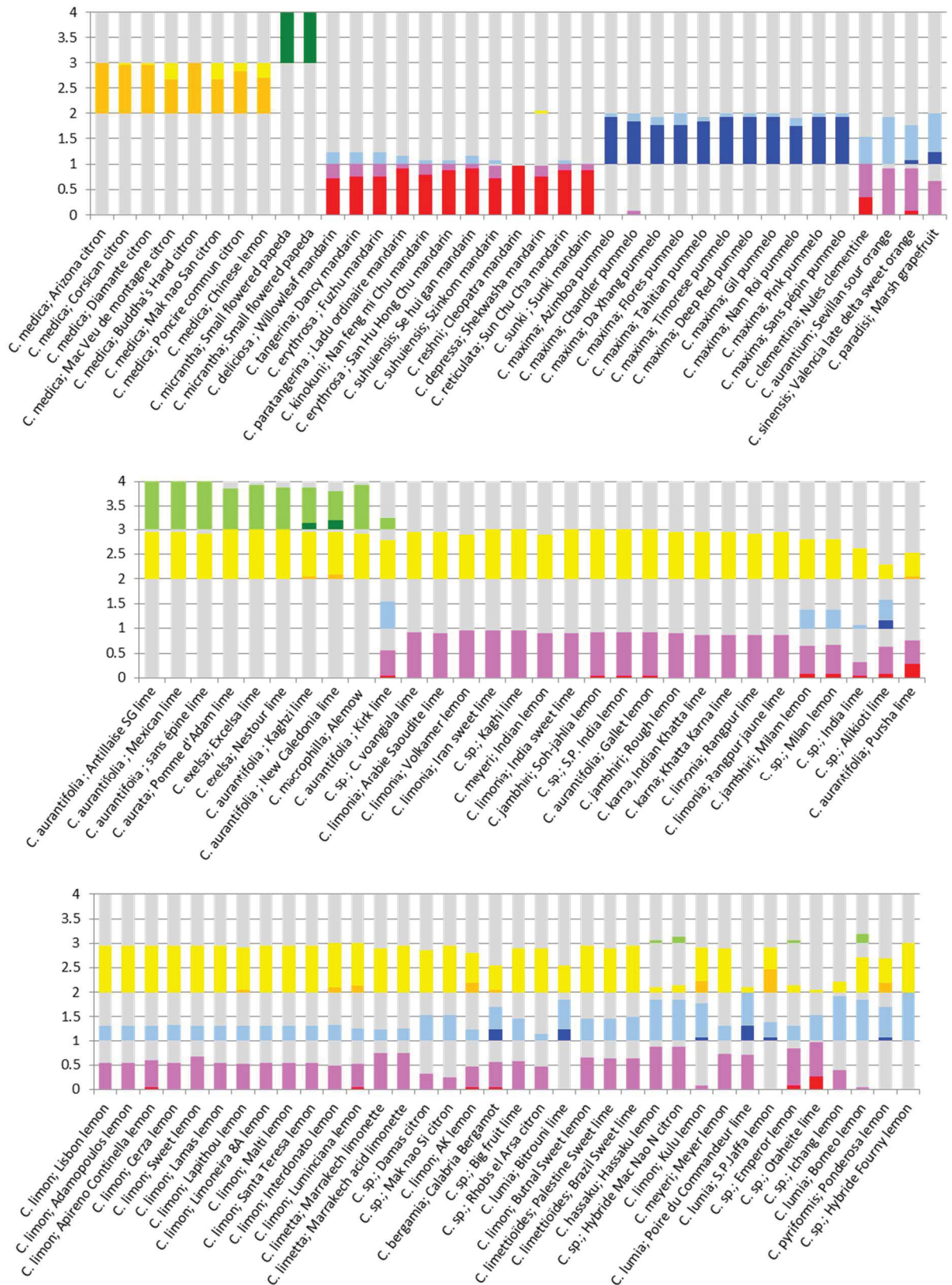


Figure III.4. Contribution of the four ancestral taxa to the 103 MLGs. Analysis of the frequency of homozygosity and heterozygosity frequency for the four specific sets of diagnostic markers. (orange: *C. medica* homozygosity; yellow: *C. medica* heterozygosity; red: *C. reticulata* homozygosity; pink: *C. reticulata* heterozygosity; deep blue: *C. maxima* homozygosity; light blue *C. maxima* heterozygosity; deep green: *C. micrantha* homozygosity; light green: *C. micrantha* heterozygosity).

The CT4 (acid mandarin cytoplasm like) displayed 3 main clusters. The first one (F) included four sub-clusters and two additional MLGs. With a lower bootstrap value the F cluster joined with the G cluster [Annex IV, Figure S.III.3]. The F and G clusters integrated all *C. limonia* ('Arabie saoudite' lime, 'Volkamer' lemon 'Iranian' sweet lime, and 'Indian' sweet lime, in cluster F, and 'Rangpur' and 'Yellow Rangpur' limes in cluster G) and *C. karna* accessions ('Indian Khatta' and 'Khata Karna' limes; in cluster G) as well as two *C. jambhiri* ('So-Jjalia' lemon and 'Rough' lemon; in cluster F), one *C. meyeri* ('Indian' lemon), one *C. aurantifolia* ('Gallet' lemon) and three *Citrus Sp.* ('Voangiata' lime, 'Kaghi' lime, and 'SP India' lemon). All these varieties displayed very similar Structure (Figure III.3) and Hom/Het diagnostic-allele (Figure III.4) patterns with close to 50/50% contribution of *C. medica* and *C. reticulata* (Figure III.3) in heterozygosity (Figure III.4). The differentiation between sub clusters and cluster F and G with heterozygous markers with different alleles suggest that each group resulted from different reticulation events of the same kind of *C. reticulata* x *C. medica* hybridization. With around 97% of congruency for the *C. reticulata* x *C. medica* model, the hypothesis is validated for 'Rangpur' lime, 'Volkamer' and 'Rough' lemons [Annex IV, Table S.III.6]. Among the different mandarins, the 'Sun Shu Cha' is the one with the adequate CT that gave the better congruencies (89.3, 94.2 and 95.1% respectively). The 5 other accessions sharing the CT4 have more complex genomic structure. Four ('Milam' and 'Milan' lemons (cluster I), 'Alikioti' and 'Pursha' limes) combine *C. reticulata*, *C. maxima* and *C. medica* contributions in variable proportions. With close to 50% *C. reticulata* contribution 'Milam' and 'Milan' lemons should result from a *C. reticulata* x (*C. maxima* x *C. medica*) hybridization. 'Alikioti' lime display homozygous *C. maxima* and *C. reticulata* alleles and only a little more than 10% of *C. medica* contribution it implies therefore a complex origin with for example one parent in *C. maxima*/*C. reticulata* heterozygosity and the other one of tri specific origin (*C. reticulata*, *C. maxima*, *C. medica*). For the 'Pursha' lime, Structure analysis inferred a 3% *C. maxima* contribution while no specific allele was observed (Figure III.3). Considering it is an artefact and in agreement with the presence of *C. reticulata* homozygosity 'Pursha' lime should result from a kind of backcross (*C. reticulata* x (*C. reticulata* x *C. medica*) or (*C. reticulata* x *C. medica*) x *C. reticulata*). Considering a real introgression of *C. maxima* it implies a more complex model (or the participation of a mandarin introgressed by *C. maxima* as already described). For the 'India' lime there is discrepancy between the structure analysis that infer *C. micrantha* contribution and the Hom/Het pattern of diagnostic SNPs. Therefore no hypothesis of origin can be proposed.

CT5 (the sour orange like CT) includes 3 main nuclear clusters. The first one (A) is only composed of *C. limon* MLGs. It includes the sub clade A1 of very close genotypes and particularly a MLG representing 27 lemons varieties (MLGCit). The second cluster (B) is branched with cluster A but at low bootstrap value. It includes the three *C. limetta* accessions ('Marrakesh acid' and à 'mamelon' limes represented by a same MLG and 'Marrakesh sweet' lime). The third cluster (E) joins two citron hybrids ('Mak Nao Si' and 'Damas'). All varieties of CT5, excepted the 'Bitrouni' lime, display a three ancestral profile (*C. reticulata*, *C. maxima* and *C. medica*). The Structure and Hom/Het patterns are very similar for all accessions of cluster A1 with relative contributions close to 49%, 31% and 20% for *C. medica*, *C. reticulata* and *C. maxima* respectively (values for the MLGCit) and heterozygosity of specific alleles. These results are in total agreement with a (*C. maxima* x *C. reticulata*) x *C. medica* hybridization. It is validated over the 123 markers with 99.2% congruency [Annex IV, Table S.III.6]. All genotypes of sub-cluster A1 are very close and considering the highly complex heterozygous interspecific structure resulting from the three-way hybridization, they could not



result from different hybridization events or by further hybridization after the yellow lemon prototype creation. These different varieties are therefore mutants or somaclonal variants that appeared during the clonal propagation. The two additional varieties of cluster A ('Interdonato' and 'Lumiciana' lemons) display a slightly higher contribution of citron (54% and 57% respectively) with specific allele in homozygosity. Interestingly the differentiation of these two lemons from the MLGCit results from homozygosity mostly for *C. medica* alleles with apparent non random distribution. Particularly 'Lumiciana' lemon displayed homozygosity for *C. medica* alleles for 5 consecutive heterozygous markers at the end of the chromosome 9, and 2 consecutive in the chromosomes 5 and 8. Such observation, while the complex structure resulting from the three-specific combination is conserved in all other genome regions and discards the possibility of sexual recombination, suggests important deletion events in these genotypes. The three *C. limetta* display similar profiles to the A cluster but with a higher proportion of *C. reticulata*. Moreover, they display heterozygosity with different alleles than in cluster A. They very probably result from the same kind of evolutive sequence than the *C. limon* of cluster A (*C. maxima* x *C. reticulata*) x *C. medica* but from different segregation in the *C. maxima* x *C. reticulata* gamete. As for yellow lemons, the congruency with the *C. aurantium* x *C. medica* model is very high over the 123 markers (98.3%; Annex IV, Table S.III.6). 'Rhobs el Arsa' citron hybrid appears to have a very similar structure and probably origin than *C. limetta*. The two accessions of cluster E ('Mak Nao Si' and 'Damas' citron hybrids) probably also resulted from a same evolution but with a higher proportion of *C. maxima* transmitted by the gamete of the *C. maxima* x *C. reticulata* parent. The 'AK' lemon displays homozygosity for *C. medica* allele and has therefore a more complex origin. *C. bergamia* displays a reduced proportion of *C. medica* (30%) and close to 50% *C. maxima* alleles as well as homozygous specific alleles of *C. maxima*, *C. reticulata* and *C. medica*. It should therefore not result from a direct interspecific hybridization. Among the three model tested with the 123 markers [Annex IV, Table S.III.6], the *C. limon* ('Lisbon' cv) x *C. aurantium* ('Sevilla' cv) was the better one (96.7 congruency). With 67% and 31% contribution of *C. maxima* and *C. medica* constitution and homozygous *C. maxima* alleles, the 'Bitrouni' lime should result from a kind of backcross *C. maxima* x (*C. maxima* x *C. medica*).

CT6 displayed two main nuclear clusters. The first one (D) associate the three *C. limettioïdes* accessions ('Palestinian' and 'Brazil' sweet limes, 'Bisri' lime, the last two varieties having the same MLG, and 'Butnal' sweet lemon). These three varieties have three species nuclear structure (*C. reticulata*, *C. maxima* and *C. medica* in heterozygosity) similar to the yellow lemon ones. With close to 50% *C. medica* contribution they are probable (*C. maxima* x *C. reticulata*) x *C. medica* hybrids but with different *C. maxima* x *C. reticulata* parent than *C. limon* or *C. limetta* as testified by the CT. 'Meyer' lemon displays similar structure and Hom/Het pattern than the *C. limettioïdes* and has probably a similar (*C. maxima* x *C. reticulata*) x *C. medica* origin. Such origin is confirmed over the 123 markers with relatively 98.3 and 95.8% congruencies [Annex IV, Table S.III.6]. However, none of the tested varieties with *C. reticulata*/*C. maxima* constitution (sweet orange, 'Poire du commandeur', 'Hassaku' lemon, 'Ichang' lemon, 'Otaheite' lime, [Annex IV, Table S.III.6]) and CT6 provided conclusive results. The second nuclear cluster (H) joins the 'Mak Nao N' and the *C. hassaku* accessions. Both have complex genome with mainly *C. maxima*/*C. reticulata* contribution in heterozygosity but also limited *C. medica* and *C. micrantha* contributions. Most of the other varieties sharing this CT displayed complex structures with three or four ancestral taxa contributions. However 'Jaffa' lemon, *C. pyryformis* ('Ponderosa')

and 'Hybride fourny' lemon display only *C. maxima/C. medica* contributions. 'Hybride Fourny' lemon should be a direct hybrid (96.7% congruency over 123 markers; Annex IV, Table S.III.6) while 'Ponderosa' and 'Jaffa' lemons displayed *C. maxima* and *C. medica* homozygous alleles suggesting a more complex origin (eventually a F2 like origin). The varieties classified as *C. lumia* appears to have very different origins; indeed *C. lumia* 'Poire du commandeur' has mostly a *C. reticulata/C. maxima* constitution (94.2% congruency with the *C. maxima* x *C. reticulata* model over the 123 markers; [Annex IV, Table S.III.6]) while 'Jaffa' lemon has a *C. medica/C. maxima* constitution and 'Borneo' lemon a *C. maxima / C. medica / C. micrantha* constitution.

The global genetic structuration appears relatively disconnected from the 2 main taxonomic classifications of limes and lemons as confirmed by the F stat analysis (Table III.3). Indeed  $F_{ST}$  values for Swingle and Reece and Tanaka are respectively very low (0.08) and low (0.18) while the high negative  $F_{IS}$  values (0.30 and 0.48, respectively) probably testify for fixed heterozygosity within species due to vegetative propagation. The F parameters, considering the main genetic clusters of our study as subpopulation, revealed an improved intergroup organization compared with classical taxonomic classifications ( $F_{ST}=0.323\pm 0.034$  and  $F_{IS}=-0.82\pm 0.057$ ).

Table III.3. Genetic structuration of lime and lemon germplasm according to Tanaka (1961), Swingle and Reece (1967) and our cluster analysis.

	$F_{ST}$	$F_{IS}$	$F_{IT}$
Tanaka subdivision	0.179±0.021	-0.48±0.051	-0.21±0.056
Swingle subdivision	0.085±0.021	-0.30±0.054	-0.18±0.059
Genetic clusters	0.323±0.034	-0.82±0.057	-0.25±0.066

### Nuclear diversity at triploid and tetraploid level

Ten triploid and one tetraploid limes belonging to *C. aurantifolia* and *C. latifolia* species where analysed. For SSR and Indel heterozygous genotypes, allele dosage was calculated using the MAC-PR (microsatellite DNA allele counting-peak ratio) method (Esselink *et al.*, 2004), validated in citrus by Cuenca *et al.*, (2011). For SNPs it was estimated from relative allele signal according to the method described by Cuenca *et al.* (2013b). An example is given (Figure III.5) for the 9P25060404 AG SNP marker displaying a citron specific allele (G). Among the four triploid limes in the considered plate, an AGG genotype was inferred for 'Tanepao' and 'Coppenhad' limes and AAG for 'Tahiti' and 'Persian' lime. When it was not possible to infer allele doses for heterozygous genotypes, data were considered as missing. Missing data ranged from 4% to 30% with an average of 11% for SSRs and indels and between 0 and 6% for SNPs with 2% as average (Table III.4).

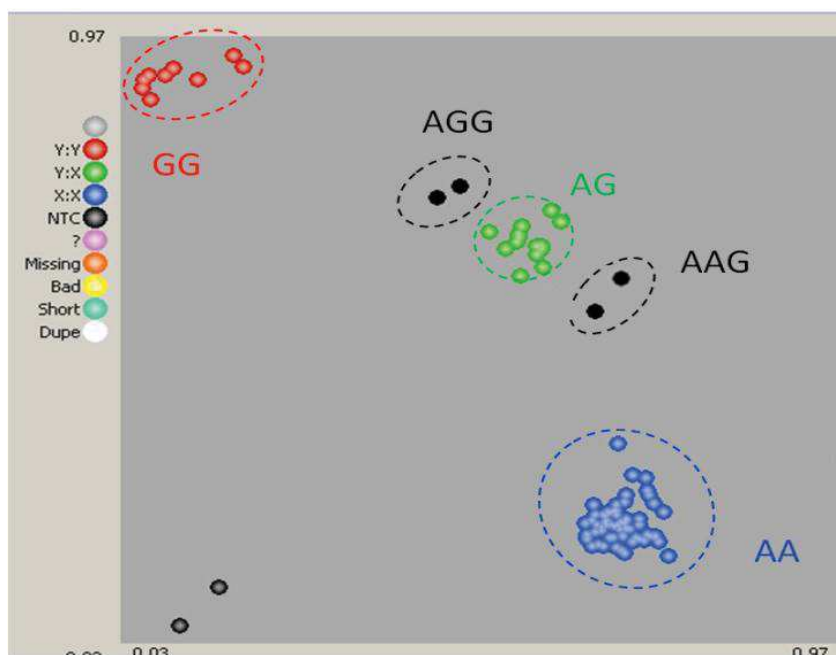


Figure III.5. Pattern of relative allele fluorescence for the 9P25060404 SNP marker and inference of allele doses for four triploid limes. AGG: ‘Tanepao’ and ‘Coppennhad’ limes; AAG: ‘Tahiti’ and ‘Persian’ limes).

Among the SSRs and indels, the *C. latifolia* and *C. aurantifolia* accessions, displayed relatively 29% and 20% homozygous markers, 49% and 37% heterozygous markers with two alleles and 22% and 43% heterozygous markers with three alleles. *C. latifolia* varieties were also more heterozygous for the diallelic SNP markers with 54% of heterozygous markers in average compared with 0.42 for the triploid *C. aurantifolia*. For the tetraploid ‘Giant Key’ lime, no markers displayed three alleles and the average heterozygosity was 0.41. For SSRs and indels, the relative peaks ratio of ‘Giant Key’ lime were equivalent to the ‘Mexican’ lime ones and the relative allele florescence for SNPs were also similar to the ‘Mexican’ lime ones. Therefore it was considered that the allele doses at heterozygous loci were 2/2. With these dose inferences ‘Giant key’ appears to be a probable doubled diploid of a *C. aurantifolia* ‘Mexican’ like type.

Table III.4. heterozygosity and multi-allelism of the polyploid limes and lemons

Tanaka classification names	Common name	Ploidy	N	SSR and Indels			SNPs		
				Hom	Di	Tri	N	Hom	Het
<i>C. aurantifolia</i>	Madagascar lemon	3	0.15	0.26	0.57	0.17	0	0.57	0.43
<i>C. aurantifolia</i>	Ambilobe lime	3	0.15	0.3	0.48	0.22	0.01	0.59	0.41
<i>C. aurantifolia</i>	Coppennhad lime	3	0.07	0.28	0.52	0.2	0.02	0.56	0.44
<i>C. aurantifolia</i>	Mohtasseb lime	3	0.3	0.32	0.42	0.26	0.02	0.6	0.4
<i>C. aurantifolia</i>	Tanepao lime	3	0.11	0.29	0.46	0.25	0.02	0.56	0.44
<i>C. latifolia</i>	Bears lime	3	0.07	0.2	0.36	0.44	0.02	0.46	0.54
<i>C. latifolia</i>	Persian lime	3	0.07	0.2	0.36	0.44	0	0.46	0.54
<i>C. latifolia</i>	El Kseur lime	3	0.11	0.21	0.38	0.42	0.06	0.46	0.54
<i>C. latifolia</i>	IAC-5 lime	3	0.11	0.21	0.38	0.42	0.02	0.46	0.54
<i>C. latifolia</i>	Tahiti lime	3	0.07	0.2	0.36	0.44	0.03	0.46	0.54
<i>C. aurantifolia</i>	Giant Key lime	4	0.04	0.19	0.81	0	0.02	0.59	0.41

The hierarchical classification of the 10 triploid limes displayed two strong clusters corresponding to the *C. latifolia* and *C. aurantifolia* species of Tanaka (Figure III.6). The diversity within each cluster appeared very scarce.

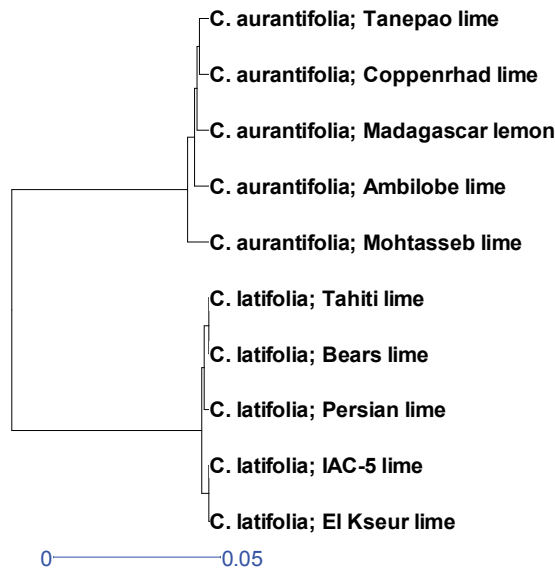


Figure III.6. Hierarchical classification of triploid limes based on 123 genetic markers.

The frequency for homozygosity, 2/3, 1/3 and 0 doses for the diagnostic alleles of the four specific marker sets (total of 73 markers) were evaluated (Figure III.7). None of the specific alleles were found in homozygosity. The five triploid *C. aurantifolia* displayed very similar patterns with most of the *C. medica* diagnostic alleles founded with double doses and *C. micrantha* allele in single doses (Figure III.7). Such pattern and the *C. micrantha* like cytoplasm suggest that these triploid limes should result from an interspecific backcross origin (*C. micrantha* x *C. medica*) x *C. medica* with a double contribution (2x gamete) of the interspecific hybrid. A F2 origin was discarded by the presence of triallelic SSR and indel markers. These triallelic markers also suggested that, considering the BC model, two different *C. medica* genotypes were implied in their origin. We tested the hypothesis of the interspecific BC model by estimating the proportion of the 123 markers that can generate the triploid *C. aurantifolia* genotypes from a 2x ‘Mexican’ lime gamete and a x citron gamete (from the citron population). 96.3% to 98.8% markers fitted with the model [Annex IV, Table S.III.6].

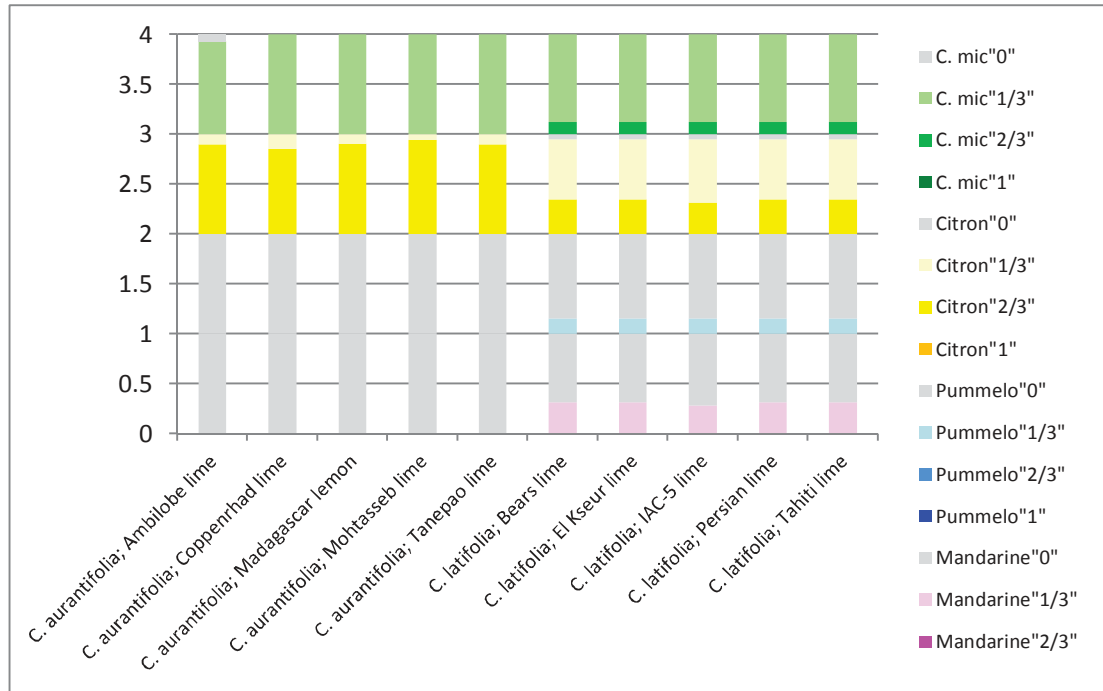


Figure III.7: Contribution of the four ancestral taxa to the 10 triploid limes. Analysis of the frequency of homozygosity and heterozygosity doses 2/3, 1/3 for the four specific sets of diagnostic markers.

The *C. latifolia* accessions displayed a much more complex genomic structure with a contribution of the four basic taxa. Specific alleles of *C. medica* and *C. micrantha* were found in single (respectively around 60% and 87%) or double doses (respectively 35% and 13%) and some *C. reticulata* (32%) and *C. maxima* (15%) specific alleles were observed in single doses. Considering that *C. latifolia* accessions share the sour orange / yellow lemon CT, a ((*C. maxima* x *C. reticulata*) x *C. medica*) x *C. (micrantha* x *C. medica*) model with a 2x male gamete should explain the genomic structures of the *C. latifolia* accessions. We tested this hypothesis considering a *C. limon* ('Lisbon' lemon) x *C. aurantifolia* ('Mexican' lime) model (*C. limon* cv 'Lisbon' being a potential ((*C. maxima* x *C. reticulata*) x *C. medica*) hybrid and *C. aurantifolia* cv 'Mexican' lime a *C. micrantha* x *C. medica* hybrid. 99% of the 123 markers fitted with this model [Annex IV, Table S.III.6]. Moreover all specific *C. maxima* and *C. reticulata* alleles observed in the *C. latifolia* accessions were also found in *C. limon* cv 'Lisbon'.

## Discussion

### **The lime and lemon citrus horticultural group is genetically highly complex, involving four ancestral species with diploid, triploid and tetraploid compartments**

While the other main citrus horticultural groups (sweet oranges, sour oranges, grapefruits, small citrus –mandarins, tangors and tangelos-) result only from *C. reticulata* and *C. maxima* gene pools at diploid level (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b), the genomic structure and the origin of limes and lemons appeared much more complex.

Limes and lemons varieties shared four of the six cytoplasmic types identified within the *Citrus* genus, testifying for four maternal phylogenetic origins. Most *C. limon*, all *C. limetta* and *C. latifolia* as well as one *C. lumia* and several unclassified varieties had identical CT than *C. aurantium*. Earlier molecular marker studies also concluded for the CT identity between *C. aurantium* and “yellow lemons” (Nicolosi *et al.*, 2000; Bayer *et al.*, 2009; Froelicher *et al.*, 2011; Curk *et al.*, in press). Froelicher *et al.* (2011), also shown that *C. aurantium*, “yellow lemons” and triploid limes (‘Bears’ and ‘Tahiti’) shared the same mitotype. All *C. limettioides*, most *C. lumia*, 1 *C. meyeri*, the *C. pyryformis*, *C. hassaku* and several unclassified varieties shared a *C. maxima* CT also found in sweet orange and grapefruit. CpDNA (Nicolosi *et al.*, 2000) and mitochondrial analysis (Froelicher *et al.*, 2011) previously shown that *C. limettioides* and *C. meyeri* had a *C. maxima* maternal phylogeny. All *C. limonia*, *C. jambhiri*, *C. karna*, two *C. aurantifolia* and one *C. meyeri* as well as 6 unclassified varieties shared the CT of non edible mandarins. The differentiation between two main mandarin CTs as well as the association of ‘Rangpur’ lime, ‘Rough’ and ‘Volkamer’ lemons with the ‘Cleopatra’ mandarin one was previously described at mitochondrial level (Froelicher *et al.*, 2011). The *C. micrantha* CT was found for most *C. aurantifolia* and all *C. excelsa*, *C. webberii*, *C. aurata* and *C. macrophylla* varieties. The high similarity of ‘Mexican’ lime and ‘Alemow’ CT with the *C. micrantha* one were already proposed from chloroplastic (Nicolosi *et al.*, 2000; Bayer *et al.*, 2009; Penjor *et al.*, 2013; Curk *et al.*, in press; Curk *et al.*, submitted b) and mitochondrial (Froelicher *et al.*, 2011) marker studies.

Natural ploidy variation among varieties observed in limes is also uncommon in citrus. Indeed, while spontaneous tetraploid plants (doubled diploids) occurs in seedling of diploid polyembryonic genotypes (Aleza *et al.*, 2011) and 2n gametes are relatively frequent (Esen and Soost, 1971; Ollitrault *et al.*, 2008) natural polyploid genotypes are very rare. Triploidy of the *C. latifolia* varieties was already described (Bacchi, 1940) and ‘Giant Key’ lime was known to be a spontaneous tetraploid selected in a seedling of the diploid ‘Key’ lime (‘Mexican’ lime type) in Florida in 1973 by HC Barrett (US Horticultural Research Laboratory, Orlando). The triploidy of the ‘Tanepao’ like varieties was unknown and unsuspected because these triploid varieties produce seeds while triploidy is largely associated with sterility and seedlessness in citrus (Ollitrault *et al.*, 2008).

Based on 123 codominant markers (SSRs, Indels and SNPs) the nuclear analysis revealed generalized high heterozygosity of limes and lemons MLGs ( $H_o=0.49\pm 0.05$  in average) when compared with the one of the basic taxa ( $0.13\pm 0.03$ ). This confirms their probable interspecific origin, has already pointed out by several molecular studies (Herrero *et al.*, 1996; Federici *et al.*, 1998; Gulsen and Roose, 2001a; Barkley *et al.*, 2006). All individuals of a same nuclear cluster shared the same CT. Nine main clusters (bootstrap values over 90%) of limes and lemons MLGs was revealed by NJ analysis. Two of them (C and F) were subdivided with high bootstrap values (between 98 and 100%) on several sub-clusters (three and four respectively). Inter sub-cluster variability revealed several loci heterozygous in the different sub-clusters but with different allelic constitutions suggesting that they arose from different hybridization events. The cluster A displayed a sub-cluster A1 of highly related genotypes (including an MLG representative of 27 *C. limon* varieties), and two *C. limon* varieties a little more differentiated. Considering that the differentiation was only due to the loss of heterozygosity in these two varieties compared with the highly redundant MLG of sub-cluster A1 (Lisbon lemon type) and the complex tri-specific constitution of cluster A (*C. medica*/*C. reticulata*, *C. maxima*) we considered probable that all MLGs of cluster A derived from a same hybrid ancestor without additional sexual event. Nineteen MLGs remained alone without strong clustering with others limes or lemon genotypes and may have originated from independent hybridization events. Therefore we hypothesize that the analyzed limes and lemons arose from 36 independent reticulation events (cluster A, B, D, E, G, H, I; sub-clusters C1, C2, C3, F1, F2, F3, F4, 1 MLGs of cluster C, 2 MLG of cluster F and 19 'unclustered' MLGs). It points out that an essential part of lime and lemon intra horticultural group diversity results from different sexual recombination events, while for two other main secondary species and corresponding horticultural groups (*C. sinensis*, the sweet oranges, *C. paradisi*, the grapefruits) all actual varieties originated from a single clonal parent via a series of mutations or somatic variations (Barkley *et al.*, 2006; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b).

The identification of 73 markers over the 123 with high phylogenetic values (differentiation of one basic taxa from the three others) and further STRUCTURE analysis and study of the frequency of homozygous and heterozygous specific alleles for the four set of ancestral taxa diagnostic markers, provided information in the interspecific admixture of each lime and lemon genomes. While *C. medica* CT was not found in limes and lemons, the contribution of citron at nuclear level appeared essential and will be discussed further. Numerous limes and lemons MLGs appeared to be issued from direct hybridization between two ancestral taxa. *C. micrantha* x *C. medica* should be the model for 9 diploid MLGs and *C. reticulata* x *C. medica* for 15 MLGs. Four MLGs displayed only *C. medica* and *C. maxima* contribution but only the 'Hybride de Fourny' hybrid was found to be a potential direct hybrid between these two species. A three-specific origin (*C. maxima* x *C. reticulata*) x *C. medica* was coherent with 20 MLGs patterns. Contributions of these three species but probably from more complex hybridization schemes have been observed for several limes and lemons and particularly *C. bergamia*. By the end, five MLGs at diploid level and *C. latifolia* varieties at triploid level displayed a contribution of the four ancestral taxa. The origin of the main lime and lemon groups are further discussed in more detail.

The geographical distribution of the centers of origin of the different citrus species proposed by Malik *et al.* (1974) supports the assumption that most of the prototypes of limes and lemons should have originated in Asia where the diversification areas of the supposed

parent overlapped. Neither Swingle and Reece (1967) and Tanaka (1954) botanical classifications, nor the usual lime and lemon denomination correctly encompass the organization of the genetic diversity. It is logical considering for one side the too much simplified dichotomy of Swingle and Reece (*C. limon/C. aurantifolia*) or usual denomination (lemon/lime) while much more different phylogenetic origin are involved and for the other side the multiplication of taxa proposed by Tanaka separating sub-groups of similar origin (ei: *C. karna/C. jambhiri/C. limonia* or *C. limon/C. limetta*). It is also possible that some varieties were incorrectly attributed to one taxa in the germplasm depositories, contributing to decrease the estimation of inter taxa genetic differentiation.

### **Citron is the common genetic contributor of limes and lemons but never acted as female parent**

With the exception of the 'Poire du commendeur' and 'Othaheïte' lime all other lime and lemon like accessions analysed displayed a contribution of the *C. medica* genome. The first variety is classified as *C. lumia* but display very few characters of limes and lemons. The second one was unclassified and despite its usual lime appellation, it looks more like a mandarin than a lime. For other lime and lemon varieties, citrons appeared mainly as a direct parent (for 11 over the 14 identified clusters and independent sub-clusters and 6 of the non clusterized MLGs). However, none of the analysed lime and lemon varieties displayed a *C. medica* maternal phylogeny. Therefore we can conclude that citron acted as male parent. Our results expand to much more lime and lemon varieties than the previous hypothesis made by several early taxonomists on the relationship between citron, limes, and lemons (see the introduction). They also agree with more recent conclusions obtained by combining cytoplasmic and nuclear markers analysis (Nicolosi et al., 2000; Gulsen and Roose, 2001a; Abkenar et al., 2004; Barkley et al., 2006; Bayer et al., 2009; Jena et al., 2009; Froelicher et al., 2011; Curk et al., in press). Many authors consider that *C. medica* have an Indian origin (Malik et al., 1974; Scora, 1975; Jena et al., 2009). However chloroplastic phylogenetic studies leaded Beattie et al. (2008) to propose an Australasian origin for *C. medica*. Indeed their closest relatives were *Clymenia polyandra* of Papua New Guinea, *Oxanthera spp.* of New Caledonia and *Microcitrus* and *Eremocitrus* species of Papua New Guinea and Australia. They proposed that early species dispersed westward – possibly as floating fruit – some 30-35 million years ago from northeastern Australia probably aided by equatorial currents. Whatever the initial geographical origin, India is clearly a diversification centre for citron (Malik et al., 1974) and cytogenetic studies demonstrated that *C. medica* is a true species parental of limes and lemons (Carvalho et al., 2005). Indeed in a study of citron, limes and lemons Carvalho et al. (2005) have shown that *C. medica* was the only cytogenetically homozygous accession and that all of its chromosome types were clearly represented in limes and lemons, some of them forming heteromorphic pairs. Moreover, the lemons and limes were heterozygous for all rDNA sites, whereas *C. medica* was entirely homozygous. Of the citrus groups that are thought to be true *Citrus* species, the citrons had the lowest observed heterozygosity and diversity as previously observed in several molecular studies (Barkley et al., 2006; Ollitrault et al., 2011; Garcia-Lor et al., 2012). This low polymorphism may be explained by the cleistogamy of citron flowers. Indeed selfing, increase homozygosity. Barrett and Rhodes (1976) reported that citrons produce vigorous selfed seedlings and tend to be highly homozygous, which is consistent with our data. Such cleistogamy behaviour can also explain that none of the limes and lemons that display direct



relationship with *C. medica* have citron as female parent, while it is a totally monoembryonic species (no apomixes).

### Global phylogenomic structure and origin of the main lime and lemon groups are revealed

An incomplete congruency was found between the Tanaka classification and the nuclear interspecific patterns and inferred origins. Our conclusions are schematized in Figure III.8 and discussed below for the main groups of lemons and limes previously debated in the citrus literature.

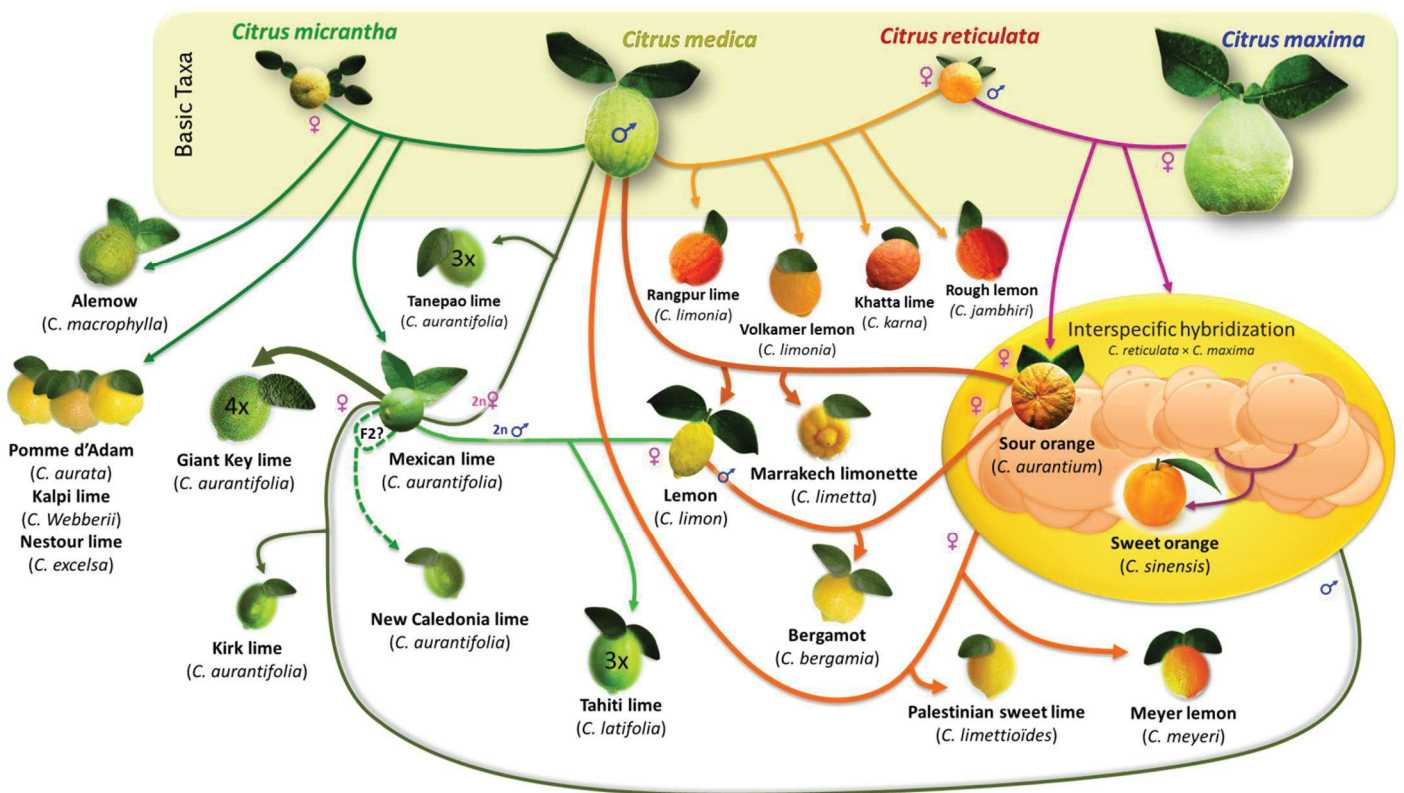


Figure III.8: Origin of the main lime and lemon varietal sub-groups

#### *C. micrantha* x *C. medica*

*C. micrantha* x *C. medica* should be the model for *C. macrophylla*, *C. aurata*, *C. webberii*, *C. excelsa* and diploid *C. aurantifolia*, including the widely cultivated ‘Mexican’ lime. For ‘Mexican’ lime like cultivars our conclusions agree with the hypothesis of a papeda × citron hybridization and more specifically *C. micrantha* × *C. medica* hybridization proposed by several authors from biochemical data (Scora, 1975), molecular markers (Nicolosi *et al.*, 2000; Yamamoto *et al.*, 2007; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b) and cytogenetic observations (Carvalho *et al.*, 2005). Our results discard the hypothesis of a trihybrid intergeneric cross involving *C. medica*, *C. grandis*, and a species of *Microcitrus* proposed by Barrett and Rhodes (1976) or of a direct hybridization between *C. medica* and

*C. grandis* proposed by Liang *et al.* (2007) from an AFLP study. The origin of 'Alemow' (*C. macrophylla* in Tanaka classification) is more controversial. Swingle and Reece (1967) considered *C. macrophylla* to be a hybrid of *C. celebica*, or some other species of the subgenus *Papeda*, with a species of the subgenus *Citrus*, probably *C. maxima*. Tanaka placed it in the section *Limonellus* along with *C. aurantifolia*. Federici *et al.* (1998) found that *C. macrophylla* clustered with *C. aurantifolia*, and the papedas *C. hystrix* and *C. micrantha* and Nicolosi *et al.* (2000) mentioned that 'Mexican' Lime and *C. macrophylla* had similar *C. medica*/*C. micrantha* constitution. A possible *C. micrantha* × *C. medica* origin was also proposed from SNPs data by Ollitrault *et al.* (2012b). Very few or no data are available on the origin of the other Tanaka species that should have a similar *C. micrantha* (or closely related papeda) × *C. medica* origin. For the 'New Caledonian' and 'Kaghzi' lime, a F2 (*C. micrantha* × *C. medica*) × (*C. micrantha* × *C. medica*) fit better, than a direct cross between ancestral taxa, to the observed pattern with homozygous *C. medica* and *C. micrantha* specific alleles.

### ***C. reticulata* × *C. medica***

*C. reticulata* × *C. medica* is proposed as model from our data for *C. limonia*, *C. jambhiri*, *C. karna*, one *C. aurantifolia* and one *C. meyeri*. The two last accession correspond eventually to erroneous classification. The origins of 'Rough' lemon (*C. jambhiri*), 'Volkamer' lemon (*C. volkameriana*) and 'Rangpur' lime (*C. limonia*) have been widely debated. Most of the authors recognized mandarin as one parent of 'Rangpur' lime but combined with different second parents: lime (Webber, 1943; Tatum *et al.*, 1974; Barkley *et al.*, 2006), sour orange (Barrett and Rhodes, 1976), 'Rough' lemon (Handa and Oogaki, 1985) and citron as the conclusion of our study (Federici *et al.*, 1998; Gulsen and Roose, 2001a; Li *et al.*, 2007). Scora (1975) suggested that the 'Rough' lemon was a natural hybrid of a mandarin and a citron and several more recent molecular studies agreed with this hypothesis ((Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Gulsen and Roose, 2001a; Barkley *et al.*, 2006; Ollitrault *et al.*, 2012b). The origin of 'Volkamer' lemon was more controversial. Barrett and Rhodes (1976) thought mandarin × sour orange a possible parentage. Nicolosi *et al.* (2000) considered that citron and sour orange were the ancestors. Carvalho *et al.* (2005) considered from cytogenetic evidence that 'Volkamer' lemon have a mandarin × citron origin, similar to 'Rough' lemon and 'Rangpur' lime. Our results agree with the last hypothesis and both cytoplasmic and nuclear data display a very close relationship between 'Rough' Lemon and 'Volkamer' lemon. Cytogenetic studies also provided evidences for a mandarin × citron origin of 'Volkamer' lemon, 'Rough' lemon and 'Rangpur' lime (Carvalho *et al.*, 2005). Moreover, these authors observed that one half of the 'Rough' lemon and 'Rangpur' lime karyotypes was identical to the haploid chromosome complement of *C. medica*, while the other half chromosome set perfectly matched the haploid complement of the 'Cleopatra' mandarin. This last observation is also coherent with our cytoplasmic data showing that 'Rough' lemon, 'Volkamer' lemon and 'Rangpur' limes shared the CT of acid mandarins ('Cleopatra', 'Sunki', 'Shekwasha' and 'Sun Chu Cha').

### **(*C. maxima* × *C. reticulata*) × *C. medica***

A three-specific origin (*C. maxima* × *C. reticulata*) × *C. medica* appeared as the origin of *C. limetta*, *C. limettioïdes*, *C. meyeri* CV 'Meyer' lemon and most *C. limon* varieties. Cytoplasmic data and analysis of the fit for different concrete parental models suggested that a *C. aurantium* × *C. medica* should be at the origin of *C. limetta* and most *C. limon* varieties while other parents derived from *C. maxima* and *C. reticulata* gene pools were involved for *C.*

*limettioïdes* and *C. meyeri*. Contributions of *C. maxima*, *C. reticulata* and *C. medica* with probable more complex hybridization schemes have been observed for several limes and lemons and particularly *C. bergamia*. While Scora (1975) proposed that by hybridization with the lime, citron gave rise to the lemon more recent molecular and cytogenetic studies (Nicolosi *et al.*, 2000; Gulsen and Roose, 2001a; Carvalho *et al.*, 2005; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b; Ramadugu *et al.*, 2013) agree that the 'yellow lemon' types arisen from *C. aurantium* x *C. medica* hybridization as validated by the very high fit (>99%) of our data with this model. *C. limetta* varieties probably originated in the Mediterranean Basin (Webber *et al.*, 1967) where sour orange and citron were present for a long time and our results suggested that they have the same phylogenetic origin than the 'yellow lemons'. Many hypotheses have been proposed for the origin of *C. limettioïdes*. *C. aurantifolia* was frequently proposed as one of the parents. *C. aurantifolia* by *C. limetta* or citron (Webber, 1943), *C. aurantifolia* by *C. sinensis* (Barrett and Rhodes, 1976); *C. aurantifolia* x *C. medica* (Carvalho *et al.*, 2005). Nicolosi *et al.* (2000) proposed citron and sweet orange as putative male and female parents, respectively. Our cytoplasmic data totally discarded the possibility that *C. aurantifolia* would be the female parent of *C. limettioïdes*. Moreover no evidence for *C. micrantha* contribution (one direct parent of *C. aurantifolia*) was provided by the nuclear analysis. Therefore, we consider that a *C. aurantifolia* contribution to *C. limettioïdes* is improbable. The *C. sinensis* x *C. medica* model did not fit correctly with our data (85%) but 98% of the 123 markers provided coherent patterns for a (*C. maxima* x *C. reticulata*) x *C. medica* model. Scora (1975) and Gulsen and Roose (2001a) proposed that 'Meyer' lemon (*C. meyeri*) should be a sweet orange by citron hybrid. Our cytoplasmic results agree with this hypothesis but the nuclear fit for this model was low (86%). Therefore *C. meyeri* Cv 'Meyer' lemon is a probable (*C. maxima* x *C. reticulata*) x *C. medica* hybrid (fit of 96% with our data) but with unidentified concrete parents. The origin of bergamot (*C. bergamia*) was also controversial. Gallesio (1811) proposed a sour orange x lemon origin while several molecular studies disagree and proposed hybridization between citron and lime (Chen *et al.*, 1991), between sour orange and a sweet lime (Herrero *et al.*, 1996; Federici *et al.*, 2000), or sour orange and a citron (Li *et al.*, 2010; Nicolosi *et al.*, 2000). The different hypotheses were confronted with our data and the best fit was observed for the *C. limon* x *C. aurantium* model (97%) that also agree with the cytoplasmic data.

### **Triploid lime origins**

Previous cytoplasmic studies showed that *C. latifolia* ('Tahiti' lime like varieties) shared the same cytoplasmic pattern than *C. limon* and *C. aurantium* (Bayer *et al.*, 2009; Froelicher *et al.*, 2011). However, to our knowledge this is the first molecular analysis of the nuclear structure that allows proposing a hypothesis on the origin of these triploid limes. The five *C. latifolia* analysed displayed very similar patterns with a contribution of the four ancestral taxa (*C. maxima*, *C. medica*, *C. micrantha* and *C. reticulata*) with all specific *C. medica* and *C. micrantha* allele in single or double doses while *C. maxima* and *C. reticulata* were in single dose or absent. According to the cytoplasmic and nuclear pattern we propose that 'Tahiti' lime like accessions resulted from the fertilization of a haploid lemon ovule by a diploid 'Mexican' lime like gamete. 99% of the 123 markers analysed were in accordance with this model. This diploid gamete should be originated from a natural doubled diploid of 'Mexican' like lime, as the 'Giant key' lime selected in a seedling of 'Key' lime in 1973 by HC Barrett, US Horticultural Research Laboratory, Orlando. It is also possible that the diploid *C. aurantifolia* diploid gamete was an unreduced gamete from a diploid variety. Reece and

Childs (1962) previously proposed from morphological trait segregation studies in 'Tahiti' lime seedlings that this variety should result from lime by citron or lemon hybridization but did not recognize the triploid status of 'Tahiti' lime. The identity of 'Persian' lime, 'Tahiti' lime and 'Bears' lime is explained by the diffusion of this variety (Morton, 1987). It is believed that the 'Tahiti' like lime was introduced into the Mediterranean region through Iran ('Persian' lime). Portuguese traders probably carried it to Brazil, and it was apparently taken to Australia from Brazil about 1824. It reached California from 'Tahiti' between 1850 and 1880 and had arrived in Florida by 1883. According to Webber (1943), the 'Bears' variety originated about 1895 on the place of J.T. 'Bears', a nurseryman at Porterville, California presumably as a seedling of a tree grown from seed from a fruit of 'Tahitian' origin.

We also revealed a second group of triploid lime ('Tanepao', 'Coppenhad', 'Ambilobe' and 'Mothasseb' limes and 'Madagascar' lemon) with a different phylogenetic origin. Indeed, very probably, these varieties displaying a *C. micrantha* CT and only *C. medica* and *C. micrantha* contribution at nuclear level with mostly double doses of *C. medica* specific allele and single one for *C. micrantha* resulted from a (*C. micrantha* x *C. medica*) x *C. medica* hybridization with a diploid gamete of the *C. micrantha* x *C. medica* parent. Tested with the 'Mexican' lime as genotype for the *C. micrantha* x *C. medica* parent, our data fitted for 96.3% of the 123 markers.

### **Asexual variations are important source of phenotypic variability in the apomictic limes and lemon groups**

All limes and lemons varieties display partial apomixes (polyembryonic seeds with nucellar embryos). Previous molecular (Gulsen and Roose, 2001b) and cytogenetic studies (Carvalho *et al.*, 2005) have shown that numerous lemon cultivars originated from a single clonal parent via a series of mutations. Similar conclusions were proposed by Snoussi *et al.* (2012) for several limes and lemons from a survey of Tunisian citrus germplasm. Our results lead to the same conclusions for the polymorphism observed within the different identified sub-clusters. Indeed the varieties were highly heterozygous and displayed very limited diversity within the sub-clusters, generally in the form of homozygous/heterozygous polymorphisms with a common allele. Such patterns should not be obtained from sexual hybridization of the highly heterozygous prototypes of the subgroups. Their occurrence from parallel reticulation events was totally discarded for sub-groups having a tri specific origin. It appears also improbable for the sub-groups issued of direct hybridization between two ancestral taxa, considering the diversity and heterozygosity within these taxa. Punctual mutations and transposable elements have been proposed as the source of diversity within citrus groups propagated vegetatively by apomictic seeds or by grafting (Bretó *et al.*, 2001; Barkley *et al.*, 2006; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b). In our study the loss of the non citron allele for five consecutive markers at the end of chromosome 9, two in chromosome 5 and 8, while the rest of its tri-specific genomic structure was conserved, suggested deletions events at the origin of the 'Lumiciana' lemon. Mutation or epigenetic variations were previously described as major factor of diversification in several apomictic species (Hörandl and Paun, 2007; Nybom, 2007). The contribution of somatic mutations to the evolution of other vegetatively propagated crops, such as grapes (Crespan, 2004), olives (Cipriani *et al.*, 2002), yams (Scarcelli, 2005) and cassava (Sardos *et al.*, 2008), has also been demonstrated. In addition to sexual recombination, human selection of new phenotypes and further clonal propagation are also key factors generating the inter-varietal phenotypic polymorphism within horticultural-groups (Barrett and Rhodes, 1976; Ollitrault *et al.*, 2003).

## Conclusions

The lime and lemon horticultural group is genetically highly polymorph with diploid, triploid and tetraploid varieties, maternal phylogeny involving four CT over the six encountered in the *Citrus* genus and nuclear contribution of the four *Citrus* basic taxa (*C. medica*, *C. maxima*, *C. micrantha* and *C. reticulata*). All lime and lemon varieties analysed were highly heterozygous and displayed interspecific admixture involving two, three but also the four ancestral taxa genomes for a few accessions and particularly the widely cultivated 'Tahiti' triploid limes (*C. latifolia* types). *C. medica* appeared as a common genomic component of all limes and lemons but never acted as direct female parent in relation with the cleistogamy of citron varieties. *C. medica* was very probably the direct male parent for the main lime and lemon sub-groups in combination with *C. micrantha* (*C. aurata*, *C. exelsa*, *C. webberii*, *C. macrophylla* and *C. aurantifolia* varieties of Tanaka taxa), *C. reticulata* (the *C. limonia*, *C. karna* and most of the *C. jambhiri* varieties of Tanaka taxa), *C. aurantium* (*C. limetta* and most of the *C. limon* varieties of Tanaka taxa), or *C. maxima* x *C. reticulata* hybrid (*C. limettioides*, *C. meyeri*). Other combinations involving *C. medica* hybrids were also identified. Two origins were identified for the triploid limes. *C. latifolia* varieties ('Tahiti', 'Bears', 'Persian', 'El Kseur', 'IAC 5') should result from the fertilization of an haploid gamete of *C. limon* (yellow lemon type) by a diploid gamete of *C. aurantifolia* ('Mexican' lime type; 2n gamete of a diploid parent or diploid gamete produced by a doubled diploid) while the *C. aurantifolia* triploid varieties ('Tanepao', 'Coppenhad', 'Madagascar', 'Ambilobe', 'Mothasseb') should results from an interspecific backcross (a diploid ovule of *C. aurantifolia* –'Mexican' lime type- pollinated by *C. medica*). The lime and lemon horticultural group results therefore of many independent reticulation events (36 identified in this work) that should explain that neither the botanical classifications (Swingle and Reece (1967) with two species and Tanaka (1954) with 37 taxa nor the usual denomination (limes and lemons) correctly reflects the genetic organization. Considering the very high interspecific heterozygosity of all analyzed varieties, the low intra-subgroup polymorphisms was attributed to punctual mutations, epigenetic variations or genomic fragment deletion as clearly revealed at the end of the chromosome 9 by our results for the 'Lumiciiana' lemon. The new insight on the global interspecific admixture structures and the inferred phylogenetic origins for the main sub-groups of limes and lemons will be essential for a better utilization of the citrus biodiversity to create new rootstock and acid citrus varieties with improved adaptation and resistance to biotic and abiotic constraints.

---

## Chapitre IV

---

# **Deciphering the interspecific genome structure of limes and lemons through whole genome SNP analyses**

## Titre et Résumé en français

### Déchiffrement de la structure interspécifique du génome des limettiers et des citronniers grâce à l'étude pangénomique des SNP

L'évolution réticulée des agrumes a produit des structures génomiques particulièrement complexes avec de fréquentes zones en hétérozygotie interspécifiques. Il est généralement admis que quatre taxons ancestraux (*C. medica* L. cédratiers, *C. reticulata* Blanco mandariniers, *C. maxima* (Burm.) Merr. pamplemoussiers, et *C. micrantha* Wester, papedas) sont à l'origine des toutes les espèces cultivées du genre *Citrus*. Les limettiers et les citronniers sont considérés comme le groupe variétal le plus complexe impliquant les quatre taxons de base et éventuellement un autre genre botanique, les *Microcitrus*. Le récent développement des NGS ont profondément modifiées les techniques d'études de l'évolution des plantes passant de la phylogénétique à la phylogénomique. Ainsi les structures phylogénomiques de l'oranger, du bigaradier et du clémentinier, résultant d'une combinaison de pools géniques *C. reticulata* et *C. maxima* ont été récemment décrites. L'objectif de cette étude a été (i) de développer un processus d'analyses permettant une inférence phylogénomique de fragments d'ADN, à partir de données SNP provenant de données de re-séquençage complet, pouvant prendre en compte un nombre illimité d'ancêtres potentiels, et (ii) de l'appliquer pour l'étude phylogénomique des limettiers et des citronniers. Les polymorphismes SNPs tout au long du génome sont issus de données de re-séquençage complet Illumina, ancrées sur la séquence de référence de l'haploïde de clémentinier. Ils nous ont été fournis par le consortium Citrusseq. Plus d'un million de SNP diagnostiques des 4 taxons de base et de *Microcitrus australasica* ont été sélectionnés dans les séquences géniques. Les structures phylogénomiques le long des 9 chromosomes ont été déduites à partir des fréquences des allèles diagnostiques des 5 ancêtres potentiels en homozygotie et hétérozygotie. Notre processus d'analyse a été validé par la cohérence de nos résultats pour l'oranger, le bigaradier et la clémentine avec les études précédentes issues d'analyse complète des génomes. Les structures phylogénomiques du Volkameriana, du Rough lemon et de la lime Rangpur ainsi que leur phylogénie maternelle démontrent que ces variétés sont des hybrides directs entre *C. reticulata* et *C. medica*, alors que la lime 'Mexicaine' et Alemow résultent d'une hybridation directe entre papeda (*C. micrantha* ou espèce apparentée) et *C. medica*. Nous avons confirmé l'hypothèse selon laquelle les citronniers de type 'Eureka' dériveraient d'un hybride direct entre bigaradier (*C. aurantium*) et *C. medica*. Nos résultats montrent que la lime Bears résulte de la pollinisation d'un gamète haploïde de citronnier type 'Eureka' par un pollen diploïde de *C. aurantifolia* (limettier de type Mexicain). Les caryotypes phylogénomiques de ces limettiers et citronniers et en particulier ceux issus de trois taxons de base (citronnier Eureka = (*C. maxima* × *C. reticulata*) × *C. medica*) et quatre taxons de base (limettier Bears = (*C. maxima* × *C. reticulata*) × (*C. micrantha* × *C. medica*)) ont été schématisés pour la première fois. Ce nouvel éclairage sur les origines et la structure phylogénomique des limettiers et des citronniers est important pour le développement de nouveaux programmes de création variétale au sein de ce groupe.



## Abstract

Reticulate evolution of citrus species led to complex genomic structures in frequent interspecific heterozygosity. Indeed, it is generally agreed that four ancestral taxa (*C. medica* L. (citron), *C. reticulata* Blanco (mandarin), *C. maxima* (Burm.) Merr. (pummelo), and *C. micrantha* Wester (papada)] are the ancestors of all cultivated *Citrus* species. Limes and lemons are considered as the more complex horticultural group involving the four ancestral taxa and eventually the related genera *Microcitrus*. Recently NGS technologies have deeply modified the way to analyze plant evolution, moving from phylogenetics to phylogenomics and the phylogenomic structure of sweet orange, sour orange and clementine, resulting from a combination of *C. reticulata* and *C. maxima* gene pools, were described. The objective of this work was to develop a workflow for phylogenomic inference from SNP derived from WG resequencing data, allowing taking into account an unlimited number of potential ancestors and to apply it for the phylogenomic study of limes and lemons. Re-sequencing illumina data mapped in the haploid clementine reference sequence and provided by the Citrusseq consortium were filtered for gene sequence and more than one million diagnostic SNPs for the 4 *Citrus* basic taxa and *M. australasica* were selected after the identification of introgressed genomic fragments in accessions representative of *C. reticulata* and *C. maxima*. Then, phylogenomic structures along the nine chromosomes were inferred from the patterns of the homozygosity and heterozygosity frequencies of the five ancestors diagnostic alleles. Our workflow was validated by the identity of our results for sweet orange, sour orange and clementine with previous structures inferred from WG data. 'Volkamer' lemon, 'Rough' lemon and 'Rangpur' lime phylogenomic nuclear structure and maternal phylogeny testified for an origin by *C. reticulata* × *C. medica* direct hybridization while 'Mexican' lime and 'Alemow' appeared to result from Papada (*C. micrantha* or closely related) × *C. medica* direct hybridization. The *C. aurantium* × *C. medica* hypothesis for 'Eureka' lemon was confirmed and we concluded that 'Bears' lime resulted from the pollination of *C. limon* ('Eureka' type) haploid gamete by a diploid gamete of *C. aurantifolia* ('Mexican' lime type). The 'phylogenomic karyotype' of these limes and lemons varieties and particularly the one issued from three ('Eureka' lemon= (*C. maxima* × *C. reticulata*) × *C. medica*) and four ('Bears' lime= (*C. maxima* × *C. reticulata*) × (*C. micrantha* × *C. medica*)) ancestral species are provided for the first time. These new insight in limes and lemon origin and phylogenomic structures are decisive to develop innovative breeding schemes for these horticultural groups.

## Introduction

Reticulate evolution poses serious challenges in phylogenetic analyses. Indeed, hybrid speciation, introgression and lateral gene transfer, lead to evolutionary histories that cannot be adequately represented in the form of phylogenetic trees (Stebbins, 1950; Grant, 1981; Arnold, 1997; Doolittle, 1999; Otto and Whitton, 2000). In many species, it is rather a network, resulting in phylogenetic incongruences between cytoplasmic and nuclear genomes but also between nuclear genome areas (Pamilo and Nei, 1988; Rieseberg and Soltis, 1991; Linder and Rieseberg, 2004; Beiko and Hamilton, 2006). When vegetative propagation, such as apomixis, takes place immediately or a few generations after a reticulation event, large parts of the genome often remain in interspecific heterozygosity. Genome-wide molecular analyses are therefore needed to decipher the complex interspecific mosaic genomes resulting from such evolution.

The genus *Citrus* is a good example of such a gene pool with reticulate evolution, where apomixis and vegetative propagation have fixed ancient reticulation events and limited further interspecific recombinations. Indeed, molecular marker analyses have brought decisive information for the comprehension of domestication and the relations between the various cultivated species of *Citrus* (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Li *et al.*, 2010; Garcia-Lor *et al.*, 2012; Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b; Ramadugu *et al.*, 2013). Four ancestral taxa [*C. medica* L. (citron), *C. reticulata* Blanco (mandarin), *C. maxima* (Burm.) Merr. (pummelo), and *C. micrantha* Wester (papeda)] have been identified as the ancestors of all cultivated *Citrus* (Nicolosi *et al.*, 2000; Garcia-Lor *et al.*, 2012; Ramadugu *et al.*, 2013). All *Citrus* taxa are still fully sexually compatible, but previous studies testified that the four basic taxa have reached an advanced stage toward speciation. Indeed, cytogenetic studies (Nair and Randhawa, 1969; Raghuvanshi, 1969; Guerra, 1993) evidenced a high karyotypic diversity among *Citrus* species and flow cytometry analyses of nuclear genome size of *Citrus* species displayed a differentiation reaching 10% between *C. reticulata* (0.74 pg/2C) and *C. medica* (0.81 pg/2C) (Ollitrault *et al.* 2003). So called secondary species [*C. sinensis* (L.) Osb. (sweet orange), *C. aurantium* L. (sour orange), *C. paradisi* Macf. (grapefruit), *C. limon* (L.) Burm. (lemon), and *C. aurantifolia* (Christm.) Swing. (lime)] arose from hybridizations between the four basic taxa (Nicolosi *et al.*, 2000; Garcia-Lor *et al.*, 2013b). A major part of the phenotypic diversity of edible citrus results from the initial differentiation between the basic taxa and the interspecific mosaic structure is the main component driving the ideotype of the secondary species. Deciphering the interspecific admixture structure of citrus germplasm is therefore essential for an efficient mobilization of citrus biodiversity in innovative breeding schemes.

NGS have deeply modified the way to analyze plant evolution, moving from phylgenetics to phylogenomics based on whole genome variability analysis and new insight have been provided into the domestication history of several fruit crops (Myles *et al.*, 2011; Cornille *et al.*, 2012) and cereals (Huang *et al.*, 2012; Hufford *et al.*, 2012). The release of the first high quality citrus reference genome by the international citrus genome consortium (ICGC, 2014; Wu *et al.*, 2014) implemented from a haploid clementine was a fundamental step to develop phylogenomic in citrus. Recently resequencing WGS data revealed unexpected *C. maxima* introgressions in traditional mandarins genomes (Wu *et al.*, 2014) and revealed the origin of sour orange, sweet orange and clementine (Xu *et al.*, 2013; Wu *et al.*, 2014). Moreover the mosaic interspecificity between *C. maxima* and *C. reticulata* of the concerned varieties was deciphered all along the citrus genome by Wu *et al.* (2014). Even if introgressions in mandarin cultivars made the analysis more complex, the previous study

was limited to relatively simple situation with only two supposed ancestral taxa. In the present work, we propose an approach allowing taking into account an unlimited number of potential ancestors. It is focused on the “acid citrus” limes and lemons (*C. aurantifolia* and *C. limon*) at diploid and triploid level. Indeed, previous molecular studies suggested that the four ancestral taxa of edible citrus contributed to the genesis of limes and lemons (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Bayer *et al.*, 2009; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b; Curk *et al.*, accepted; Curk *et al.*, submitteda; Curk *et al.*, submittedb). However, if most authors agree on the role of *C. medica* in the origin of all “acid citrus”, the phylogeny of most limes and lemons remains controversial. In addition to *Citrus* species, the related genera *Microcitrus* was proposed as potential parent of ‘Mexican’ lime (Barrett and Rhodes, 1976).

In this work, we take advantage of some publicly available WG resequencing data (Xu *et al.*, 2013) and of WG resequencing data of the Spanish Citrusseq consortium kindly made available for this study to identify species-diagnostic markers of the four *Citrus* basic taxa and *Microcitrus* to analyse the phylogenomic structure of two diploid and one triploid *C. aurantifolia* accessions and four diploid of *C. limon* accessions. Three accessions of mandarins, pummelos and citrons and one of *C. micrantha* were used to identify diagnostic SNP of the four *Citrus* ancestral taxa. One accession of *M. australisica* was used for *Microcitrus* diagnostic points. Sour orange supposed to be the female parent of “yellow lemons” (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b) was added to the study. We also analysed sweet orange and clementine already described by Wu *et al.* (2014) to validate our phylogenomic inference approach.

Researchers of the Citrusseq consortium mapped the re-sequencing data (Terol *et al.*, 2012) in the nine main nuclear scaffolds (ie: 9 chromosomes of the haploid citrus set) of the reference clementine sequence provided by the International Citrus Genomic Consortium (Wu *et al.*, 2014); [http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org\\_Cclementina](http://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Cclementina) and in the reference chloroplast sequence of sweet orange (Bausher *et al.*, 2006), performed a rigorous variant calling and generated the SNP matrix (Carbonell-Caballero *et al.*, submitted; Terol *et al.*, submitted) that allowed us to perform the phylogenomics analyses.

Therefore both maternal phylogeny and nuclear phylogenomic were analysed. Considering the karyotypic diversity of citrus species and the potential important variability in repeated sequence areas we decided to develop the phylogenomic study taking into account only the gene sequences as annotated along the nuclear genome in phytozome. For the first time, phylogenomic karyotypes are proposed for seven limes and lemons varieties.

## Material and Method

### SNP matrix

The SNP matrix used in this work was provided by the Citrusseq consortium/Genomic Center of IVIA. The matrix was derived from Illumina data after rigorous SNP calling following standard procedures of the Citrusseq consortium (Carbonell-Caballero *et al.*, submitted; Terol *et al.*, submitted). The matrix included 20 accessions belonging to the *Citrus* genus and one to *Microcitrus australisica* (Table IV.1). We adopted the Swingle and Reece (1967) botanical classification for scientific names. Ten accessions of the four horticultural groups highly related with the four ancestral taxa (or so called basic taxa) were used (three mandarins, three pummelos, three citrons and one *C. micrantha*). Representatives of limes and lemons included two diploid and one triploid *C. aurantiifolia* accessions and four *C. limon* accessions. Sour orange (*C. aurantium*) was included as putative parent of yellow lemons (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b; Curk *et al.*, submitted). We also add one sweet orange (*C. sinensis*) and one clementine (*C. clementina*) already described by Wu *et al.* (2014) to validate our method of basic species-diagnostic SNP identification and secondary species phylogenomic deciphering.

Table IV.1. Genome of citrus accessions from which the SNP Matrix was derived.

Common name	Swingle and Reece classification
'Cleopatra' mandarin	<i>Citrus reticulata</i> Blanco
'Willowleaf' mandarin	<i>Citrus reticulata</i> Blanco
'Huanglingmiao' mandarin	<i>Citrus reticulata</i> Blanco
'Chandler' pummelo	<i>Citrus maxima</i> (Burm.) Merr
'Guanxi' pummelo	<i>Citrus maxima</i> (Burm.) Merr
'Shatian' pummelo	<i>Citrus maxima</i> (Burm.) Merr
'Corsican' citron	<i>Citrus medica</i> L.
'Buddha's hand' citron	<i>Citrus medica</i> L.
'Humpang' citron	<i>Citrus medica</i> L.
Citrus micrantha	Citrus micrantha Wester
'Clemenules' clementine	<i>Citrus reticulata</i> Blanco
'Washington' navel sweet orange	<i>Citrus sinensis</i> (L.) Osbeck
'Sevillan' sour orange	<i>Citrus aurantium</i> L.
'Mexican' lime	<i>Citrus aurantiifolia</i> (Christm.) Swingle
'Alemow'	<i>Citrus aurantiifolia</i> (Christm.) Swingle
'Rangpur' lime	<i>Citrus limon</i> (L.) Burm. F.
'Bears' lime	<i>Citrus aurantiifolia</i> (Christm.) Swingle
'Eureka' lemon	<i>Citrus limon</i> (L.) Burm. F.
'Volkamer' lemon	<i>Citrus limon</i> (L.) Burm. F.
'Rough' lemon	<i>Citrus limon</i> (L.) Burm. F.
'Australian finger' lime	<i>Microcitrus australasica</i> (F.Muell.) Swingle

### Maternal phylogeny study

Diallelic SNPs over the whole chloroplast genome were selected and suspicious SNP positions with heterozygotic patterns eliminated. Neighbor-joining analysis (Saitou and Nei, 1987) was computed using DARwin software version 5.0 (Perrier and Jacquemoud-Collet, 2006). Genetic dissimilarities were calculated using the simple matching dissimilarity index ( $d_{i-j}$ ) between pairs of accessions:

$$d_{i-j} = 1 - 1/L \sum_{l=1}^L m_l / 2$$

with  $d_{i-j}$ , the dissimilarity between units  $i$  and  $j$ ;  $L$ , the number of loci;  $m_l$ , the number of matching alleles for locus  $l$ . Weighted neighbor-joining trees were computed from the dissimilarity matrix with 1000 bootstraps to test branch robustness.

### Genetic parameters

The search of diagnostic SNPs of each taxon was based on  $G_{ST}$  parameter estimations (Nei, 1973) for the concerned taxon considering two subpopulations: (1) the concerned taxa ( $T_i$ ) and (2) a subpopulation joining the 4 others ancestral taxa ( $T-i$ ). It was performed from the estimated allele frequency of each taxa considering same population size for each taxa to estimate the frequency of the two subpopulations ( $T_i$  and  $T-i$ ) and the whole population ( $Tot$ ) frequency.

$$G_{ST\ Taxai} = (He_{Tot} - (He_{T_i} + He_{T-i})/2) / He_{Tot}$$

where  $He$  is the expected proportion of heterozygous loci per individual ( $He = 1 - \sum p_i^2$ , where  $p_i$  is the frequency of a given allele in the considered population or subpopulation).

Values of  $G_{ST}$  range from zero to one, with low values indicating that little variation is proportioned among subpopulations while, high values, denote that a large amount of variation is found among subpopulations. In our study  $G_{ST\ Taxai}=1$  means that the taxa  $i$  is totally differentiated from the four other ancestral taxa and probably fixed for a mutant allele that probably occurs in the taxa  $i$  after the separation of the considered taxon from the others.

Observed heterozygosity ( $H_o$ ), expected heterozygosity ( $He$ ) and  $G_{ST}$  estimations were computed in excel.

Genetic relationships were studied by Neighbour-Joining analysis (NJA), based in the SNP data with DARwin software (Perrier and Jacquemoud-Collet, 2006), using the simple matching dissimilarity index ( $d_{i-j}$ ) described above. This index was also used to infer the intra- and inter taxa average differentiation and for Factorial Analysis using DARwin software. The similarity index used to identify introgressions in accessions representative of basic taxa is calculated as  $S_{i-j}=1-d_{i-j}$ .

## Results and discussion

### Maternal phylogeny

Phylogenetic inferences from chloroplastic data are currently established by the Citruseq consortium. Therefore the objective in the present work was not to infer the evolutive sequence leading to the divergence of basic taxa, nor the times of divergence between taxa but to identify the different chlorotypes and to study the relations of the chlorotype of the ancestral taxa with the ones of the analysed lemons and limes. Diallelic SNPs over the whole chloroplast genome were selected and suspicious SNP positions with heterozygotic patterns eliminated. 718 diallelic SNP positions, without suspicious heterozygous genotypes, were found over the whole chloroplast genome. Five main phylogenetic clusters were observed on the NJ analysis (Figure IV.1). The *C. maxima* cluster includes three sub-clusters. One joining the three pummelo accessions, the second one is represented by sweet orange and the third one associate sour orange, 'Eureka' lemon and 'Bears' lime. The differentiation between *C. aurantium* and pummelo accession cytoplasm type (CT) was already detected by cpSSR markers as well as the identity of 'Eureka' lemon and 'Bears' lime CT with the *C. aurantium* one (Curk *et al.*, submittedb). Earlier molecular marker studies also concluded for the CT identity between *C. aurantium* and "yellow lemons" (Nicolosi *et al.*, 2000; Bayer *et al.*, 2009; Froelicher *et al.*, 2011; Curk *et al.*, in press; Curk *et al.*, submittedb). Froelicher *et al.* (2011), also shown that *C. aurantium* "yellow lemons" and triploid limes (Bears, Tahiti) shared the same mitotype. The second main cluster includes *C. micrantha*, 'Mexican' lime (very close to *C. micrantha*) and 'Alemow', a little more differentiated. The high similarity of 'Mexican' lime and 'Alemow' CT with the *C. micrantha* one were already proposed from chloroplastic (Nicolosi *et al.*, 2000; Bayer *et al.*, 2009; Penjor *et al.*, 2013; Curk *et al.*, in press; Curk *et al.*, submittedb) and mitochondrial (Froelicher *et al.*, 2011) marker studies. The differentiation observed from the WGS data indicates that the maternal parent of 'Alemow' was different than the one of 'Mexican' lime. Without data on the *C. micrantha* intraspecific diversity, we cannot conclude if the maternal parents of 'Mexican' lime and 'Alemow' were *C. micrantha* genotypes different of our analysed accession or another papeda species closely related to *C. micrantha*. The *C. reticulata* cluster includes two main sub-clusters. The first one associates 'Cleopatra' mandarin and 'Rangpur' lime in a first branching and then, a little more differentiated, the 'Rough' and 'Volkamer' lemons (identical CT). The second main subcluster is formed by two edible mandarins ('Willowleaf' and 'Huanglingmiao') and 'Clemenules' clementine sharing totally identical CT. The differentiation between two main mandarin CT as well as the association of 'Rangpur' lime, 'Rough' and 'Volkamer' lemons with the 'Cleopatra' mandarin one was previously described at mitochondrial (Froelicher *et al.*, 2011) and chloroplastic levels (Curk *et al.*, submittedb). However in these previous study the three lime and lemon CTs were found totally identical with 'Cleopatra' mandarin. The fourth and fifth clusters are constituted respectively by the three citron varieties (*C. medica*) and 'Australian finger' lime (*M. australisica*) and do not include any of the *C. aurantifolia* or *C. limon* accessions. Therefore regarding lime and lemon maternal phylogeny these results from WGS data definitively confirm previous main conclusions obtained with conventional molecular markers. They show additionally that (i) 'Volkamer' and 'Rough' lemons should have an identical maternal parent but different from the 'Rangpur' lime one and (ii) that 'Alemow', 'Mexican' and *C. micrantha* have closely related but different CTs.

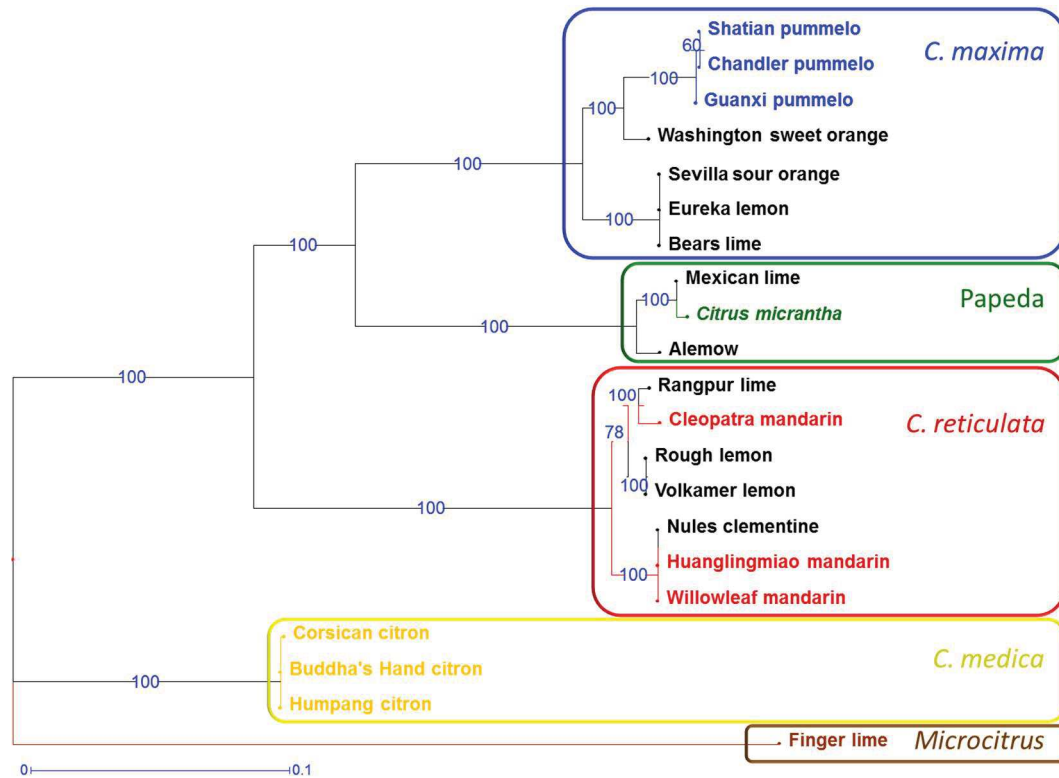


Figure IV.1. Maternal phylogeny of 7 limes and lemons and potential ancestor representatives; NJ tree based on simple matching dissimilarity index from 718 Chloroplastic SNPs.

### Search for ancestral taxa diagnostic markers in gene sequences

The workflow to identify diagnostic markers from the WGS SNP matrix is schematized in Figure IV.2. The first step is to filter the WGS data on gene sequence. The second one is to identify introgressed areas in the genome of the representative of the potential ancestors working only with the SNP positions displaying interspecific structuration (at list a basic horticultural group (BHG) with a  $Gst_{HG}$  value -differentiation between the considered basic horticultural group and a theoretical population of all others BHGs- over 0.5). Indeed we are not working with real ancestors but actual varieties resulting from the domestication process and recent studies have evidenced interspecific introgressions in varieties previously considered as pure *C. reticulata* or pure *C. maxima* (Wu *et al.*, 2014; Curk *et al.*, submitteda). Therefore the identification of ancestral taxa diagnostic SNPs requires to remove these introgressed areas for the concerned varieties before to estimate the allelic frequencies in the ancestral taxa and the interspecific differentiation parameter  $Gst_{taxa}$ . In the last step we have filtered the SNP positions by the value of the estimated  $Gst_{taxa}=1$  for one of the five potential ancestral taxa.

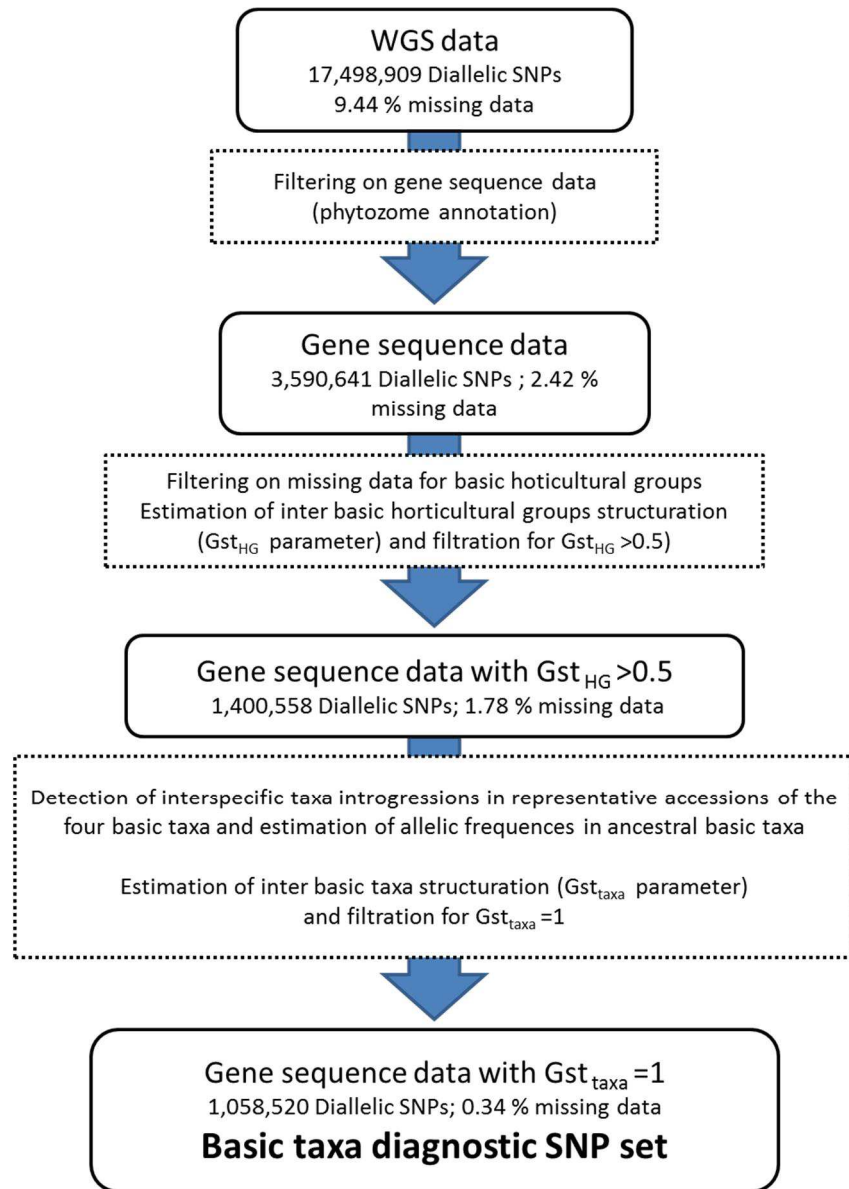


Figure IV.2. Workflow for the identification of ancestral taxa diagnostic points

### ***SNP polymorphism of nuclear gene sequences***

Over the 288,587,547 bases of the 9 mains scaffolds (= 9 chromosomes of the haploid set) of the haploid clementine reference genome sequence (phytozome), 17,498,909 diallelic positions displayed an alternative allele of the reference one, in at least one of the 21 citrus varieties analyzed. For these positions the rate of missing data over the 21 accessions was 9.44 %. The SNP positions of the WG matrix were then filtered over the 24357 genes annotated in the 9 mains scaffolds of the reference genome sequence that cover 74.447,926 kb. A total of 3,590,641 SNP positions were retained (Table IV.2) with 2.42 % of missing data.



Table IV.2. Single nucleotide polymorphisms distribution (gene sequence data)

Scaffold	Gene sequence size	Gene sequence SNPs	SNPs/kb	homozygot SNPs <sup>1</sup>	Non taxa SNPs <sup>2</sup>	Hetero Non Taxa SNPs	Homo Non Taxa SNP
1	7,823,994	366,931	46.90	36,037	23,452	23,359	93
2	9,692,704	470,944	48.59	46,976	37,235	37,102	133
3	14,134,733	677,669	47.94	63,114	49,330	48,923	407
4	7,761,013	377,018	48.58	35,769	30,824	30,696	128
5	9,008,204	422,908	46.95	44,735	31,067	30,658	409
6	6,402,475	320,599	50.07	33,883	24,481	24,092	389
7	6,637,531	321,472	48.43	32,929	26,530	26,138	392
8	5,957,414	289,994	48.68	28,473	21,066	20,947	119
9	7,029,858	343,106	48.81	33,899	27,265	27,164	101
Total	74,447,926	3,590,641	48.23	355,815	271,250	269,079	2,171

<sup>1</sup>: all genotypes are homozygote; <sup>2</sup>: SNP positions where all ancestral taxa are homozygote and identical;

The SNP rate by kb in gene sequence (48.23) is in the range of the one observed in previous citrus nuclear gene studies from Sanger sequencing (52.9 SNP/kb, Garcia-Lor *et al.*, 2013b; Ramadugu *et al.*, 2013), and 454 sequencing (36.7 SNP/kb; Curk *et al.*, submitteda). 355,815 (10 %) SNPs display the alternative allele only in homozygosity and correspond mainly (238,032, 67%) to specific polymorphism of *Microcitrus*. 271,250 SNP polymorphic positions (7.5% of total polymorphic ones) displayed alternative alleles only in secondary citrus species and just 2,171 of them were found only in homozygosity. The average frequency of these alleles not present in the basic horticultural groups over the secondary species accession was 1.3%. Higher rates were found in 'Alemow' (2.6%), and sour orange (2.5%) and 'These low rates of additional polymorphisms in secondary species, despite the very limited size of the analysed basic taxa populations, validate the hypothesis that the ancestors of the analysed secondary species are part of the considered basic taxa. The rate of heterozygosity per kb is higher in all limes and lemons accessions (11.5 to 15.3 SNP/kb) compared to the BHG accessions (Table IV.1) in agreement with their supposed interspecific origins.

To analyze the diversity within and between the BHGs, positions with missing data for *C. micrantha* and *M. australisica* were removed as well as the ones with more than one missing data for the three other BHGs. Intrataxa diversity parameters (Table IV.3) were estimated considering only the SNP positions displaying polymorphisms among the set of 11 accessions of the BHGs. The observed heterozygosity ( $H_o$ ) is very similar in mandarins, pummelos and *C. micrantha* (close to 0.07) and three time higher than in citrons (0.02). *Microcitrus* has the highest observed heterozygosity of the BHGs (0.17). In the same way the intraspecific diversity ( $H_e$  index, Table IV.4) is significantly lower (close than half) in citron than pummelos and mandarins. Previous molecular studies also pointed out the low heterozygosity of *C. medica* accessions (Gulsen and Roose, 2001; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2013b; Curk *et al.*, submitteda). Self-pollination is favored by the flower structure of citron and produce vigorous selfed seedlings (Barrett and Rhodes, 1976) while self-incompatibility is prevalent in pummelos and partial apomixis (nucellar polyembryony)

can fix complex heterozygous genomic structure in mandarins. Reproductive biology is therefore a key element of the higher homozygosity in *C. medica*. Blondel (1978) suggested that the cultivar 'Buddha's Hand' may have arisen by hybridization of *C. medica* with some other *Citrus* species. The very low heterozygosity of 'Buddha's Hand' and its very close relationship with the two other citrons definitively rule out this hypothesis. The very special fruit morphology of 'Buddha's Hand' results clearly from punctual genome modification or epigenetic changes.

Table IV.3. Basic horticultural group diversity (Ho, gene sequence data SNPs)

Scaf.	Ho mandarins (n=3)	Ho pummelos (n=3)	Ho citrons (n=3)	Ho <i>C. micrantha</i> (n=1)	Ho <i>Microcitrus</i> (n=1)
1	0.0568±0.0006	0.0733±0.0006	0.0226±0.0004	0.0647±0.0008	0.1629±0.0012
2	0.0579±0.0005	0.0701±0.0006	0.0191±0.0003	0.0655±0.0007	0.1557±0.0011
3	0.0723±0.0005	0.0669±0.0005	0.0225±0.0003	0.0687±0.0006	0.1876±0.0010
4	0.0779±0.0006	0.0568±0.0006	0.0192±0.0004	0.0641±0.0008	0.1710±0.0013
5	0.0547±0.0005	0.0594±0.0006	0.0245±0.0004	0.0772±0.0008	0.1747±0.0012
6	0.0862±0.0007	0.0727±0.0007	0.0241±0.0005	0.0698±0.0009	0.1374±0.0013
7	0.0625±0.0007	0.0800±0.0008	0.0301±0.0005	0.0806±0.0010	0.1461±0.0014
8	0.0843±0.0007	0.0776±0.0008	0.0315±0.0005	0.0665±0.0009	0.1588±0.0041
9	0.0613±0.0006	0.0733±0.0007	0.0279±0.0005	0.0681±0.0009	0.1629±0.0014
Total	0.0676±0.0006	0.0691±0.0006	0.0240±0.0004	0.0692±0.0008	0.1666±0.0012

Ho: Observed heterozygosity

Table IV.4. Basic horticultural group diversity (He gene sequence data SNPs)

Scaf.	He mandarins (n=3)	He pummelos (n=3)	He citrons (n=3)	He <i>C. micrantha</i> (n=1)	He <i>Microcitrus</i> (n=1)
1	0.0465±0.0004	0.0622±0.0005	0.0283±0.0004	0.0323±0.0004	0.0814±0.0006
2	0.0463±0.0004	0.0614±0.0004	0.0283±0.0003	0.0327±0.0004	0.0769±0.0005
3	0.0561±0.0003	0.0558±0.0003	0.0278±0.0003	0.0344±0.0003	0.0938±0.0005
4	0.0593±0.0005	0.0560±0.0005	0.0268±0.0003	0.0321±0.0004	0.0855±0.0006
5	0.0450±0.0004	0.0519±0.0004	0.0302±0.0003	0.0386±0.0004	0.0861±0.0006
6	0.0700±0.0005	0.0553±0.0005	0.0268±0.0004	0.0349±0.0005	0.0687±0.0006
7	0.0509±0.0005	0.0640±0.0005	0.0289±0.0004	0.0403±0.0005	0.0730±0.0007
8	0.0757±0.0006	0.0589±0.0005	0.0321±0.0004	0.0333±0.0005	0.0794±0.0007
9	0.0486±0.0004	0.0613±0.0005	0.0244±0.0003	0.0340±0.0004	0.0814±0.0007
Total	0.0545±0.0004	0.0582±0.0005	0.0281±0.0004	0.0346±0.0004	0.0833±0.0006

He: expected heterozygosity

### **Identification of introgressed genome area in representative of basic taxa and estimation of basic taxa allelic frequency and differentiation**

The search of introgressed areas in the representative accessions of the ancestral taxa was made by analysing the patterns of three parameters along each chromosome:

- Observed heterozygosity estimated from the matrix of genic SNP positions.
- Similarity of the considered accessions with the centroid of a theoretical population of hybrids between each pair of BHGs defined by the average allelic frequency of the two considered BHGs.
- Similarity between each pair of BHG accessions. This similarity was estimated considering only the polymorphic positions of the two BHGs concerned by each comparison (e.g. polymorphic positions in the 6 accessions of pummelo and mandarin for the estimation of similarity between a pummelo and a mandarin).

For the analysis of the patterns of the two last parameters, in order to enhance the resolution of interspecific differentiation, the analysis was made after filtering the gene matrices for SNP positions with a  $Gst_{HG} > 0.5$  for one of the BHGs. Each dot of the three curves is the average of 500 SNPs positions (249 before and 250 after). These average values were evaluated every 250 SNP positions. The Figure IV.3 illustrates the case of the chromosome 6 for 'Willow leaf' mandarins.

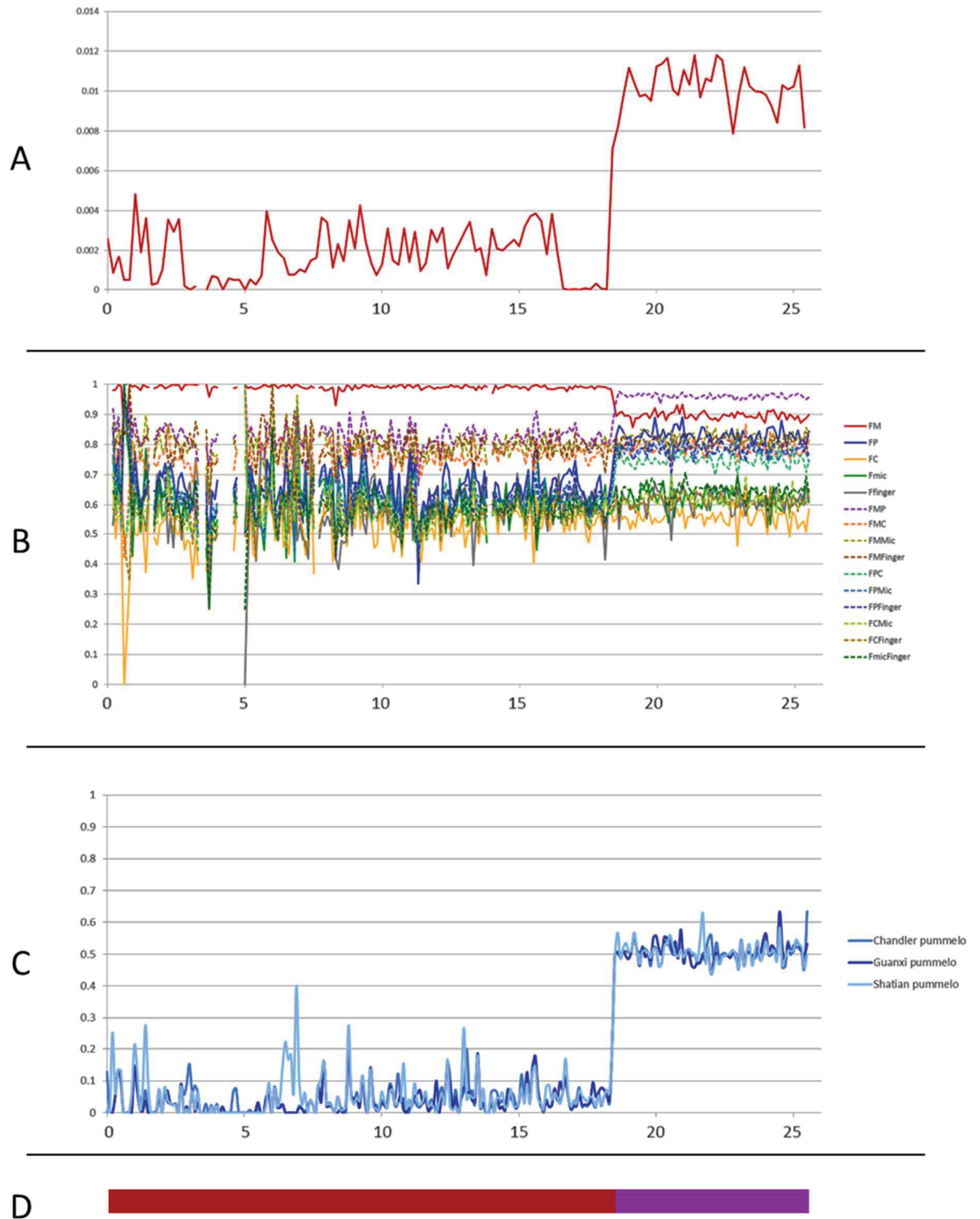


Figure IV.3. identification of interspecific introgressions in accessions representative of the basic horticultural groups: example of chromosome 6 of 'Willowleaf' mandarin (WLM). The x axis is the position in the scaffold 6 in Mb (phytozome). A: observed heterozygosity of WLM in genic area; each position is an estimation within a window of 200 kb; B: similarities of WLM with centroids of basic horticultural groups and theoretical inter horticultural groups hybrids within a window of 100 kb; C: similarity of WLM with the three accessions of pummelos within a window of 100 kb; D: deduced phylogenomic structure of 'Willowleaf' mandarin chromosome 6 (red: *C. reticulata* homozygosity; purple: *C. maxima/C. reticulata* heterozygosity).

Observed heterozygosity (Figure IV.3A) display a strong change of pattern around a position of 18.4 Mb moving from an average heterozygosity oscillating around 0.002 to 5 time higher one (variations around 0.010). This pattern suggests an interspecific heterozygosity at the end of the chromosome 6 of ‘Willowleaf’ mandarin. The analysis of similarity of ‘Willowleaf’ with BHG and theoretical inter-BHG hybrid centroids (Figure IV.3B) reveal a very high similarity with mandarin centroid until 18.4 Mb and then the mandarin × pummelo hybrid centroid appears as the closer theoretical structure. The comparison of ‘Willowleaf’ mandarin with the three pummelos (Figure IV.3C) confirm the *C. reticulata*/*C. maxima* heterozygosity after 18.4 Mb, with similarity values oscillating around 0.5 while they are close to 0 in the first part of the chromosome. From these three patterns we can conclude that for the chromosome 6, ‘Willowleaf’ mandarin is in *C. reticulata* phylogenomic homozygosity until 18.4 Mb and then in *C. reticulata*/*C. maxima* phylogenomic heterozygosity.

Over the 9 chromosomes, ‘Willowleaf’ mandarin was found introgressed in heterozygosity by *C. maxima* in chromosomes 2, 3, 4, 6 and 8 (Figure IV.4).

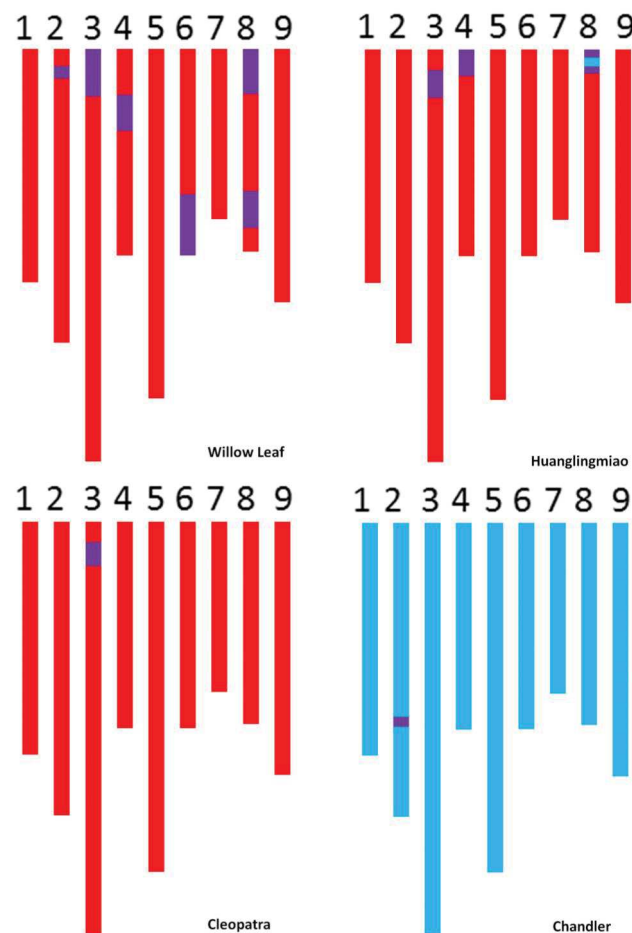


Figure IV.4. Interspecific introgressions in the nine chromosomes of four accessions of the basic horticultural taxa (red: *C. reticulata*; blue: *C. maxima*; purple: interspecific *C. reticulata*/*C. maxima* heterozygosity).

For ‘Huanglingmiao’ mandarin, *C. maxima* introgressions were found in heterozygosity in chromosome 3, 4, 8 and in homozygosity for a small segment (1.2 Mb) of chromosome 8 (Figure IV.4). Only a small *C. maxima* introgression in chromosome 3 was deduced for ‘Cleopatra’ mandarin. For the three pummelos only one small introgression (1.1 Mb) of *C. reticulata* in heterozygosity was observed for the chromosome 2 of Chandler (Figure IV.4). No interspecific introgressions were observed for the accessions of citron, *C. micrantha* and *M. australisica*. The introgression patterns deduced in our study for ‘Willowleaf’ mandarin, ‘Huanglingmiao’ mandarin and ‘Chandler’ pummelo are identical to the ones described in Wu *et al.* (2014). ‘Cleopatra’ mandarin appears to be the purest *C. reticulata* genome deduced from WGS data. That confirm previous inference from structure analysis based on SNP markers study (Curk *et al.*, submitteda) and by the pattern of haplotype phylogeny for 16 genes fragments of chromosome 2 (Curk *et al.*, accepted).

To estimate the average differentiation (SNP/kb) between two accessions at Intra-taxa and Inter taxa level we removed the SNP positions introgressed in at list one accession of the BHGs. Within taxa it varied between 1.51 for *C. medica* to 2.84 for *C. maxima* (Table IV.5). Between taxa, it ranged from 10.76 between *C. reticulata* and *C. maxima* to 15.06 between *C. micrantha* and *M. australisica*. The values obtained within the *Citrus* genus are very similar to the one obtained by Sanger sequencing of 16 nuclear genes (Garcia-Lor *et al.*, 2013b) and 454 sequencing of 57 nuclear gene fragments (Curk *et al.*, submitteda).

Table IV.5. Average differentiation (SNP/kb) between two accessions at Intra-taxa and Inter taxa level (Gene sequence without introgression areas)

	<i>C. reticulata</i>	<i>C. maxima</i>	<i>C. medica</i>	<i>C. micrantha</i>
<i>C. reticulata</i>	2.32			
<i>C. maxima</i>	10.76	2.84		
<i>C. medica</i>	12.74	11.99	1.51	
<i>C. micrantha</i>	12.18	11.39	13.24	
<i>M. australisica</i>	13.88	15.76	13.93	15.06

NJ analysis done with the same data confirmed the strong intertaxa structuration and did not reveal clear clustering structuration of the ancestral taxa (Figure IV.5).

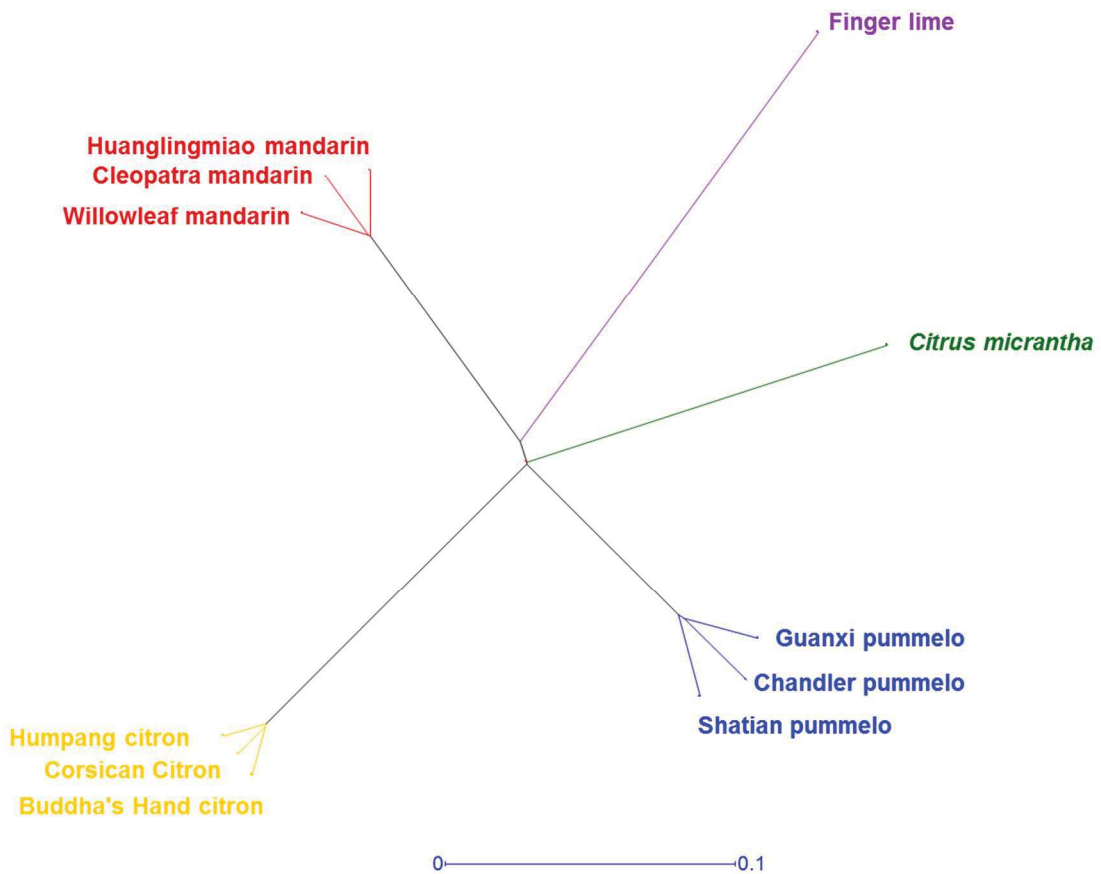


Figure IV.5. Nuclear gene sequence NJ tree between ancestral taxa without introgressed areas

The allelic frequencies of the basic taxa, and the  $Gst_{Tax}$  values for each SNP position were estimated after removing the introgressed areas for the concerned varieties. The important inter-taxa differentiation allowed to identify more than one million (1,058,520) SNP position that totally differentiate one of the ancestral taxa from the four other ones ( $Gst_{Tax}=1$ ; Table IV.6). For the three basic taxa represented by 3 cultivars it is probable that the lower intra specific diversity of *C. medica* contributed to identify higher number of diagnostic SNP in this taxa (298,977). At chromosome level, the lower number of diagnostic point was found in the chromosome 8 for *C. maxima* (8539). Regarding *C. micrantha* and *M. australasica* represented only by one accession, part of the diagnostic SNPs positions of these two taxa should correspond to allelic variations only found in the considered taxa but not fixed in the taxa. These five set of diagnostic SNPs of the five considered ancestral taxa have been used to decipher the phylogenomic structure of the 21 analysed varieties.

Additional re-sequencing data exploring a wider within diversity of the ancestral taxa will allow refining the identification of phylogenetic diagnostic points not only for the main branching between the 4 *Citrus* basic taxa but also for sub-clade within these taxa.

Table IV.6. Number of diagnostic point for each ancestral taxa and chromosome over all genome.

Scaffold Gene sequence	<i>C. reticulata</i>	<i>C. maxima</i>	<i>C. medica</i>	<i>C. micrantha</i>	<i>Microcitrus</i>	Total
1	18,333	12,595	32,127	25,870	25,190	114,115
2	20,384	14,241	38,499	33,532	32,098	138,754
3	33,674	20,524	55,924	45,681	41,190	196,993
4	16,428	11,436	31,761	25,735	23,649	109,009
5	20,331	13,313	34,822	28,496	28,582	125,544
6	15,179	10,063	28,110	21,794	24,040	99,186
7	13,423	8,689	25,896	20,184	23,252	91,444
8	11,073	8,539	23,601	20,070	20,599	83,882
9	14,617	9,827	28,237	23,999	22,913	99,593
Total gene sequence	163,442	109,227	298,977	245,361	241,513	1,058,520

### Admixture phylogenomic structure of secondary diploid species

The frequency of the ancestral taxa specific mutations were estimated for the 20 diploid varieties (Table IV.7). For the BHG, it was in total agreement with the previous analysis of interspecific introgression. Among the secondary species five displayed equilibrated frequency close to 0.50 for specific alleles of two taxa: Sour orange for *C. reticulata* and *C. maxima* alleles, 'Mexican' lime for *C. medica* and *C. micrantha*. 'Rangpur' lime, 'Volkamer' lemon and 'Rough' lemon for *C. reticulata* and *C. medica* alleles. The sweet orange and clementine displayed a higher frequency for specific mandarin alleles (respectively 0.553 and 0.791) than for *C. maxima* ones (0.399 and 0.155). 'Alemow' had a frequency 0.485 for *C. medica* alleles and a little lower for *C. micrantha* alleles. Lemon is the only diploid cultivar to display a very clear contribution of three ancestral taxa (*C. medica*, *C. reticulata* and *C. maxima* with respective specific allele frequencies of 0.480, 0.303 and 0.165). In addition to the mentioned contribution very low values (< 0.3%) are observed in the secondary species for the others basic taxa. The analysis of the distribution of specific mutations along the genome will allow concluding if it is representative of a small introgressed fragment (grouping of these specific alleles in a small genomic region) or rather artefactual data. This analysis along the genome will also reveal the areas in homozygosity or heterozygosity for the specific alleles of each ancestral taxa. The reconstruction of these "phylogenomic caryotype" based on the ancestral diagnostic SNPs identified in genic sequence is confronted to the patterns of the global heterozygosity estimated over the whole genome to check any incongruence.



Table IV.7. frequency of diagnostic alleles of the 5 ancestral taxa in the 20 diploid cultivars

	<i>C. reticulata</i>	<i>C. maxima</i>	<i>C. medica</i>	<i>C. micrantha</i>	<i>M. australasica</i>
	Specific Mutation	Specific Mutation	Specific Mutation	Specific Mutation	Specific Mutation
'Cleopatra' mandarin	0.991	0.004	0.000	0.000	0.000
'Willowleaf' mandarin	0.933	0.050	0.000	0.000	0.000
'Huanglingmiao' mandarin	0.978	0.022	0.000	0.000	0.000
'Chandler' pummelo	0.001	0.999	0.000	0.000	0.000
'Guanxi' pummelo	0.000	1.000	0.000	0.000	0.000
'Shatian' pummelo	0.000	1.000	0.000	0.000	0.000
'Corsican' citron	0.000	0.000	1.000	0.000	0.000
'Buddha's hand' citron	0.000	0.000	1.000	0.000	0.000
'Humpang' citron	0.000	0.000	1.000	0.000	0.000
<i>Citrus micrantha</i>	0.000	0.000	0.000	1.000	0.000
'Australian finger' lime	0.000	0.000	0.000	0.000	1.000
'Sevillan' sour orange	0.487	0.449	0.003	0.003	0.003
'Washington navel' sweet orange	0.553	0.399	0.002	0.002	0.002
'Clemenules' clementine	0.791	0.155	0.001	0.001	0.001
'Mexican' lime	0.001	0.001	0.476	0.493	0.001
'Alemow'	0.003	0.004	0.485	0.438	0.003
'Eureka' lemon	0.303	0.165	0.480	0.002	0.002
'Rangpur' lime	0.489	0.001	0.478	0.001	0.001
'Volkamer' lemon	0.491	0.002	0.473	0.001	0.001
'Rough' lemon	0.492	0.002	0.475	0.001	0.002

The factorial analysis based on *Citrus* ancestral taxa diagnostic points (considering the null or very low contribution of *M. australasica*, the corresponding diagnostic SNP were removed) display a classical picture of the relationships between secondary species and basic taxa (Figure IV.6). Clementine (11), sweet orange (12) and sour orange (13) are in a line between *C. reticulata* and *C. maxima* clusters with increasing distances with *C. reticulata*. 'Mexican' lime (14) and 'Alemow' (15) are in intermediary position between *C. micrantha* and *C. medica* cluster. 'Rangpur' lime (16), 'Rough' lemon and 'Volkamer' lemon have very close positions intermediary between *C. reticulata* and *C. medica* while Eureka lemon is slightly displaced in direction of *C. maxima* cluster.

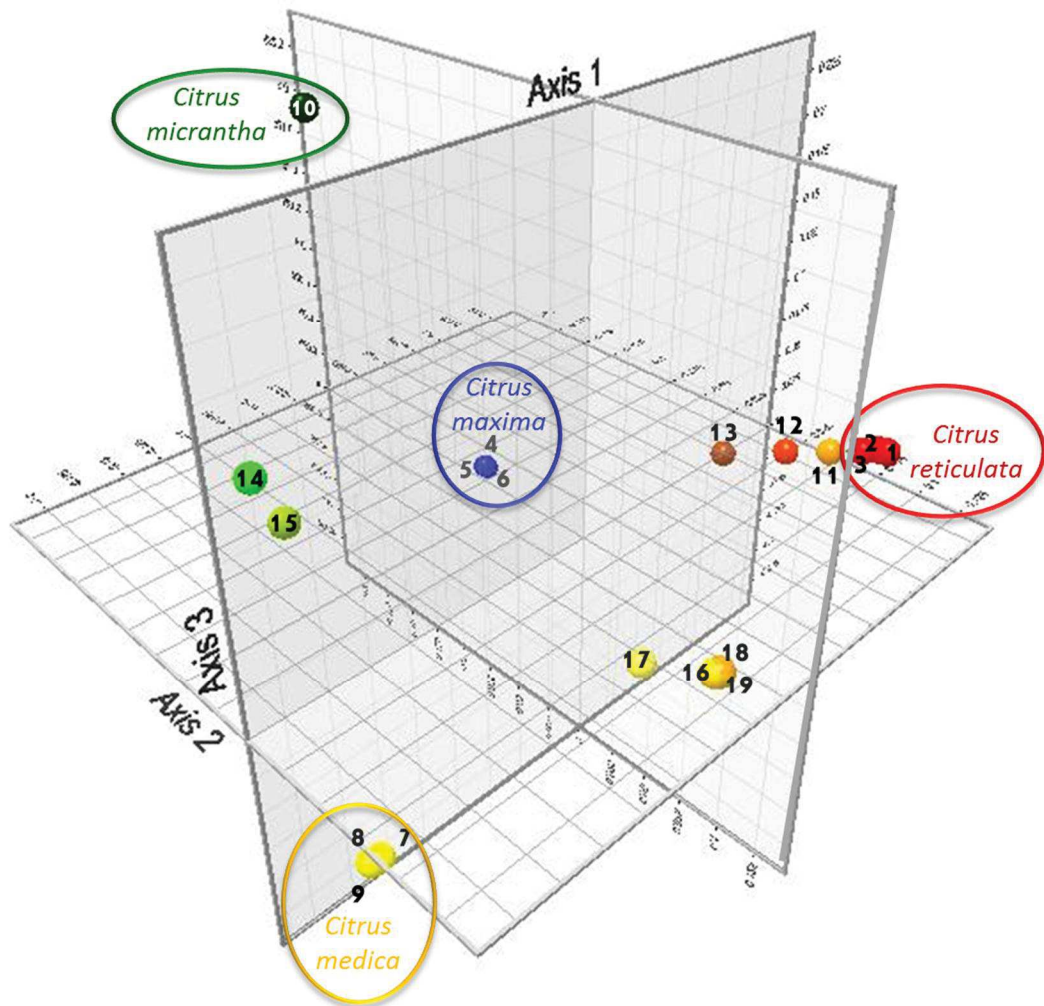


Figure IV.6. Factorial analysis from nuclear data (ancestral taxa diagnostic SNPs) 1- 'Cleopatra' mandarin; 2- 'Willowleaf' mandarin; 3- 'Huanglingmiao' mandarin; 4- 'Chandler' pummelo; 5- 'Guanxi' pummelo; 6- 'Shatian' pummelo; 7- 'Corsican' citron; 8- 'Buddha's hand' citron; 9- 'Humpang' citron; 10- *Citrus micrantha*; 11- 'Clemenules' clementine; 12- 'Washington navel' sweet orange; 13- 'Sevillan' sour orange; 14- 'Mexican' lime; 15- 'Alemow'; 16- 'Rangpur' lime; 17- 'Eureka' lemon; 18- 'Volkamer' lemon; 19- 'Rough' lemon.

### Inferred phylogenomic structures of sour orange, sweet orange and clementine validate our workflow

The methods applied to decipher phylogenomic structures is illustrated by the analysis of the chromosome 2 of Sweet Orange (Figure IV.7). The curves of heterozygosity and homozygosity frequencies of the five sets of basic taxa diagnostic SNPs (Figure IV.7B) are based on the average value of 2000 diagnostic point (1000 before and 1000 after) with average values evaluated every 1000 SNP positions.

The heterozygosity pattern (Figure IV.7A) suggests that two genomic areas are in interspecific heterozygosity and three fragments in phylogenomic homozygosity. The diagnostic allele study (Figure IV.7B) demonstrate that the area in heterozygosity have a *C. reticulata*/*C. maxima* structure while the two first homozygous fragments correspond to *C. reticulata* homozygosity and the last one to *C. maxima* homozygosity.

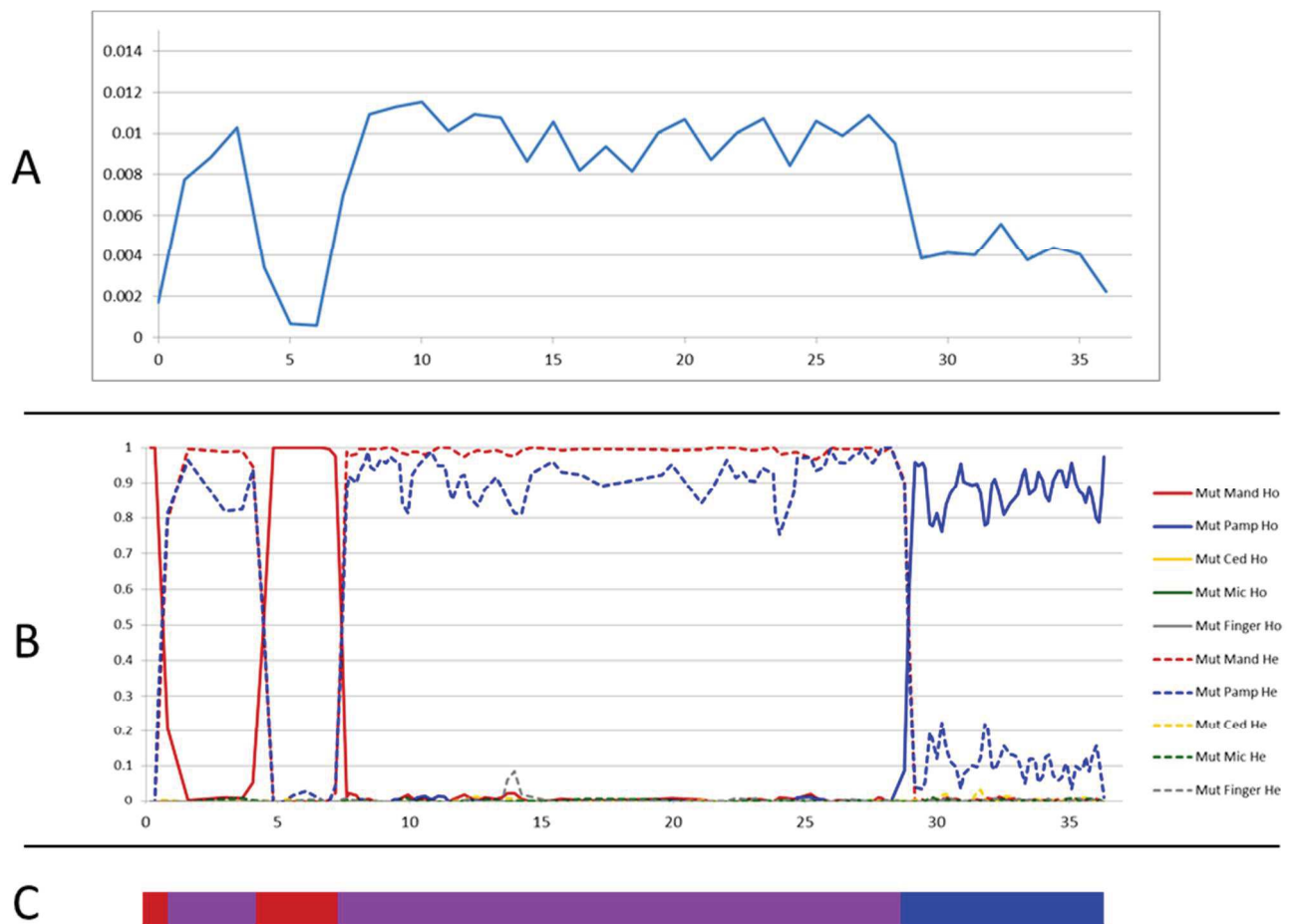


Figure IV.7. Deciphering the sweet orange phylogenomic structure of chromosome 2 by the evolution of two parameters along the chromosome. A: heterozygosity/kb; B: frequency of homozygosity and heterozygosity for the five sets of ancestral taxa diagnostic SNPs; C: phylogenomic structure of sweet orange chromosome 2 (red: *C. reticulata* homozygosity, blue: *C. maxima* homozygosity and purple: *C. maxima/C. reticulata* heterozygosity).

The nine chromosome of sour orange cv ‘Sevillano’, sweet orange cv ‘Washington Navel’ and clementine cv ‘Clemenules’ were analyzed with the same method (Figure IV.8). These cultivars only imply the contribution of two basic taxa (*C. reticulata* and *C. maxima*).

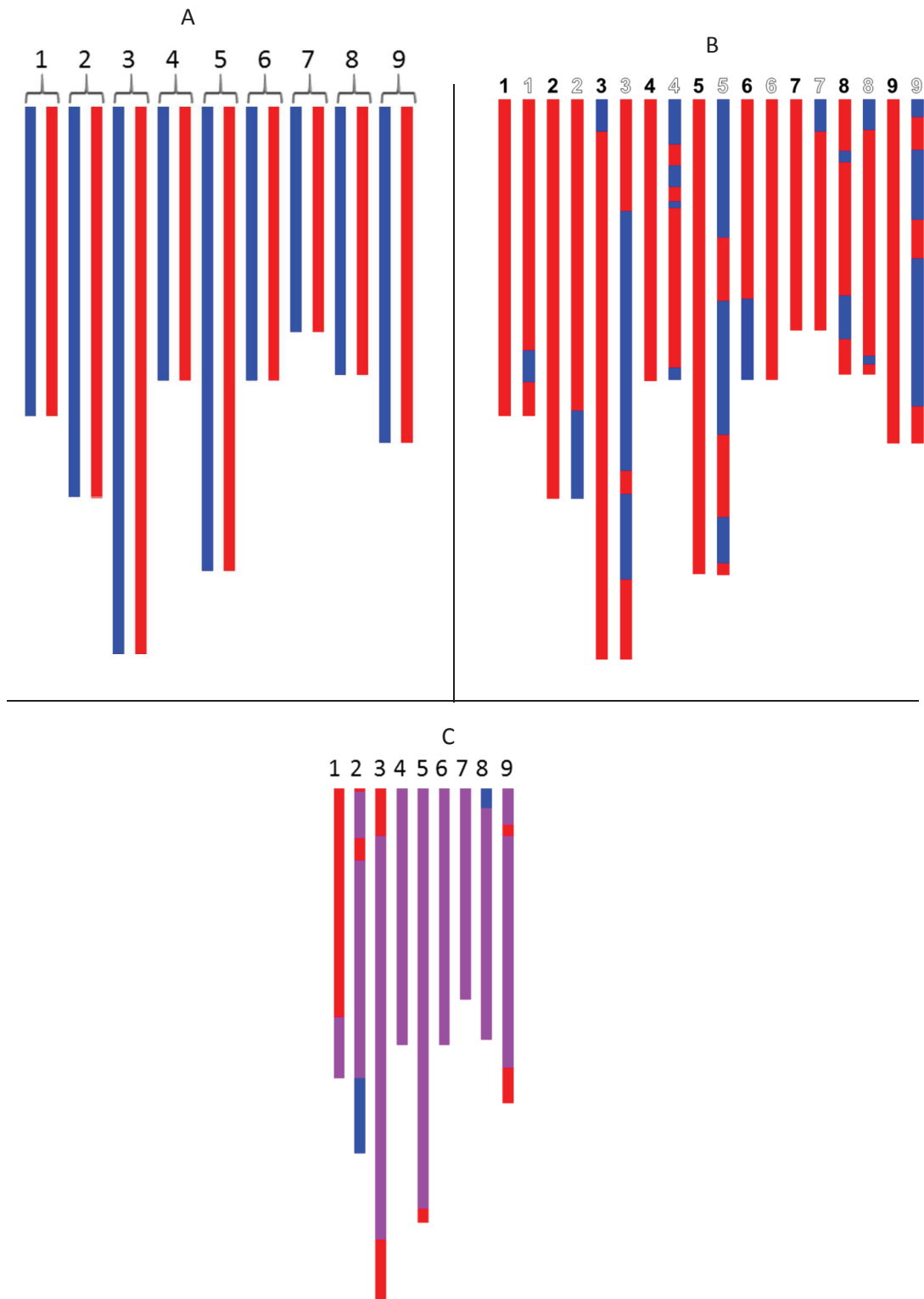


Figure IV.8. Admixture phylogenomic structures: A: Sour orange (*C. aurantium*) cv 'Sevillan' (red: *C. reticulata*, blue: *C. maxima*); B: clementine cv 'Clemenules' (red: *C. reticulata*, blue: *C. maxima*, black numbers: Willowleaf mandarin inherited chromosomes, white numbers: sweet orange inherited chromosomes); C: Sweet orange (*C. sinensis*) cv 'Washington' navel (red: *C. reticulata* homozygosity, blue: *C. maxima* heterozygosity and purple: *C. maxima* / *C. reticulata* heterozygosity)

Sour orange appears in complete heterozygosity between *C. reticulata* and *C. maxima* but the frequency of the *C. maxima* diagnostic allele in heterozygosity is slightly lower than for mandarins. Has demonstrated by the maternal phylogeny analysis, the *C. maxima* parent of the sour orange was differentiated from our representative varieties. Therefore it is probable that part of our *C. maxima* diagnostic SNPs are specific of our pummelo sample but not shared with all *C. maxima* and particularly not with the parent of sour orange. The origin of sour orange by direct *C. maxima* × *C. reticulata* hybridization proposed by Garcia *et al.* (2013b) and Wu *et al.* (2014) is validated by our study and its phylogenomic caryotype associate one *C. maxima* and one *C. reticulata* for each chromosome pair (Figure IV.8A).

Sweet orange and Clementine display more complex structure with *C. reticulata*/*C. maxima* phylogenomic heterozygosity, *C. reticulata* phylogenomic homozygosity and also for sweet orange two genomic fragments (end of chromosome 2 and start of chromosome 8) in *C. maxima* phylogenomic homozygosity. Without knowledge of the phase between heterozygous SNPs, it is not possible to have directly a complete inference of the 18 chromosome phylogenomic caryotype. However for clementine the availability of genotyping data of the 2n clementine and of the haploid clementine (the citrus reference genome sequence) allowed to identify the phylogenomic caryotype of the haploid and the caryotype of its complement to obtain the diploid clementine pattern. Moreover we take advantage of the available data of the two parents of clementine ('WillowLeaf' mandarin and sweet orange) to identify the recombination events that affected the gamete at the origin of the clementine haploid (by analyzing heterozygous clementine position in opposed homozygosity in 'WillowLeaf' and sweet orange). It was therefore possible to propose a 18 chromosome phylogenomic caryotype for clementine (Figure IV.8B), with the left chromosome coming from 'Willowleaf mandarine' and the right one from sweet orange. For sweet orange we had to keep the 9 chromosome set representation (Figure IV.8C) as previously done for the four introgressed ancestors (Figure IV.4).

The obtained patterns for these three varieties are in full agreement with the ones described by Wu *et al.* (2014) and validate our workflow.

Previous molecular studies (Bretó *et al.*, 2001; Barkley *et al.*, 2006; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b) have shown that all sweet orange cultivars originated from a single clonal parent via a series of mutations. Similar conclusion was obtained by Curk *et al.* (submitted a) for cultivars of *C. aurantium* displaying strong phenotypic differentiation (such as 'Sevillano', 'Bouquet de Fleur' and 'Chinois'). Therefore the phylogenomic pattern observed for the 'Washington navel', the 'Clemenules' clementine and 'Sevillano' sour orange should be extended to the others accessions of *C. sinensis*, clementine and *C. aurantium* respectively.

**An example of phylogenomic analysis of a complex hybrid deriving from three ancestral taxa: *C. limon* cv Eureka**

The chromosome 2 of ‘Eureka’ lemon is given as example in Figure IV.9.



Figure IV.9. Deciphering the ‘Eureka’ lemon phylogenomic structure of chromosome 2 by the pattern of two parameters along the chromosome. A: heterozygosity; B: frequency of homozygosity and heterozygosity for the five sets of ancestral taxa diagnostic SNPs; C: phylogenomic structure of ‘Eureka’ lemon chromosome 2 (red: *C. reticulata*, blue: *C. maxima* and yellow: *C. medica*).

‘Eureka’ lemon display high heterozygosity (Figure IV.9A) over all the chromosomes suggesting complete interspecific heterozygosity. It appears that close to 100% of the *C. medica* specific alleles are in heterozygosity all over the chromosome 2 (Figure IV.9B, discontinuous yellow line). These *C. medica* alleles are associated with a succession of segments with *C. reticulata* and *C. maxima* specific alleles in heterozygosity (respectively red and blue discontinuous line). Neither contribution of *C. micrantha*, nor *M. australisica* was observed. The nine chromosome of ‘Eureka’ lemon display similar patterns with heterozygosity for *C. medica* specific alleles all over the genome and a succession of *C. maxima* and *C. reticulata* fragments in heterozygosity (Figure IV.10 and Annex V, Figure S.IV.1a). Only the chromosome 9 was found in total *C. reticulata*/*C. medica* phylogenomic heterozygosity. The complete *C. medica* heterozygosity over the nine chromosomes indicates that it is a direct parent of ‘Eureka’ lemon. Therefore it is possible to infer the phylogenomic structure of the 18 chromosomes of *C. limon* cv ‘Eureka’ as described in the Figure IV.10.

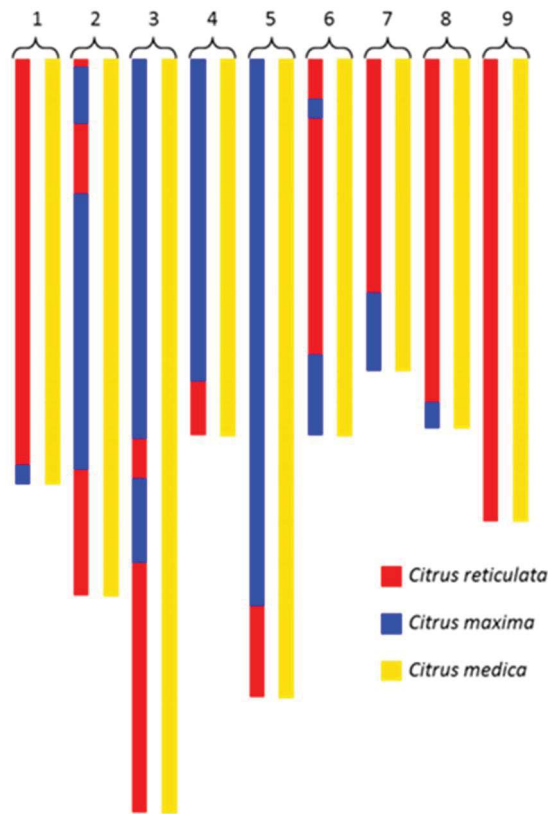


Figure IV.10. Admixture phylogenomic structure of *C. limon* cv 'Eureka'

This nuclear phylogenomic structure and the maternal phylogeny are coherent with the hypothesis of *C. aurantium* × *C. medica* origin (Nicolosi *et al.*, 2000; Gulsen and Roose, 2001; Ollitrault *et al.*, 2012b; Curk *et al.*, submitteda) and definitively exclude the *C. aurantifolia* × *C. medica* hypothesis (Scora, 1975; Barrett and Rhodes, 1976). Indeed (see below) 'Mexican' lime has no *C. maxima* nor *C. reticulata* contribution. The *C. aurantium* maternal phylogeny is attested by our results. Moreover 46% of the alleles found in sour orange but not in the BHG representative are shared by 'Eureka' lemons and found mostly in the *C. maxima*/*C. medica* genomic fragments providing additional strong indices for the *C. aurantium* × *C. medica* hypothesis. Under this hypothesis, the maternal sour orange gamete experienced 15 recombinations over the 9 chromosomes (1; 4; 3; 1; 1; 3; 1; 1; 0). Previous molecular (Gulsen and Roose, 2001; Curk *et al.*, submitteda) and cytogenetic studies (Carvalho *et al.*, 2005) have shown that numerous lemon cultivars originated from a single clonal parent (the *C. aurantium* × *C. medica* hybrid) via a series of mutations. Therefore the phylogenomic pattern observed for the 'Eureka' cultivar should be considered representative of all these cultivars.

### Phylogenomic structures of five5 diploid limes and lemons indicate that they are direct interspecific hybrids

The 5 others lime and lemons analyzed displayed patterns consistent with an origin by direct hybridization between two basic taxa (Figure IV.11).

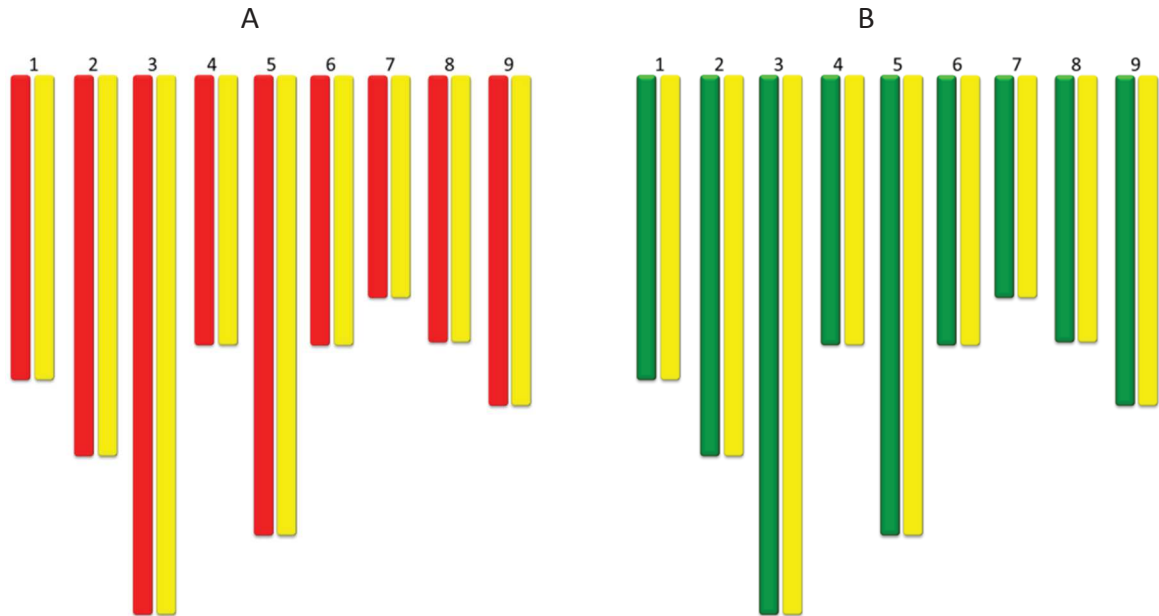


Figure IV.11. Phylogenomic structures of five limes and lemons. A: direct interspecific hybrids between *C. reticulata* (red) and *C. medica* (yellow) ('Rough' lemon, 'Volkamer' lemon and 'Rangpur' lime); B: direct interspecific hybrids between *C. micrantha* (green) or close species and *C. medica* (yellow) ('Mexican' lime and 'Alemow').

'Rough' lemon, 'Volkamer' lemon and 'Rangpur' lime display *C. reticulata* / *C. medica* heterozygosity all over their genome (Annex V, Figure S.IV.1b, c and d) and appears to result from a direct hybridization between these two taxa. These conclusions for 'Rangpur' lime and 'Rough' lemon agree with most authors' assumption (Scora, 1975; Potvin *et al.*, 1983; Federici *et al.*, 1998; Gulsen and Roose, 2001; Carvalho *et al.*, 2005; Barkley *et al.*, 2006; de Moraes *et al.*, 2007; Curk *et al.*, submitted). The origin of 'Volkamer' lemon was more controversial. Barrett and Rhodes (1976) thought mandarin × sour orange a possible parentage. While Nicolosi *et al.* (2000) considered that citron and sour orange were ancestors and Carvalho *et al.* (2005) considered from cytogenetic evidence that 'Volkamer' lemon have a mandarin × citron origin similar to 'Rough' lemon and 'Rangpur' lime. Our results agree with the last hypothesis and both cytoplasmic and nuclear data display a very close relationship between 'Rough' Lemon and 'Volkamer' lemon. Therefore from chloroplastic and nuclear data we conclude for direct *C. reticulata* × *C. medica* origin for these three genotypes.



'Mexican' lime and 'Alemow' display *C. micrantha*/*C. medica* heterozygosity for the nine chromosomes (Annex V, Figure S.IV.1e and f). However, while close to 100% of the specific alleles of *C. micrantha* are found in heterozygosity in 'Mexican' lime, it is the case for only 88% in 'Alemow' with relative homogeneity all over the genome. Our results totally discard the hypothesis of *Microcitrus* parentage for 'Mexican' lime proposed by Barrett and Rhodes (1976). Coupled with the maternal phylogeny, they agree with the hypothesis of a papeda × citron hybridization and more specifically *C. micrantha* × *C. medica* hybridization proposed by several authors (Scora, 1975; Nicolosi *et al.*, 2000; Yamamoto *et al.*, 2007; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013b; Curk *et al.*, in press; Curk *et al.*, submittedb). The origin of 'Alemow' (*C. macrophylla* in Tanaka classification) is highly controversial. *C. maxima*, *C. medica* and several papeda species were proposed as parents (Swingle and Reece, 1967; Barrett and Rhodes, 1976; Nicolosi *et al.*, 2000; Barkley *et al.*, 2006). As for sour orange we have not the true parent of 'Alemow' in our accessions representative of the basic taxa and maternal phylogeny indicate that the 'Alemow' parent was related but differentiated from our *C. micrantha* accession. Without information about *C. micrantha* intraspecific diversity we cannot conclude if the female parent of 'Alemow' was a *C. micrantha* or a close related papeda species. It is however evident that 'Alemow' results from direct hybridization between a Papeda species closely related to *the* "small flowered papeda" accession pollinated by a citron, in agreement with the hypothesis of Ollitrault *et al.* (2012b) recently confirmed by Curk *et al.* (in press; submitted b).

#### **Phylogenomic structure of the triploid 'Bears' lime**

The triploid 'Bears' lime was analysed with the same approach than the diploid varieties. As displayed in Figure IV.12A, it revealed a succession of genome segment in phylogenomic heterozygosity for two or three parents but also for very short fragments in homozygosity for *C. medica*. The three allelic fragments revealed the complete phylogenomic structures but for fragments displaying two ancestral taxa heterozygosity (mostly *C. medica*/*C. micrantha*) it was necessary to estimate the relative frequency of the two concerned taxa alleles from the relative numbers of reads (Figure IV.12B) to reveal the relative doses of the concerned ancestral taxa (Figure IV.12C). The same process was applied for all chromosomes (Annex V, Figure S.IV.2a,b,c,d,e,f,g,h and i).

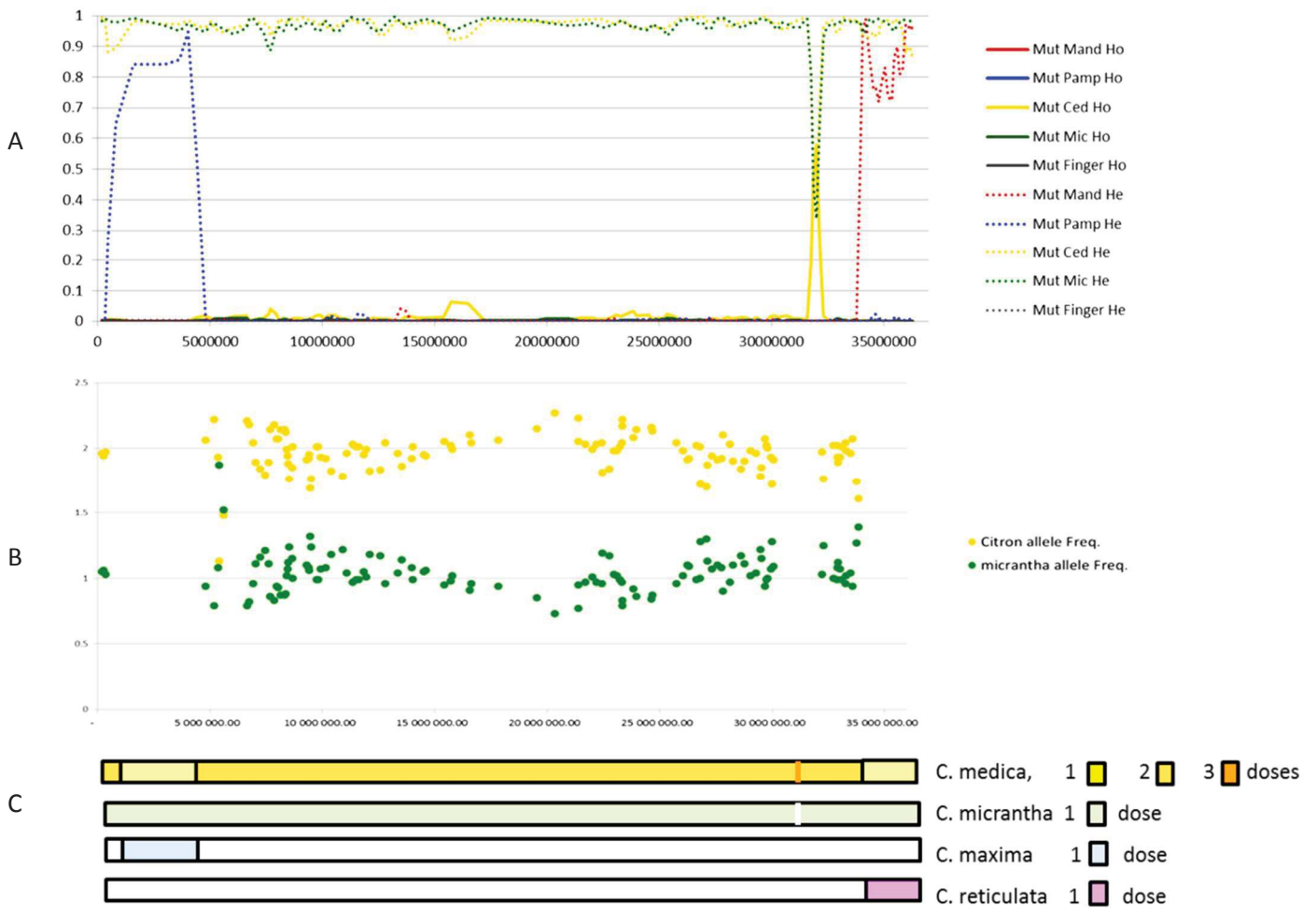


Figure IV.12. Phylogenomic structure of the triploid 'Bears' lime for the chromosome 2. A: frequency of homozygosity and heterozygosity for the five sets of ancestral taxa diagnostic alleles; B: relative numbers of reads in the genomic regions in heterozygosity between two taxa (*C. medica* and *C. micrantha* for the concerned chromosome); C : inferred proportion of each ancestral taxa along the chromosome.

According to Webber (1943), the 'Bears' variety originated about 1895 on the place of J.T. Bearss, a nurseryman at Porterville, California. While the facts are unknown, it presumably occurred as a nucellar seedling of a tree grown from seed from a fruit of Tahitian origin. Previous molecular studies indicated that 'Bears', 'Tahiti' and 'Persian' limes have a same clonal origin. These three triploid varieties represent today the main international production of lime (*Citron vert* in the French market). Before our study their origin was uncertain. Moore (2001) proposed that these triploid limes resulted probably of a tri-hybrid interspecific cross of citron (*Citrus medica*), pummelo (*Citrus grandis*), and a papeda species, *Citrus micrantha*. Recently from cytoplasmic and nuclear molecular marker study (Curk *et al.*, submittedb) proposed that Tahiti, Persian and 'Bears' lime should results from the fecundation of a lemon ('Eureka'/Lisbon' type) haploid gamete by a diploid gamete of lime ('Mexican' type). The total identity between 'Bears' lime and 'Eureka' lemon for *C. reticulata* and *C. maxima* contribution in chromosome 4, 6 and 7 and the fact that all genomic areas of *C. reticulata* and *C. maxima* found in 'Bears' lime were also observed in 'Eureka lemon' as well as the maternal phylogeny confirm the hypothesis of a 'yellow lemon' maternal origin. The frequent double dosis of *C. medica* and the occurrence of double dosis of *C. micrantha* in chromosome 3 and 5 while *C. reticulata* and *C. maxima* were found only in

single dosis demonstrated that the diploid gamete arose from a ‘Mexican’ lime like parent. We propose a phylogenomic caryotype of ‘Bears’ lime under the hypothesis of a haploid ‘yellow lemon’ ovule fecundated by a diploid ‘Mexican’ lime like pollen (Figure IV.13). For chromosome 3 and 5 it is just one of the different possibility of *C. medica* and *C. micrantha* fragment phases. Interestingly total restitution of the ‘Mexican’ lime like parent is inferred for chromosome 1, 2, 4, 6, 8 and 9. Therefore the diploid gamete of the ‘Mexican’ lime like parent should results from a doubled diploid ‘Mexican’ lime like parent (for example the tetraploid ‘Giant Key’ lime) under the hypothesis of preferential disomic segregation and eventually of an unreduced FDR gamete considering that the centromere regions of the nine chromosome (Aleza *et al.*, Submitted) were transferred in heterozygosity.

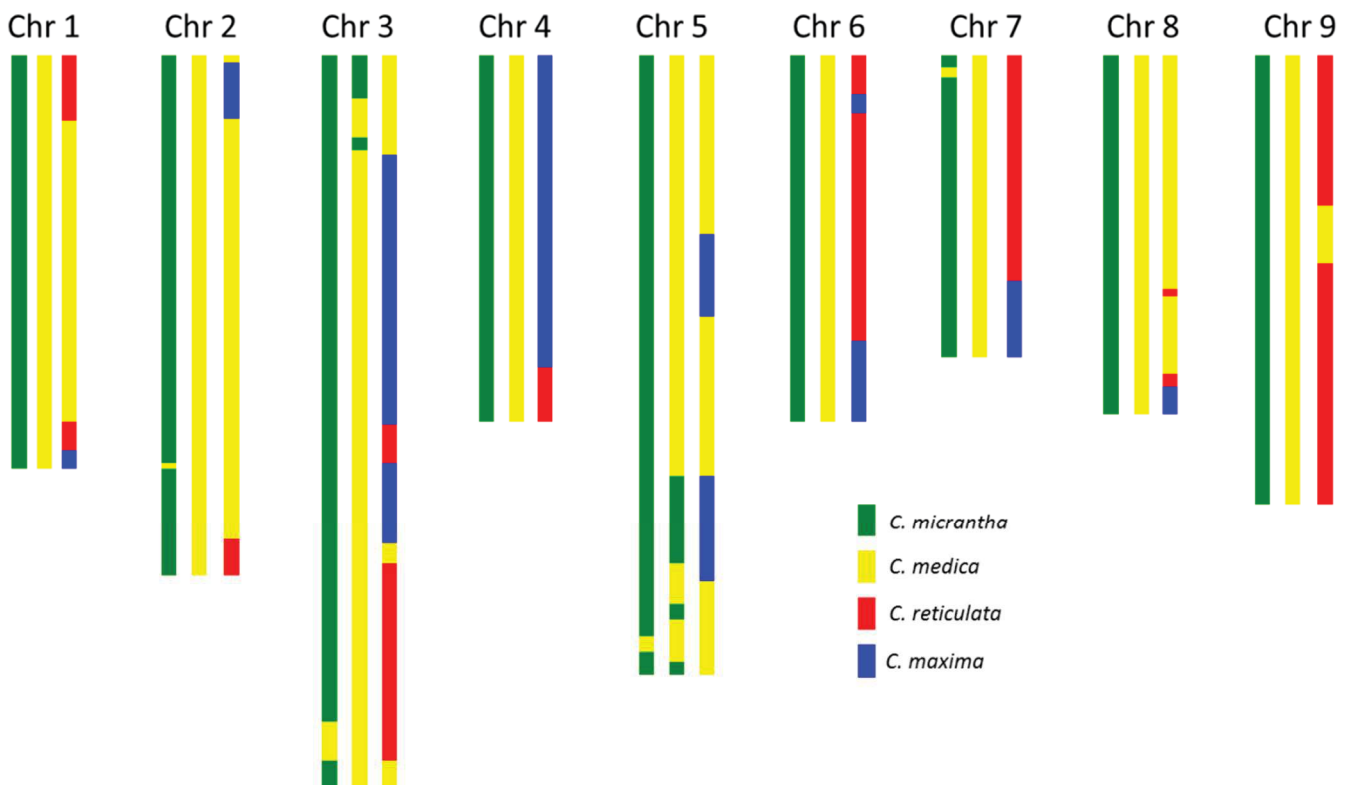


Figure IV.13. Proposed phylogenomic caryotype of ‘Bears’ lime under the hypothesis of a Yellow lemon haploid ovule by ‘Mexican’ lime like diploid pollen origin

## Conclusions

We established a successful workflow to analyse the phylogenomic structure of moderns diploid and triploid citrus varieties from WG re-sequencing data mapped in the reference citrus genome and filtered on gene sequences. *C. maxima* introgressions were identified in the three mandarin cultivars. However it is limited to one small fragment for 'Cleopatra' mandarin that appears as the purest *C. reticulata* representative. A small *C. reticulata* introgression was also observed in 'Chandler' pummelo. More than one million of diagnostic SNPs of the five analyzed ancestor were identified from their inferred allelic frequencies after the elimination of the introgressed fragments for the concerned accessions. The analysis of the frequencies of homozygosity and heterozygosity of these ancestral diagnostic SNPs along the genome revealed the phylogenomic structures of the 9 chromosomes of the 20 diploid and one triploid varieties. There complete agreement with the Wu *et al.* (2014) results for 'Willowleaf' and 'Huanglingmiao' mandarins, sweet orange, sour orange and clementine validate our workflow. The lime and lemon nuclear phylogenomic structures and maternal phylogeny revealed that five of them resulted from direct interspecific hybridization. We conclude for *C. reticulata* × *C. medica* hybridization for 'Volkamer' lemon, 'Rough' lemon and 'Rangpur' lime while 'Mexican' lime and 'Alemow' results from papeda (*C. micrantha* or closely related) × *C. medica*. More complex evolutive sequences were inferred for 'Eureka' lemon and 'Bears' lime. They confirmed the *C. aurantium* × *C. medica* origin for 'Eureka lemon' and we concluded that 'Bears' lime resulted from the pollination of *C. limon* (Eureka type) haploid gamete by a diploid gamete of *C. aurantifolia* ('Mexican' lime type). The phylogenomic karyotype of these varieties issued from three (Eureka lemon= (*C. maxima* × *C. reticulata*) × *C. medica*) and four ('Bears' lime= (*C. maxima* × *C. reticulata*) × (*C. micrantha* × *C. medica*)) ancestral species are provided for the first time. Considering the asexual intraspecific diversification of most secondary species the phylogenomic structures revealed in these study for Clementine cv 'Clemenules', sweet orange cv 'Washington Navel', sour orange cv 'Sevillano', should be extended to all clementine, *C. sinensis* and most *C. aurantium* accessions respectively. The structure of 'Eureka' lemon should also be considered as representative of many 'yellow lemons' and the one of the triploid 'Bears' lime should be shared with 'Tahiti' and 'Persian' limes. These new insight in limes and lemon origin and phylogenomic structure are decisive to develop innovative breeding schemes for these horticultural groups.

## Contributions

This chapter results from collaboration between the Center of Genomic of Ivia, the Spanish Citrusseq-Citrusgenn Consortium, Inra and Cirad. P. Ollitrault (UMR Agap) and M. Talon (IVIA) planned the experiment. Inra/Cirad provided genetic resources for the project. J. Terol from Ivia, J. Carbonell, R. Alonso and J. Dopazo from the CIPF provided SNP matrices from the Citrusseq-Citrusgenn Consortium database. P. Ollitrault and F. Curk (UMR Agap) implemented the analysis method to identify species-diagnostic SNPs and to decipher the phylogenomic structures of the considered secondary species. F. Curk analysed the data of diploid genotypes. F. Curk and V. Ibanez analyzed the structure of the triploid 'Bears' lime. F. Curk and P. Ollitrault wrote the chapter.

## Acknowledgement

This work has been possible under the agreement and thanks to the Citrusseq-Citrusgenn Consortium. The data used in this chapter came from a project which received, in addition to the Citrusseq-Citrusgenn Consortium, funding from the *Ministerio de Ciencia e Innovación* through FEDER grants PSE-060000-2009-8 and IPT-010000-2010-43 to Manuel Talón and Joaquin Dopazo; by AGL2011-30240 grant to Javier Terol and by grants BIO2011-27069 from the Spanish Ministry of Economy and Competitiveness (MINECO) and PROMETEO/2010/001 from the *Conselleria de Educació* of the Valencia Community to Joaquin Dopazo

---

## Discussion

---

Comme nous l'avons vu dans l'introduction, les agrumes, présentent une grande diversité phénotypique. Les études de marquages moléculaires indiquent que quatre taxons ancestraux (*C. reticulata* les mandariniers, *C. maxima* les pamplemoussiers, *C. medica* les cédratiers et *C. micrantha*) sont à l'origine de l'ensemble des formes cultivées suite à des événements de réticulations. Il en résulte des structures génotypiques complexes, généralement fixées par l'apomixie, fortement hétérozygotes et formées d'une mosaïque de grands fragments d'origine phylogénétique différente. Ce travail de thèse a exploré différentes approches d'analyse de la diversité des génomes afin d'identifier l'origine phylogénétique des principaux groupes horticoles et de décrypter la structure phylogénomique des variétés modernes. Une analyse plus poussée a été conduite sur les limettiers et citronniers, moins étudiés que la plupart des autres agrumes. La thèse a tiré parti de l'évolution très rapide des NGS et propose une utilisation raisonnée des outils disponibles en fonction des questions de recherche. Le séquençage parallèle d'amplicons par la technique 454 (Roche) permet d'obtenir une information haplotypique multi-loci sur des fragments d'environ 500 bases. Il a été utilisé (i) pour décrypter la structure en mosaïque interspécifique du chromosome 2 de cinquante variétés et (ii) pour identifier des marqueurs diagnostiques des quatre taxons ancestraux. En association avec des marqueurs SSR et indels, ces marqueurs ont permis, d'apporter un éclairage nouveau sur l'origine des limettiers et citronniers, par un génotypage exhaustif des collections de l'Inra/Cirad et de l'ivia. Enfin, les données de re-séquençage complet Illumina de sept variétés de limettiers et de citronniers comparées à celles de quelques représentants des taxons ancestraux nous ont permis de reconstituer la structure interspécifique de leur génome complet et de schématiser pour la première fois leur caryotype phylogénomique. L'intérêt relatif des différentes approches d'analyse des génomes en fonction des objectifs de recherche ainsi que les principales conclusions sur l'origine des espèces secondaires et des variétés modernes sont discutés dans les points suivants.

## **1. Intérêt et limites de différentes approches en fonction des questions de recherche**

### **1.1. L'Haplotypage par séquençage parallèle avec la technologie 454 (Roche)**

L'évolution réticulée de nombreuses plantes cultivées rend les études phylogénétiques plus compliquées. Les événements de réticulation sont à l'origine d'histoires évolutives qui ne peuvent pas être représentées correctement par des arbres phylogénétiques (Stebbins, 1950; Grant, 1981; Arnold, 1997; Doolittle, 1999; Otto and Whitton, 2000). Elles se traduisent en effet par de fréquentes divergences entre les relations phylogéniques, révélées par les génomes cytoplasmiques et nucléaires, mais également entre celles inférées à partir de régions différentes du génome nucléaire (Pamilo and Nei, 1988; Rieseberg and Soltis, 1991; Linder and Rieseberg, 2004; Beiko and Hamilton, 2006). Si une forme de multiplication végétative (apomixie par exemple) devient prédominante immédiatement après l'évènement de réticulation ou après un nombre limité de générations sexuées, comme cela semble avoir été le cas pour les agrumes, des portions très larges du génome restent en hétérozygotie interspécifique. Des études portant sur l'ensemble du génome sont alors nécessaires pour décrypter les structures génomiques interspécifiques complexes résultant de cette évolution.

Dans un tel contexte, les études basées sur les déséquilibres de liaison renforcent la puissance des analyses phylogénétiques. Elles ont par exemple permis de montrer que le génome du tournesol était constitué de fragments chromosomiques hérités d'espèces différentes (Rieseberg, 1996; Rieseberg *et al.*, 2003). Avec des structures génomiques en hétérozygoties interspécifiques comme pour les agrumes, la capacité à conduire des analyses haplotypiques multilocus nous apparaissait de nature à renforcer considérablement notre capacité à assigner une origine phylogénétique aux fragments génomiques par rapport à des analyses monolocus. Le séquençage parallèle d'amplicons par la méthodologie 454 (Roche) qui fournit pour des coûts abordables des lectures parallèles de 300 à 400 bases à partir de fragments uniques, semblait pouvoir apporter une réponse adaptée à notre problématique. En effet, comparé à d'autres approches de laboratoire pour obtenir des données haplotypiques, comme le clonage d'amplicons suivi de séquençage Sanger (Fortune *et al.*, 2008; Rousseau-Gueutin *et al.*, 2009; Ramadugu *et al.*, 2013) le séquençage parallèle par les NGSs d'amplicons « tagués » est une solution efficace et abordable. Chez l'homme, le séquençage 454 d'amplicons a par exemple permis de produire des milliers d'haplotypes du locus de la "beta-defensin" (Taudien *et al.*, 2010). Dans le cadre de cette thèse, ce type de séquençage a été appliqué avec succès pour déterminer les haplotypes de fragments génomiques pour des cultivars fortement hétérozygotes d'agrumes. Comme pour toute méthode basée sur la PCR, une limitation de l'approche est la possible absence d'amplification ou la compétition allélique au sein de structures hétérozygotes si des polymorphismes existent au niveau des séquences d'amorçage comme cela a été relevé par Ramadugu *et al.* (2013) dans le cadre de leur étude sur les agrumes par séquençage Sanger d'amplicons. Le niveau de différenciation identifié chez les agrumes entre les quatre taxons ancestraux (entre 10,61 et 14,8 SNP/kb) est environ sept fois plus élevé que le polymorphisme intra spécifique. Nous avons pu, pour 12 des 16 fragments de gènes du chromosome 2, identifier l'origine phylogénétique des haplotypes des 50 variétés analysées. Cette assignation phylogénétique des haplotypes a en particulier révélé des introgressions de *C. maxima* dans le génome de variétés considérées comme de pures mandarines. L'identification de ces introgressions a permis une meilleure estimation des fréquences alléliques des taxons ancestraux et par là même l'établissement d'un meilleur arbre phylogénétique ainsi qu'une identification optimisée de marqueurs/allèles diagnostiques de ces taxons. Comme l'ont recommandé Ramadugu *et al.* (2013) les arbres phylogénétiques d'espèces peuvent être améliorés par l'utilisation de locus plus structurants au niveau des arbres de chaque fragment chromosomique. Pour un tel objectif, les analyses haplotypiques sont précieuses pour écarter les régions génomiques présentant des introgressions interspécifiques. L'approche séquençage 454 d'amplicons par la méthodologie 454 répond donc bien à notre problématique de recherche. Nous n'avons pourtant pas pu l'exploiter au mieux car nous avons été confrontés à une importante hétérogénéité du nombre de lectures par fragments, avec moins de 20 lectures pour plus de 7 % des fragments séquencés et moins de 50 lectures pour près de 20 % de fragments. De plus, les erreurs de séquençage, dont le taux peut être élevé notamment sur les zones riches en homopolymères (répétition de la même base), ont fait apparaître de faux indels entre génotypes, ce qui a particulièrement complexifié le traitement des données. Enfin, la reconstitution des haplotypes et leur alignement n'ont finalement pas pu être automatisés avec succès et la gestion « séquence par séquence », dont les résultats nous ont permis l'étude du chromosome 2, a été particulièrement chronophage et n'a pu être étendue à l'ensemble des chromosomes.



## 1.2. Les marqueurs SNP diagnostiques des taxons ancestraux

Compte tenu de leurs caractéristiques (polymorphisme élevé, dispersion sur tout le génome, caractère codominant), de nombreux marqueurs SSR ont été développés chez les agrumes ces 15 dernières années<sup>1</sup>. Toutefois, l'utilité de ces marqueurs pour tracer l'origine phylogénétique des fragments d'ADN est limitée du fait des phénomènes d'homoplasie (Barkley *et al.*, 2009). Ainsi, diverses études de génotypage attestent de la présence très fréquente d'allèles communs à au moins deux taxons ancestraux : *C. reticulata* et *C. maxima* (Garcia-Lor *et al.*, 2012; 2013b). Les polymorphismes microsatellites, qu'ils soient génomiques ou inclus dans des EST, ne permettent donc pas d'atteindre un niveau suffisant de spécificité pour répondre efficacement à nos objectifs. Garcia-Lor *et al.* (2012) et Ollitrault *et al.* (2014) ont montré que les indels et les SNP présentaient un potentiel bien plus élevé que les SSR pour sélectionner des marqueurs diagnostiques des taxons ancestraux des agrumes. Ainsi, les études récentes de phylogénies nucléaires à partir de séquençage Sanger de fragments de gènes (Garcia-Lor *et al.*, 2013b; Ramadugu *et al.*, 2013) ont révélé des SNP différenciant les quatre taxons de base dont certains ont fait l'objet d'un développement réussi de marqueurs de type Kaspar (Garcia-Lor *et al.*, 2013a). Plusieurs marqueurs SNP différenciant *C. maxima* et *C. reticulata* ont également été développés à partir de données de Bac-end de clémentinier (Ollitrault *et al.*, 2012). Par ailleurs, les données de re-séquençage du génome de divers mandariniers et pamplemoussiers ont révélé des centaines de milliers de SNP différenciant *C. maxima* et *C. reticulata* (Wu *et al.*, 2014). Cependant le nombre de marqueurs diagnostiques d'espèces utilisables en routine restait très limité, en particulier pour *C. medica* et *C. micrantha*. Nous avons identifié de nouveaux SNP diagnostiques des quatre taxons ancestraux à partir des données de séquençage d'amplicons (454) de 26 variétés représentatives des quatre taxons ancestraux pour 57 fragments de gènes répartis sur les 9 chromosomes. Deux cent soixante et onze des 1053 SNP identifiés présentaient une spécificité allélique très élevée pour l'un des taxons de base ( $G_{ST} > 0.9$ ) et 73 marqueurs Kaspar diagnostiques de l'un des taxons ont été développés avec succès. En combinaison avec 32 SNP diagnostiques précédemment développés, ils ont permis de révéler la constitution interspécifique de nombreux porte-greffe et variétés actuels. De fortes corrélations ont été observées entre ces estimations réalisées à l'aide des marqueurs diagnostiques et celles obtenues à partir des données de re-séquençage complet de sept variétés (Wu *et al.*, 2014), validant la valeur phylogénétique de ces marqueurs. Nos résultats confirment que l'on peut réaliser une sélection efficace de marqueurs diagnostiques d'espèces à partir des polymorphismes SNP et que ces marqueurs sont performants pour tracer l'origine phylogénétique des fragments d'ADN qui les portent. De plus, le développement de ce type de marqueurs peut potentiellement se faire à très grande échelle et de nombreuses solutions de génotypage existent en fonction des quantités de marqueurs et d'individus à analyser. Ils peuvent ainsi être utilisés en routine pour le génotypage d'un nombre important de variétés et donc pour la caractérisation des différentes collections variétales en vue d'évaluer la contribution des taxons ancestraux aux différentes variétés et porte-greffe. Toutefois une limitation très importante des marqueurs SNP développés est leur di-allélisme et leur incapacité à révéler des polymorphismes nouveaux par rapport à ceux issus du panel d'identification (*SNP mining* à partir des données 454 de 26 variétés dans notre cas). Ils peuvent donc conduire à des images biaisées de la diversité si ils sont appliqués à des populations trop différenciées de celles du panel d'identification

<sup>1</sup> cf. point 7.1.9. de l'introduction : SSR (*Simple Sequence Repeats*)

(Clark *et al.*, 2005; Rosenblum and Novembre, 2007; Albrechtsen *et al.*, 2010; Negrini *et al.*, 2010; Bradbury *et al.*, 2011; Trask *et al.*, 2011). Les marqueurs que nous avons développés n'ont donc pas d'intérêt pour des études élargies aux genres apparentés (*Poncirus*, *Fortunella*, *Microcitrus* etc.) et pourraient conduire à des conclusions erronées pour certains papedas (*Citrus* sauvages) compte tenu de leur très faible représentativité dans le panel d'identification. Enfin, nous avons sélectionné ces marqueurs pour leur capacité à différencier les taxons ancestraux. De fait, ils révèlent très peu de polymorphisme intra taxa. C'est donc la combinaison des deux types d'informations (SNP et SSR, polymorphisme respectivement majoritairement inter et intra spécifique) qui permettra une étude équilibrée des collections de ressources génétiques. C'est cette approche que nous avons appliquée pour l'analyse de la diversité et de l'origine des limettiers et citronniers.

### 1.3. Les données de re-séquençage WGS

Les NGS ont modifié considérablement la manière d'analyser l'évolution des plantes, passant d'études phylogénétiques à des études phylogénomiques basées sur l'analyse de la diversité sur l'ensemble du génome. Des connaissances nouvelles ont ainsi été apportées sur la domestication des différentes espèces fruitières (Myles *et al.*, 2011; Cornille *et al.*, 2012) et céréalières (Huang *et al.*, 2012; Hufford *et al.*, 2012). La libération par le consortium international de génomique des agrumes (ICGC, 2014; Wu *et al.*, 2014) de la première séquence de référence de haute qualité des agrumes établie à partir d'un haploïde de clémentinier a été une étape clef pour le développement de la phylogénomique chez les agrumes. Ainsi récemment, le re-séquençage de diverses variétés de mandariniers, de pamplemoussiers, de clémentiniers, d'orangers et de bigaradiers (Wu *et al.*, 2014) a révélé des introgressions de *C. maxima* dans des variétés de mandariniers généralement considérées comme représentatives de *C. reticulata* ainsi que la structure en mosaïque interspécifique de différentes variétés. Cette étude portait sur un groupe d'espèces et de variétés issues des pools géniques de deux taxons ancestraux : *C. maxima* et *C. reticulata*. Dans le cadre de cette thèse nous avons proposé une approche permettant de prendre en compte un grand nombre d'ancêtres potentiels et nous nous sommes focalisés sur l'analyse des agrumes acides pour lesquels une contribution des quatre taxons ancestraux était supposée (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Bayer *et al.*, 2009; Ollitrault *et al.*, 2012; Garcia-Lor *et al.*, 2013b; Curk *et al.*, accepted; Curk *et al.*, submitted a; Curk *et al.*, submitted b). Nous avons travaillé à partir de matrices de SNP incluant toutes les positions polymorphes pour au moins l'une des variétés relativement à la séquence de références. Ces matrices fournies par le Centre de génomique de l'lvia (à partir des bases de données de re-séquençage complet d'agrumes du consortium CITRUSSEQ) sont issues de l'ancrage sur la séquence de référence des données de re-séquençage illumina et d'une inférence des SNP par *multicalling* (GATK). Afin de limiter les erreurs de génotypage, la matrice de SNP a été filtrée sur les séquences géniques mieux conservées.

Nous avons mis en place un processus d'analyse efficace basé sur (1) l'identification de SNP diagnostiques des différents ancêtres supposés et (2) l'analyse de la fréquence des allèles spécifiques de chaque taxon ancestral en homozygotie et hétérozygotie sur tout le génome. Pour cela il a fallu, dans un premier temps, rechercher les introgressions interspécifiques dans le génome des variétés considérées comme représentatives des espèces ancestrales sur la base de SNP structurants les différentes populations. Une fois les zones introgressées identifiées, les fréquences alléliques des taxons ancestraux ont été

réévaluées en éliminant du calcul les variétés introgressées dans ces zones concernées pour un taxon donné. Nous avons retenu comme SNP diagnostiques ceux différenciant totalement un taxon ancestral des autres sur la base de ces fréquences réévaluées. Plus d'un million de SNP diagnostiques ont été ainsi identifiés pour les cinq taxons considérés (*C. reticulata*, *C. maxima*, *C. medica*, *C. micrantha* et *M. australasica*). La parfaite cohérence de nos résultats avec ceux de Wu *et al.* (2014) pour les quatre variétés communes à notre étude (oranger Washington Navel, bigaradier Sevillano, clémentinier Clemenules, mandariniers Willow leaf et Huanglimiao) valide notre processus d'analyse. Il a permis de révéler les structures phylogénomiques de variétés diploïdes et triploïdes de limettiers et citronniers, combinant les génomes de deux, trois et même quatre taxons ancestraux pour ce qui concerne la lime triploïde Bears. Les structures révélées pour les limettiers et citronniers sont discutées point 2 de la discussion. Les données de re-séquençage complet du génome ont donc permis d'étudier de manière fine les structures phylogénomiques des agrumes sans *a priori* sur les polymorphismes. La très haute densité d'information est sans commune mesure avec ce que pourraient apporter des méthodes de génotypage classique, même à haut débit. L'ensemble des données de génotypage disponibles pour cette étude a été obtenu par ancrage sur l'assemblage de l'haploïde de clémentinier produit par le consortium international ICGC (2014). Il est donc important de prendre en compte le fait que le génome du clémentinier serait jusqu'à 10% plus petit que celui du cédratier par exemple (Ollitrault and Navarro, 2012). Nous n'avons pour l'instant pas d'informations détaillées sur les différences structurelles qui pourraient exister entre les taxons ancestraux, ni sur l'importance et la nature des variations présentes au niveau des zones non codantes (dont les séquences répétées par exemple). Il s'agit d'une limite importante de cette approche, basée sur l'ancrage sur un génome principalement représentatif de *C. reticulata*, confirmée par la grande quantité de séquences non ancrées et le « bruit de fond », observés lors de l'étude des SNP diagnostiques répartis sur tout le génome. Afin d'atténuer fortement les effets de ce « bruit de fond » l'étude a ainsi été recentrée sur les SNP diagnostiques situés sur les zones géniques considérées comme mieux conservées et soumises à moins d'aléas en terme d'ancrage et d'inférence des polymorphismes. Au-delà des limites dépendantes de la qualité du séquençage, de l'ancrage et des méthodes d'inférences des polymorphismes (*SNP* et *indel calling*), les coûts de séquençage restent encore élevés pour des études portant sur de grandes populations. Enfin, l'immense quantité de données générée, requiert encore aujourd'hui des outils informatiques puissants et des compétences spécifiques en bio-informatique pour générer des matrices de SNP. Voilà pourquoi, malgré l'intérêt incontestable d'avoir accès aux données de l'ensemble du génome, les approches couplant les NGS et les techniques basées sur la réduction de la complexité des génomes telles que le Génotypage par Séquençage (GBS : Elshire *et al.*, 2011; Deschamps *et al.*, 2012) ou les *Restriction site associated DNA (RAD sequencing* : Baird *et al.*, 2008; Davey and Blaxter, 2011; Peterson *et al.*, 2012) semblent prometteuses. Ces techniques permettent le multiplexage de nombreux individus, ne nécessitent pas obligatoirement de génome de référence et sont particulièrement compétitives en terme de coût. L'intérêt de la GBS pour l'étude des structures interspécifiques des agrumes cultivés est en cours d'évaluation au sein de notre groupe de recherche.

## 2. Origine et structure des différentes espèces cultivées

Les différentes approches que nous avons choisies pour étudier la structure génétique des espèces cultivées nous ont conduites à des conclusions convergentes. Nos résultats confirment les hypothèses antérieures concernant l'évolution réticulée des espèces secondaires tout en mettant en évidence de fréquentes introgressions de *C. maxima* dans le génome des variétés de mandariniers, pourtant considérées comme représentatives de *C. reticulata*.

Alors que presque tous les autres groupes d'espèces secondaires d'agrumes (orangers, bigaradiers, pomélos, hybrides de mandariniers, dont le clémentinier, les tangors et les tangelos) proviennent d'hybridations n'impliquant que deux taxons de base, *C. reticulata* et *C. maxima* (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Ollitrault *et al.*, 2012; Garcia-Lor *et al.*, 2013b), le groupe des limettiers et citronniers apparaît comme étant bien plus complexe. En plus de regrouper des individus diploïdes, triploïdes et tétraploïdes, on y retrouve quatre types cytoplasmiques sur les six identifiés pour tous les agrumes du genre *Citrus*. Enfin, toutes les variétés de limettiers et de citronniers que nous avons étudiées ont été trouvées avec un taux d'hétérozygotie élevé, avec des combinaisons interspécifiques impliquant deux, trois et parfois même les quatre génomes des taxons ancestraux pour certaines variétés.

Grâce à l'accès aux données de re-séquençage complet du génome de sept limettiers et citronniers nous avons pu, pour la première fois, proposer le caryotype phylogénomique d'un certain nombre de variétés diploïdes ainsi que celui d'un limettier triploïde (Chapitre IV).

Nos résultats démontrent que ni la classification de Swingle et Reece (1967) qui organise les limes et les citrons en deux espèces, ni la classification de Tanaka (1954), ni même les dénominations vernaculaires de limes et de citrons ne correspondent à la réelle structuration phylogénétique de ce groupe.

Nous confirmons certaines hypothèses antérieures et en proposons de nouvelles sur l'origine de presque toutes les variétés du groupe des limettiers (diploïdes, triploïdes et tétraploïdes) et citronniers présentes dans les collections de l'lvia de Valence, en Espagne et du CRB Citrus Inra-Cirad (NF S96-900) de San Giuliano en Corse. Tous ces résultats sont commentés et organisés dans les points suivants en fonction de la classification de Swingle et Reece (1967) pour les groupes principaux et selon la classification de Tanaka (1961) pour les sous-groupes. Enfin, une synthèse est présentée figure D1.

### 2.1. Les taxons de base

#### 2.1.1. *Citrus medica*, les cédratiers

Aucune introgression des autres taxons de base n'a été trouvée dans le génome des vrais cédratiers dont nous avons des données d'haplotypage du chromosome 2 (chapitre I), de marquage moléculaire (chapitres II et III) ou de séquençage complet (chapitre IV). En revanche, certaines variétés portant le nom usuel de « cédrat » sont en fait des hybrides d'origine interspécifique. Ainsi, les cédratiers de Damas, Mak Nao Si, ou Rhobs el Arsa, ont un type cytoplasmique identique aux citronniers de type Lisbonne et bigaradiers (figure III.1 et Annexe IV, figure S.III.2). Les marqueurs nucléaires utilisés mettent en évidence des structures hybrides (*C. reticulata* × *C. maxima*) × *C. medica* (figures III.3 et III.4) comme précédemment identifié par Luro *et al.* (2012). A l'inverse, le citronnier chinois n'est pas un citronnier mais bien un vrai *C. medica*.

### 2.1.2. *C. maxima*, les pamplemoussiers

Les pamplemoussiers, bien que mono-embryonnés, paraissent globalement bien représentatifs de *C. maxima* et peu introgressés. Le pamplemoussier Chandler par exemple présente un unique tronçon introgressé en hétérozygotie *C. reticulata* sur son chromosome 2 (figure IV.4). Cette introgression révélée par les données de séquençage complet n'avait pas pu être mise en évidence lors de l'étude du chromosome 2 par séquençage 454 d'amplicons (Chapitre I). La présence d'introgressions *C. maxima* chez de nombreuses mandarines dans cette zone, a empêché l'attribution d'une origine spécifique des haplotypes étudiés pour un grand nombre de variétés (figure I.7).

### 2.1.3. *C. reticulata*, les mandariniers

Nous avons estimé de manière approximative les proportions relatives des taxons ancestraux dans le génome de 17 variétés traditionnelles de mandariniers à l'aide de marqueurs nucléaires diagnostiques des taxons ancestraux (chapitre II). Sept d'entre elles ont présenté des introgressions significatives de *C. maxima* pouvant atteindre des proportions élevées (supérieures à 20%) comme pour les mandariniers Satsuma, King et Carvalhal. Une estimation beaucoup plus fine a été obtenue pour trois mandariniers (Willowleaf, Cléopâtre et Huanglingmiao) et le clémentinier à partir de données de séquençage complet (chapitre III). Dans cette dernière étude aucun des trois mandariniers considérés comme représentatifs du taxon *C. reticulata* n'a été trouvé pure *C. reticulata*.

- Mandarinier Cléopâtre (*C. reshni*) : il s'agit du mandarinier qui a le génome le moins introgressé avec un seul petit fragment en hétérozygotie *C. maxima* au début du chromosome 3 (figure IV.4),
- Mandarinier Willowleaf (*C. deliciosa*) : nous avons observé des tronçons en hétérozygotie *C. maxima* dans 5 de ses chromosomes (figure IV.4), conformément aux observations de Wu *et al.* (2014),
- Mandarinier Huanglingmiao (*C. reticulata*) : il présente des introgressions *C. maxima* en hétérozygotie dans 3 de ses chromosomes ainsi qu'une petite zone d'environ 1,2 Mb en homozygotie *C. maxima* sur son chromosome 8 (figure IV.4),

### 2.1.4. *C. micrantha*, les papedas

Nous n'avons que très peu de diversité disponible pour ce groupe. Les *C. micrantha* des collections de Corse et d'Espagne se sont révélés identiques. Aucune introgression interspécifique n'a été identifiée chez *C. micrantha*.

## 2.2. Les espèces secondaires

### 2.2.1. *C. aurantium*, les bigaradiers

Conformément aux hypothèses précédentes (Swingle and Reece, 1967; Scora, 1975; Barrett and Rhodes, 1976; Green *et al.*, 1986; Scora, 1988; Yamamoto *et al.*, 1993; Nicolosi *et al.*, 2000; Uzun *et al.*, 2009) nous avons systématiquement observé, quel que soit les outils utilisés (figures I.7, II.10, III.3, III.4 et IV.8A), une contribution équivalente des deux taxons de base *C. maxima* et *C. reticulata* à hauteur de 50% chacun dans le génome des bigaradiers. De plus, les marqueurs cytoplasmiques (figures III.1 et IV.1) indiquent que le parent femelle du bigaradier serait un *C. maxima* différent du parent femelle de l'oranger. Nous pouvons donc conclure que le bigaradier est un hybride direct *C. maxima* x *C. reticulata*.

### 2.2.2. *Citrus sinensis*, les orangers

Nos résultats sont en accord avec les toutes dernières hypothèses (Garcia-Lor *et al.*, 2013b; Wu *et al.*, 2014) sur l'origine de l'oranger. Ils confirment que les deux parents de l'oranger sont d'origine interspécifique (*C. maxima* x *C. reticulata*) puisque nous observons, en plus d'un cytoplasme de type *C. maxima* (différent de celui du bigaradier, figures III.1 et IV.1), un génome à environ 60% *C. reticulata* et 40% *C. maxima* (figures II.10 et III.3) avec une structure complexe comprenant des allèles diagnostiques des deux taxons en hétérozygotie mais également en homozygotie (figures I.10, III.4 et IV.8C).

### 2.2.3. *Citrus paradisi*, les pomelos

Nous avons mis en évidence la structure génomique nucléaire des pomelos dans laquelle la contribution de *C. maxima* est supérieure à celle de *C. reticulata* (environ 60% contre 40%, figures II.10 et III.3). On y trouve des allèles diagnostiques *C. maxima* en homozygotie en plus des allèles diagnostiques en hétérozygotie *C. maxima* et *C. reticulata* (figure III.4). Les marqueurs cytoplasmiques (figures III.1) nous indiquent que le pomelo a une origine maternelle *C. maxima* semblable à celle de l'oranger. Ces résultats sont en accord avec l'hypothèse selon laquelle le pomelo serait un hybride entre l'oranger (*C. sinensis*) et un pamplemoussier (*C. maxima*) (Barrett and Rhodes, 1976; Scora *et al.*, 1982; de Moraes *et al.*, 2007; Li *et al.*, 2010; Ollitrault *et al.*, 2012).

### 2.2.4. Hybrides de mandariniers, tangors, tangelos et orangelo

Les proportions relatives approximatives de la contribution des taxons ancestraux aux génomes de variétés modernes de petits agrumes ont été estimées à l'aide de marqueurs nucléaires diagnostiques des taxons ancestraux (Chapitre II).

- Tangors (*C. reticulata* x *C. sinensis*), tangelos (*C. reticulata* x *C. paradisi*) : tous présentent, dans des proportions différentes, une partie de leur génome en hétérozygotie interspécifique entre *C. reticulata* et *C. maxima* ainsi qu'une grande proportion de marqueurs en homozygotie *C. reticulata* et pour certains quelques marqueurs en homozygotie *C. maxima*.

- Orangelo (*C. sinensis* x *C. paradisi*): la variété étudiée (orangelo Triumph présente une contribution majoritaire de *C. maxima*, avec de nombreux marqueurs en homozygotie, et de *C. reticulata* en hétérozygotie.
- Le clémentinier (*C. clementina*) : les données cytoplasmiques (figures III.1 et IV1) et nucléaires (Chapitres I, II, III et IV) confirment le lien de parenté direct de ce génotype avec le mandarinier Willowleaf (parent femelle) et l'oranger (parent mâle). Grace aux données de séquençage complet du génome de ces trois variétés, associées aux données de séquençage de l'haploïde de clémentinier, nous avons pu schématiser son caryotype phylogénomique complet (figure IV.8b).

### 2.2.5. *C. amblycarpa*, le mandarinier Nasnaran

Nos données nucléaires du chapitre II (figure II.10) confirment les récentes études proposant que ce mandarinier est un hybride *C. micrantha* x *C. reticulata* (Froelicher *et al.*, 2011; Ollitrault *et al.*, 2012).

### 2.3. Cas particulier des limettiers et citronniers

Afin de clarifier plus largement la structuration de la diversité des limettiers et citronniers et d'identifier l'origine des différents groupes, nous avons fait le choix de travailler sur toute la population de citronniers et de limettiers présente dans les collections de Corse (CRB Citrus Inra-Cirad) et de Valence en Espagne (Ivia). Ces limettiers et citronniers sont classés en 2 espèces selon Swingle et Reece (1967) ou 16 espèces selon la classification de Tanaka (1977). Nous avons également intégré dans notre étude des variétés phénotypiquement proches des limettiers et des citronniers mais dont nous ne connaissons pas la classification, en les rangeant dans le groupe des *Citrus sp.* (Annexe IV, tableau S.III.1). Nos données nucléaires et cytoplasmiques confirment les hypothèses précédentes sur l'origine d'une partie des limettiers et citronniers mais permettent également de proposer de nouvelles hypothèses pour d'autres.

Toutes nos données confirment un lien de parenté des limettiers et des citronniers avec *C. medica* comme proposé par Federici *et al.* (1998), Nicolosi *et al.* (2000), Barkley *et al.* (2006), Pessina *et al.* (2011), Ollitrault *et al.* (2012), Garcia-Lor *et al.* (2013b). Nos données cytoplasmiques n'indiquent jamais le cédratier comme un possible parent femelle, cela confirme également les études chloroplastiques (Nicolosi *et al.*, 2000; Bayer *et al.*, 2009) et mitochondriales (Froelicher *et al.*, 2011) antérieures. *C. medica* apparaît bien comme un élément commun à la majorité des variétés classées dans les limettiers et citronniers et en est très certainement le parent mâle direct, en combinaison avec *C. micrantha*, *C. reticulata*, *C. aurantium* et des hybrides (*C. maxima* x *C. reticulata*) pour la majorité des limettiers et citronniers.

Les limettiers et citronniers de notre étude résultent d'au moins 36 évènements de réticulations différents. Les structures génomiques et les hypothèses sur les origines des principaux groupes de limettiers et citronniers sont détaillées ci-après. La discussion est structurée selon les deux taxa de Swingle et Reece, subdivisées selon les taxa de Tanaka.

### 2.3.1. *C. aurantifolia*, selon Swingle et Reece

#### 2.3.1.1. Les limettiers diploïdes à petits fruits verts et à pépins, *C. aurantifolia* selon Tanaka

Tous nos marqueurs nucléaires (figures I.7, II.10, III.3, III.4, et IV.11) et cytoplasmiques (figures III.1 et IV.1) confirment les hypothèses précédentes (Scora, 1975; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012; Garcia-Lor *et al.*, 2013b) affirmant que le limettier Mexicain (*C. aurantifolia* selon Tanaka) est un hybride direct entre le *C. micrantha* et le cédratier (*C. medica*). La lime Mexicaine, selon Webber (1967), serait originaire de l'Insulinde, zone incluant le sud des Philippines d'où serait également originaire *C. micrantha* (Swingle et Reece, 1967). Mais tous les limettiers diploïdes donnant des fruits de petit calibre et à pépins, classés comme *C. aurantifolia* selon Tanaka et Swingle et Reece n'ont pas la même origine. Ainsi, le limettier de Nouvelle Calédonie et la lime Kaghzi, qui possèdent un cytoplasme de type *C. micrantha* (figures III.1 et III.4), présentent des allèles spécifiques *C. medica* et *C. micrantha* en homozygotie impliquant une origine plus complexe ; ce type de structure génomique est compatible avec une origine F2 (autofécondation) d'une lime de type Mexicaine. Le limettier Kirk, qui présente la contribution des quatre taxons ancestraux (figures III.3 et III.4) pourrait être un hybride entre un limettier type Mexicain, dont il possède le même type cytoplasmique (figure III.1) et le pollen d'un hybride *C. reticulata* × *C. maxima*. Le parent mâle de ce limettier n'a pas pu être identifié parmi les variétés que nous avons étudiées (Annexe IV, tableau S.III.6). La diversité observée pour ces limettiers à petits fruits n'est pas surprenante. En effet, les limettiers de type Mexicain sont multipliés par semis dans un grand nombre de régions du monde et même si les pépins de la lime Mexicaine sont polyembryonnés, il arrive qu'un embryon zygotique se développe et donne ainsi naissance à un nouvel hybride qui, en fonction de son intérêt local, pourra perdurer ou pas, en étant multiplié végétativement à son tour. C'est ce qui est certainement arrivé dans le cas de la lime de Nouvelle Calédonie et de la lime Kirk.

#### 2.3.1.2. Les limettiers triploïdes à gros fruits verts et à pépins, *C. arantifolia* selon Tanaka

La présence de pépins dans ces limettiers, de type Ambilobe, Coppenrhad, du Mohtasseb, de Madagascar ou Tanepao, a longtemps fait passer leur triploïdie inaperçue. Ce niveau de ploïdie est en effet généralement associé à des stérilités mâles et femelles et donc à l'aspermie (Ollitrault *et al.*, 2008). Nos résultats nous permettent de proposer une hypothèse sur l'origine de ces limettiers. Ils ont un lien de parenté avec le limettier de type Mexicain dont ils partagent le type cytoplasmique (figure III.1). Au niveau nucléaire, nous observons les contributions de *C. micrantha* et de *C. medica* avec double dose d'allèles spécifiques *C. medica* (figure III.7). Ces limettiers pourraient donc être issus d'un gamète femelle diploïde de *C. aurantifolia* fécondé par un pollen haploïde *C. medica*.

#### 2.3.1.3. Les limettiers triploïdes à gros fruits verts et sans pépins, *C. latifolia* selon Tanaka

La triploïdie de ces limettiers est connue depuis plus longtemps que pour le groupe précédent (Bacchi, 1940). L'origine des limettiers triploïdes, à gros fruits sans pépins, de type Tahiti, Bears, El Kseur, IAC-5, ou de Perse, est mal connue (Webber *et al.*, 1967). Tout



ce que l'on sait c'est que le nom de Tahiti viendrait du fait que ce fruit fut introduit en Californie depuis Tahiti entre 1850 et 1880 et que la lime de Perse tirerait son nom du fait qu'elle serait arrivée en Méditerranée par la Perse, même si Chapot, en 1965 n'en retrouva aucune lors de sa prospection en Iran. Ces limettiers étaient supposés avoir une origine hybride avec un lien de parenté avec le cédratier au même titre que les autres limettiers et citronniers (Mabberley, 2004; Bayer *et al.*, 2009). Reece et Childs (1962) pensaient même avoir identifié un des deux parents comme étant le limettier de type Mexicain mais hésitaient entre le citronnier (*C. limon*) et le cédratier (*C. medica*) pour l'autre parent. Nos résultats confirment les résultats de Bayer *et al.* (2009) et Froelicher *et al.* (2011) démontrant que ces limettiers ont un type cytoplasmique *C. maxima*, du même type que les bigaradiers et les citronniers de type Lisbonne (figure III.1). Notre étude du génome nucléaire montre que les quatre taxons de base contribuent à la structure génétique de ces limettiers avec uniquement les allèles spécifiques *C. medica* et *C. micrantha* en double ou simple dose, alors que les allèles *C. maxima* et *C. reticulata* sont soit en simple dose soit absents (figure III.7 et chapitre IV). Ces observations nous permettent de proposer l'hypothèse suivante : les limettiers triploïdes de type Tahiti seraient issus de la fécondation d'un gamète femelle haploïde de citronnier, par un gamète mâle diploïde d'un limettier de type Mexicain. 99% des 123 marqueurs étudiés sont en adéquation avec ce modèle (Annexe IV, tableau S.III.6). En revanche, on ne sait pas s'il s'agit d'un gamète non réduit d'un limettier diploïde, ou d'un gamète mâle diploïde produit par un limettier tétraploïde, tel que le limettier Giant Key par exemple. L'étude du génome complet du limettier Bears nous a permis de schématiser l'ensemble de son caryotype phylogénomique (figure IV.13) en tenant compte des dosages alléliques des quatre taxons de base *C. micrantha*, *C. maxima*, *C. medica* et *C. reticulata*.

#### 2.3.1.4. Le limettier tétraploïde à gros fruits verts, *C. aurantifolia* selon Tanaka

La tétraploïdie du limettier Giant Key a été confirmée par cytométrie en flux (Chapitre III). L'étude des marqueurs SSR et SNP (Chapitre III) n'a pas mis en évidence de tri allélisme et les profils alléliques de la Giant-Key sont systématiquement identiques à ceux de la lime Mexicaine. Le génome de ce limettier correspond donc au doublement de celui d'un limettier Mexicain. Le fait qu'un limettier de type Mexicain soit à la fois :

- ✓ à l'origine du gamète mâle diploïde des limettiers triploïdes de type Tahiti,
- ✓ à l'origine du gamète femelle diploïde des limettiers triploïdes de type Tanepao,
- ✓ et à l'origine du limettier tétraploïde,

ne nous permet pas de savoir si les gamètes qui ont donné les limettiers triploïdes sont issus de gamètes non réduits d'un individu diploïde ou issus d'un individu tétraploïde similaire à ce limettier Giant Key.

#### 2.3.1.5. Les limettiers à fruits aplatis, *C. Excelsa*, *C. Webberii* et *C. aurata* selon Tanaka,

Ces limettiers regroupent des variétés à fruits particulièrement reconnaissables par leur forme et leur aréole qui, chez certaines variétés, ressemble à des traces d'incisives laissées dans une pomme (d'où le nom de la pomme d'Adam [*C. aurata*, classé chez les *C. limon* par Swingle et Reece], mais également les limes Excelsa et Nestour [*C. excelsa*] et la lime Kalpi [*C. webberii*] classées parmi les *C. aurantifolia* par Swingle and Reece). Nos résultats montrent que ces variétés classées dans 3 espèces différentes selon Tanaka et

deux selon Swingle et Reece, ont une origine commune voir identique. Toutes ces limes ont le même type cytoplasmique *C. micrantha* (figure III.1) et une contribution génétique partagée entre *C. micrantha* et *C. medica* (figures III.3 et III.4). Nous faisons l'hypothèse que ces variétés sont des hybrides directs entre un papeda apparenté à *C. micrantha* et *C. medica*. Du fait du niveau d'adéquation relativement bas obtenu pour le modèle *C. micrantha* x *C. medica* sur l'ensemble des marqueurs étudiés (88 %, Chapitre III, Annexe IV, tableau S.III.6), il est clair que les *C. micrantha* dont nous disposons ne peuvent pas être les parents directs de ces limes. En l'absence de connaissance sur la diversité intraspécifique de *C. micrantha*, il n'est pas possible de conclure entre l'intervention d'un autre génotype de *C. micrantha* ou celle d'une espèce apparentée.

#### 2.3.1.6. Alemow, *C. macrophylla* selon Tanaka

Le lien de parenté entre Alemow et les papedas a déjà été évoqué (Barkley *et al.*, 2006). Les résultats obtenus à l'aide des 5 indels mitochondriaux et des 3 SSR chloroplastiques (figure III.1) indiquent qu'Alemow possède bien un cytoplasme de type *C. micrantha* mais ceux obtenus avec les 718 SNP chloroplastiques (Chapitre IV), démontrent qu'il s'agit d'un papeda proche du *C. micrantha* étudié mais pas strictement identique (figure IV.1). Nos résultats nucléaires mettent en évidence un génome composé pratiquement de 50% de *C. micrantha* et de 50% de *C. medica* (que ce soit sur le chromosome 2, figure I.7, ou sur l'ensemble du génome figures III.3, III.4 et chapitre IV) confirmant ainsi les précédentes hypothèses le considérant comme un hybride direct *C. micrantha* (ou apparenté) x *C. medica* (Barrett and Rhodes, 1976; Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012). L'hypothèse selon laquelle *Citrus celebica* Koord (un autre papeda, différent de *C. micrantha* et non présent dans nos collections) serait à l'origine de cette variété (Webber *et al.*, 1967) est donc plausible mais reste à confirmer.

#### 2.3.1.7. Les limettiers à fruits doux, *C. limettioïdes* selon Tanaka

Ces limettiers à fruits doux, dont la lime douce de Palestine, du Brésil ou encore la lime Bisri, ont longtemps été considérés comme des hybrides de limettier Mexicain (*C. aurantifolia*) soit par *C. limetta* (Webber, 1943) soit par *C. medica* (Webber, 1943 ; Carvalho *et al.*, 2005). En 2000, Nicolosi *et al.* proposent la combinaison orange (*C. sinensis*) x cédratier. Nos résultats cytoplasmiques (Chapitre III) écartent totalement la possibilité que le limettier Mexicain soit le parent femelle puisque nous identifions un type cytoplasmique *C. maxima* similaire à celui de l'oranger et différent du bigaradier (figure III.1). De plus, nous n'avons retrouvé aucune trace de *C. micrantha* dans l'ensemble de l'étude génétique nucléaire de ces génotypes. Ils ont des profils génétiques impliquant trois taxons de base, *C. medica*, à hauteur de 50% (figures II.10, III.3 et III.4) ainsi que *C. maxima* et *C. reticulata*. Ces limettiers sont donc probablement issus de croisements (*C. maxima* x *C. reticulata*) x *C. medica*. Nous n'avons pas pu identifier le parent femelle car aucun des hybrides (*C. maxima* x *C. reticulata*) de notre étude en combinaison avec *C. medica* ne permet de générer le profil génétique de ces variétés (Annexe IV, tableau S.III.6). L'oranger est également écarté du rôle de parent femelle (*C. maxima* x *C. reticulata*) puisqu'il ne donne que 85 % d'adéquation contre 98 % pour un modèle (*C. maxima* x *C. reticulata*) x *C. medica* (Annexe IV, tableau S.III.6).

### 2.3.1.8. La bergamote, l'arbre à parfum, *C. bergamia* selon Tanaka

La bergamote partage le même type cytoplasmique que la bigarade et les citrons jaunes (figure III.1) conformément aux résultats de Gulsen and Roose, (2001b). Sa structure nucléaire est un mélange à hauteur de 30 % de *C. medica*, environ 40 % de *C. maxima* et 30 % de *C. reticulata* (marqueurs SNP et Indels ; chapitre III). De plus, elle apparaît homozygote pour certains allèles spécifiques de ces trois taxons de base, suggérant que la bergamote ne serait pas un hybride interspécifique direct. En testant différents modèles d'origine relatifs à la diversité étudiée, l'hypothèse la plus probable est celle d'un croisement entre un citronnier jaune de type Lisbonne et un bigaradier (*C. limon* × *C. aurantium*) (Annexe IV, tableau S.III.6). Cette hypothèse est en accord avec Gallesio (1811), mais contredit de nombreuses autres études (Herrero *et al.*, 1996; Federici *et al.*, 2000; Nicolosi *et al.*, 2000; Li *et al.*, 2010).

### 2.3.2. *C. limon* selon Swingle et Reece

#### 2.3.2.1. Les citronniers à fruits jaunes, *C. limon* selon Tanaka

Nos résultats montrent que 27 citronniers (également appelés citrons jaunes et incluant les variétés commerciales classiques Lisbonne et Eureka et d'autres moins connues mais d'intérêt agronomique, comme le citronnier Adamo considéré comme tolérant au Mal Secco (Recupero *et al.*, 2010), des collections de Corse et d'Espagne, présentent un profil génétique identique. Selon nos différentes études cytoplasmiques (Chapitre III et IV) ils ont un cytoplasme de type *C. maxima*, identique au bigaradier. Nous avons également pu déterminer la structure génétique de ces citronniers à l'aide d'indels et de SNP (figures II.10, III.3 et III.4) dont le génome est à 50% composé de *C. medica*, à 30% de *C. reticulata* et à 20% *C. maxima*. Ces résultats sont en accord avec l'hypothèse de Nicolosi *et al.* (2000) qui propose une origine par hybridation *C. aurantium* × *C. medica*, confirmée par d'autres études moléculaires (Gulsen and Roose, 2001a; Ollitrault *et al.*, 2012; Garcia-Lor *et al.*, 2013b). Ce modèle est cohérent pour 99,2 % des 123 marqueurs étudiés dans le chapitre III (Annexe IV, tableau S.III.6).

L'étude du génome complet du bigaradier et du citronnier nous a permis de vérifier la cohérence du modèle proposé (bigaradier × cédratier), à l'origine des citronniers en recherchant chez le citronnier les allèles rares du bigaradier, non partagés avec les taxons de base. Ils sont partagés à 46 % par le citronnier, essentiellement dans les zones hétérozygotes *C. maxima* / *C. medica* (Chapitre IV).

Au-delà de confirmer des hypothèses sur l'origine de ces citronniers, hybrides entre le bigaradier (*C. maxima* × *C. reticulata*) et le cédratier (*C. medica*), nous apportons une nouvelle information qui est celle de la structure phylogénomique complète du citronnier Eureka par la schématisation de son caryotype phylogénomique (figure IV.10) sur lequel toutes les zones hétérozygotes *C. reticulata* / *C. medica* et *C. maxima* / *C. medica* sont clairement identifiées et localisées ainsi que les 15 points de recombinaison du gamète bigaradier sur les 9 chromosomes (respectivement 1, 4, 3, 1, 1, 3, 1,1 et 0 sur les chromosomes de 1 à 9).

### 2.3.2.1. Le limettes à fruits doux ou acides et à mamelon, *C. limetta* selon Tanaka

Ces limettiers sont classés chez les *C. limon* par Swingle et Reece. Ils ont une forme très particulière qui a donné son nom à une variété (limette à Mamelon, figure 23 de l'Introduction), et sont très recherchés en Afrique du Nord pour leur parfum (en Tunisie une variété est d'ailleurs appelée bergamote de Tunis et, au Maroc, une autre a pris le nom de la ville où on l'apprécie particulièrement, notamment pour parfumer les tajines : la limonette de Marrakech, figure 38 de l'introduction). D'après Webber (1967) ces limettes seraient nées en Méditerranée. Il précise également qu'il s'agirait d'un groupe de mutants doux d'un type de limettier acide. Les indels mitochondriaux et les SNP chloroplastiques (figure III.1) nous indiquent que ces limes ont un type cytoplasmique *C. maxima* équivalent aux bigaradiers et aux citronniers de type Lisbonne. Les marqueurs nucléaires font apparaître des origines impliquant trois taxons de base : *C. maxima*, *C. reticulata* et *C. medica* (figures III.3 et III.4). Nous avons testé l'hypothèse (*C. maxima* × *C. reticulata*) × *C. medica*, en considérant le bigaradier (*C. aurantium*) comme hybride interspécifique direct *C. maxima* *C. reticulata*. Cette combinaison est cohérente pour 98,3 % des 123 marqueurs nucléaires étudiés au chapitre III (Annexe IV, tableau S.III.6).

### 2.3.1.6. Les citronniers à fruits oranges et à peau rugueuse, *C. limonia*, *C. karna* et *C. jambhiri* selon Tanaka

Ce groupe comprend des variétés appelées à la fois limettiers (c'est le cas pour la lime Rangpur, figure 21 de l'introduction, ou le limettier d'Inde, d'Arabie Saoudite et d'Iran [*C. limonia*], ou limettier Khatta [*C. karna*]) et citronniers (c'est le cas pour le Volkameriana [*C. limonia*] ou le Rough lemon [*C. jambhiri*], figure 22 de l'introduction), le tout classé dans 3 espèces différentes selon Tanaka. Webber (1967) décrit les *C. limonia*, comme d'anciennes variétés indiennes d'origine inconnue. Il évoque en les décrivant des caractères proches du bigaradier et du cédratier. La plupart des auteurs considère la mandarine comme un des parents possibles de la lime Rangpur croisée par, soit un limettier (Webber, 1943; Tatum *et al.*, 1974; Barkley *et al.*, 2006), soit un bigaradier, (Barrett and Rhodes, 1976), soit le Rough lemon (Handa and Oogaki, 1985), soit un cédratier (Federici *et al.*, 1998; Gulsen and Roose, 2001a; Li *et al.*, 2010). De récentes études (Federici *et al.*, 1998; Nicolosi *et al.*, 2000; Gulsen and Roose, 2001a; Barkley *et al.*, 2006; Ollitrault *et al.*, 2012) ont confirmé l'hypothèse de Scora (1975) qui proposait que le Rough lemon soit un hybride naturel entre un mandarinier et un cédratier. L'origine du Volkameriana a suscité plus de débat, Barrett et Rhodes (1976) pensaient qu'il s'agissait d'un hybride entre un mandarinier et un bigaradier. Nicolosi *et al.* (2000) considéraient que le bigaradier et le cédratier en étaient les parents. Enfin, Carvalho *et al.* (2005), émirent l'hypothèse, à partir d'une étude cytogénétique, que le Volkameriana était un hybride mandarinier x cédratier comme le Rough lemon ou la lime Rangpur. D'après nos résultats (Indels mitochondriaux et SNP chloroplastiques, Chapitre III) ces variétés ont un cytoplasme *C. reticulata* de type acide, ce qui est cohérent avec les résultats de Carvalho *et al.* (2005) qui avaient observé les caryotypes de ces variétés et mis en évidence leur parfaite correspondance entre les jeux de chromosomes complémentaires haploïdes cédrat et mandarine Cléopâtre. Nos résultats nucléaires (Chapitre III) mettent en évidence une origine hybride n'impliquant que deux taxons ancestraux : *C. reticulata* et *C. medica*. D'après la diversité de *C. reticulata* dont nous disposons, la mandarine la plus

proche du *C. reticulata* à l'origine de ce groupe serait la mandarine Sun Chu Cha (Annexe IV, tableau S.III.6).

### 2.3.2.2. Les citronniers à fruits oranges et à peau lisse, *C. meyeri* selon Tanaka

Le citronnier Meyer, importé aux États-Unis en 1908 depuis Pékin, par Frank N. Meyer de l'*U.S. Department of Agriculture* (Webber *et al.*, 1967), a longtemps été considéré comme un hybride naturel entre une orange et un citron (Scora, 1975; Gulsen and Roose, 2001a). Nos résultats (SNP, figure II.10 et indels et SNP, figures III.3 et III.4) mettent en évidence une structure nucléaire impliquant trois taxons ancestraux (*C. reticulata* × *C. maxima*) × *C. medica*. Le citronnier Meyer partage le cytoplasme *C. maxima* de l'oranger (figure III.1) mais, d'après nos données nucléaires, ce dernier ne serait pas le parent de ce citronnier. La diversité étudiée ne nous a pas permis d'identifier le parent femelle (*C. reticulata* × *C. maxima*) à l'origine de ce type de citron (Annexe IV, tableau S.III.6).

## 2.4. Bilan sur les classifications botaniques

Nos résultats indiquent que la classification de Swingle et Reece (1967) regroupant tous les limettiers et citronniers en deux espèces (*C. limon* et *C. aurantifolia*) ne correspond pas à la réalité des différentes origines de ces espèces secondaires. La classification de Tanaka (1961, 1969), quant à elle, présente un trop grand nombre de subdivisions correspondant pour l'essentiel à des groupes variétaux d'origine clonale (en raison de l'apomixie facultative) issus de différents événements de réticulation. Comme le dit Mabberley en 2004, la subdivision en un nombre réduit d'espèces différentes est plus juste du point de vue de l'évolution des agrumes mais le regroupement proposé par Swingle et Reece n'est pas cohérent avec les réelles origines hybrides de la plupart de ces variétés. C'est pour cela que Mabberley (1997) a essayé de simplifier encore cette classification en ne proposant que 3 principales espèces pour les agrumes commerciaux du genre *Citrus* : *C. medica*, *C. reticulata* et *C. maxima*, et des noms d'hybrides du type *Citrus* × *aurantium* pour les orangers, les bigaradiers et les pomelos par exemple, *Citrus* × *taitensis* pour le Rough lemon ou *Citrus* × *limon* pour les citrons considérés comme des rétrocroisements d'un limettier par un cédratier. Cette classification, en cherchant à se rapprocher de l'histoire évolutive des agrumes du genre *Citrus*, propose une démarche intéressante mais n'est pas encore acceptable car elle est basée sur des hypothèses de l'origine des variétés aujourd'hui contredites par les données de génomiques ou non vérifiées. Les résultats de cette thèse permettent d'éclairer plus précisément la phylogénomique des agrumes du genre *Citrus* et contribueront ainsi à établir une classification plus proche de la véritable histoire évolutive de ces variétés.

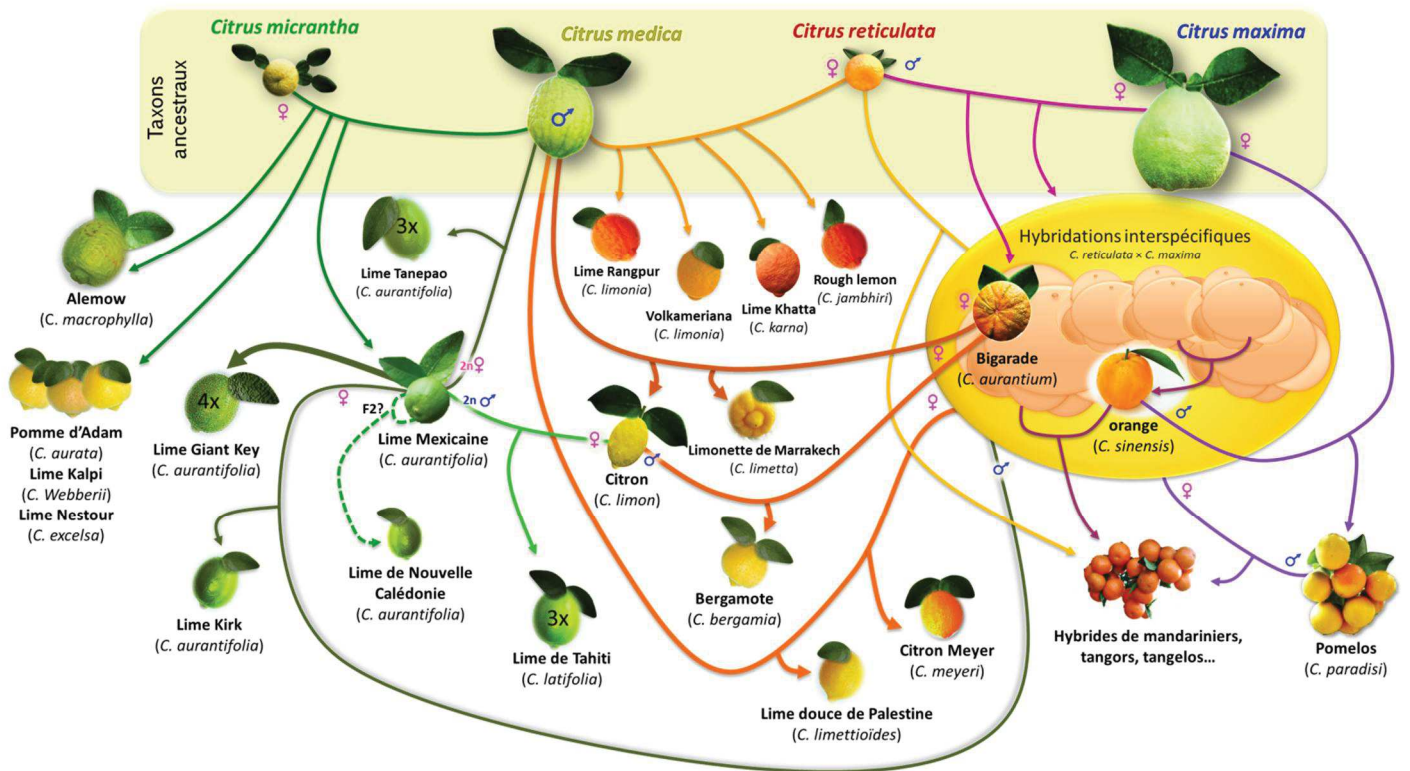


Figure D1 : Hypothèses sur l'origine phylogénétique des principales espèces secondaires de *Citrus* (Conception F. Curk et P. Ollitrault ; Photos et photomontage © F. Curk – Inra)

### 3. Perspectives pour la gestion des ressources génétiques et l'amélioration des agrumes

#### 3.1. Caractérisation, valorisation et stratégie de conservation de la diversité génétique

L'un des principaux objectifs des centres de ressources biologiques végétaux, comme le CRB Citrus Inra-Cirad de San Giuliano, est la conservation de la diversité génétique des plantes cultivées. Même si en général les CRB se limitent à un groupe d'espèces, la collectionniste peut ne pas avoir de limite. En effet, en fonction des stratégies de conservation, de la diversité disponible et des moyens mis en œuvre on peut facilement tomber dans la philosophie des zoos du XIXe siècle qui cherchaient à avoir le maximum d'espèces animales différentes (Mullan and Marvin, 1987). En général, les moyens (personnel, coûts de maintenance...) et les surfaces allouées aux CRB sont de véritables facteurs limitants à l'expansion à outrance. C'est pourquoi la notion de « collection-noyau » ou *core collection* est de plus en plus évoquée dans les stratégies de conservation rationalisée de la diversité. L'objectif est alors d'avoir la plus grande diversité possible à partir d'un nombre limité d'introductions (ou accessions). Des études ont montré que 85 % de la diversité génétique d'une collection de riz, pourrait être conservée dans 18 % du nombre total d'accessions (Pessoa-Filho *et al.*, 2010). Mais ce résultat ne peut être atteint que grâce à une connaissance approfondie de l'organisation de la diversité des complexes d'espèces. Ce travail de thèse a permis d'identifier de très nombreux SNP diagnostiques des quatre espèces ancestrales et de développer des marqueurs de type Kaspar efficaces. Ces marqueurs contribueront, en association avec des SSR nucléaires et les marqueurs cytoplasmiques, à une caractérisation exhaustive de la collection du CRB de Corse. Leur aptitude à révéler les structures interspécifiques sera un élément important dans la définition d'une éventuelle *core collection*.

Mais la *core collection* ne doit pas être le seul moteur de la gestion de la diversité génétique. D'autres notions d'intérêts scientifiques, économiques, culturels et sociaux doivent être prises en compte au moment de faire des choix sur la diversité génétique à conserver. Ainsi, le CRB Citrus Inra-Cirad de San Giuliano, très riche en mandariniers, est relativement pauvre en accessions de cédratiers. Pourtant, le cédratier est un taxon particulièrement intéressant. En effet, comme nous l'avons vu, son génome est impliqué dans pratiquement l'ensemble des génotypes du groupe des limettiers et citronniers. De plus, il présente plusieurs caractéristiques singulières pouvant soit être des thèmes d'études, soit servir de référence comportementale par rapport aux interactions génotype / environnement :

- ✓ il est particulièrement sensible au froid et présente une déficience en mécanisme de régulation du stress photo-oxydatif (Santini *et al.*, 2012 ; 2013),
- ✓ il est très sensible aux stress salin (Hussain *et al.*, 2008 ; 2012),
- ✓ il présente des compositions particulières en huiles essentielles recherchées en cosmétique (Venturini *et al.*, 2010; Luro *et al.*, 2012),
- ✓ il est utilisé comme indicateur biologique de la présence de virus ou amplificateur de viroïdes pour faciliter leur détection (Roistacher *et al.*, 1977),
- ✓ il est sensible au mal secco, maladie fongique qui vient tout juste d'être signalée en Corse (Fredon Corse, 2014) et qui s'attaque également aux citronniers et aux limettiers (Bové and Duran-Vila, 2013),

- ✓ il s'agit également de l'agrume cultivé ayant le plus grand génome ; environ 10% plus grand que celui du clémentinier, référence génomique des agrumes (Ollitrault *et al.* 1992). Nos études sur les données de re-séquençage complet du génome ne traitent d'ailleurs pas de cette partie du génome des cédratiers encore à découvrir,
- ✓ enfin, il présente des attraits autres que scientifiques : il existe un fort attachement historique du cédrat à la Corse, le symbolisme religieux de ce fruit est important dans la religion juive et il est considéré comme le premier agrume introduit en Méditerranée.

Ce taxon est donc utile pour différents programmes de recherche et pourrait être un modèle d'études. Il serait intéressant de se pencher sur son évolution hors de sa zone d'origine et sur les facteurs anthropiques comme éléments de structuration de sa diversité ; comme cela a déjà été partiellement abordé par Nicolosi *et al.* (2005) et Luro *et al.* (2012)

Au-delà de la conservation de ressources phytogénétiques, les collections ex-situ sont des outils privilégiés pour analyser le support moléculaire de la diversité phénotypique par génétique d'association. Garcia lor *et al.* (2012) ont montré, au niveau du genre *Citrus* dans son ensemble, que les déséquilibres gamétiques généralisés, dus à l'histoire évolutive des agrumes cultivés, ne permettraient pas de telles approches. Ils suggéraient en revanche que le groupe des petits agrumes (type mandariniers) pourrait être d'avantage adapté à ce type d'étude. Nos résultats (chapitre III), montrant un continuum des contributions relatives de *C. reticulata* et *C. maxima* aux variétés de ce groupe (mandariniers, tangors, tangelos, orangelos) confortent cette hypothèse. Compte tenu de la très forte différenciation phénotypique entre *C. reticulata* et *C. maxima* pour la plupart des caractères clefs de qualité des fruits, ce groupe variétal pourrait permettre d'aborder des études de génétique d'association basées sur les structures phylogénomiques. Enfin, dans ce cadre d'analyse des déterminants de la variabilité, des études conduites sur d'autres espèces végétales ont montré que la confrontation interspécifique au sein d'un même génome constituait une forme de choc de nature à induire des reformatages épigénétiques importants et par là même conduire à des hérédités non mendéliennes. Les ressources génétiques naturelles et les différents hybrides générés par l'Inra et le Cirad constituent un matériel unique pour analyser l'impact des structures interspécifiques des agrumes sur le formatage épigénétique, la régulation transcriptomique et *in fine* l'élaboration du phénotype. A ce niveau, nos marqueurs Kaspar diagnostiques d'espèces, développés dans les exons de nombreux gènes, pourront contribuer à l'analyse de l'expression allèle spécifique de ces gènes dans des contextes interspécifiques.

Le CRB Citrus Inra-Cirad de San Giuliano est une extraordinaire chance pour l'équipe de généticiens que je viens de rejoindre. Mon expérience en Espagne m'a appris que la plupart des laboratoires au travers le monde, à l'aide de moyens financiers suffisants, peuvent réaliser l'essentiel des études génétiques et moléculaires sur de l'ADN de n'importe quel type de plantes. Mais en ce qui concerne les agrumes, très peu de ces laboratoires ont accès à des plantes adultes aux champs pouvant être étudiées phénotypiquement ou être utilisées comme génitrices pour la constitution de populations recombinantes, indispensables aux études de QTL par exemple. Les défis futurs ne sont plus liés à l'obtention de données de génotypage haut débit, mais bel et bien à l'accès à la diversité et à la plante *in vivo* pour pouvoir travailler sur l'expression des gènes et la physiologie de la plante dans des approches de biologie intégrative.



### 3.2. Connaissance des structures génomiques et stratégies de création variétale

L'origine interspécifique de groupes horticoles comme celui des orangers, des pomelos ou des citronniers ne permet pas d'en aborder l'amélioration par hybridation sexuée intragroupe. L'idéotype phénotypique définissant ces espèces repose en effet sur des structures génomiques qui sont brisées par la recombinaison sexuée. Une approche pour diversifier ces groupes horticoles pourrait être de reconstruire ces idéotypes à partir du génome des espèces ancestrales. Cette diversification est d'autant plus importante que les variétés de ces groupes horticoles cultivées à grande échelle au plan international ont une diversité génétique très étroite (diversification par mutation ou variation épigénétique ponctuelle dans le cadre d'une propagation clonale) et sont confrontées à des contraintes biotiques et abiotiques croissantes. Nos résultats permettent de comprendre l'évolution des principales espèces secondaires. Ils sont ainsi précieux pour développer des stratégies d'amélioration, par reconstruction de ces espèces.

Nous avons vu qu'un grand nombre de ces espèces secondaires apomictiques résultait très probablement d'hybridation directe entre deux taxons ancestraux comme le bigaradier (*C. maxima* x *C. reticulata*), le limettier Mexicain (*C. micrantha* x *C. medica*), ou les limettiers Rangpur et Khatta, et les citronniers Rough lemon et Volkameriana (*C. reticulata* x *C. medica*). Il est donc envisageable de recréer relativement simplement ces idéotypes par hybridation en sélectionnant, au sein du germplasm des taxons ancestraux correspondant, des géniteurs apportant les caractères d'adaptation ou de résistance souhaités.

Par ailleurs, pour des origines plus complexes impliquant trois ou quatre taxons ancestraux et des recombinaisons interspécifiques, les méthodes développées pour analyser les données pangénomiques et nos résultats fournissent des informations sur les principaux points de recombinaisons survenues après les premiers événements de réticulations entre les taxons de base. Il est donc envisageable d'obtenir des génotypes présentant des structures phylogénomiques proches par sélection génomique. Dans le cas des citronniers de type Lisbonne ou Eureka issus d'hybridation entre *C. aurantium* et *C. medica* (Chapitre III et IV), le schéma est relativement simple dans la mesure où *C. aurantium* est lui-même issu d'hybridation directe entre *C. maxima* et *C. reticulata* (Whu et al., 2014, et nos résultats chapitre IV). Pour l'oranger dont les deux parents sont eux-mêmes d'origine interspécifique, la reconstruction apparaît plus complexe mais reste envisageable. L'identification de l'origine des limes triploïdes de type Tahiti ouvre également la voie à la diversification de ces variétés.

En vue d'optimiser les schémas d'amélioration des agrumes, que ce soit au niveau diploïde ou polyploïde, il est par ailleurs essentiel d'analyser l'impact des structures interspécifiques, révélées par la thèse, sur le fonctionnement méiotique des géniteurs diploïdes et tétraploïdes. Cette composante dynamique de l'interspécificité est aujourd'hui traitée en collaboration entre l'UMR Agap et l'Ilvia pour les diploïdes doublés des principales espèces secondaires d'agrumes (Thèse d'Houssem Rouis). Les travaux analysent en particulier les tendances disomiques ou tétrasomiques des ségrégations des différents chromosomes en fonction de leurs structures interspécifiques. Au niveau triploïde, l'étude des ségrégations par marquage moléculaire permettra de comprendre les mécanismes

méiotiques sous-tendant la fertilité des limettiers de type Tanepao. L'analyse de l'impact de l'interspécificité sur la fertilité et la recombinaison au niveau diploïde constitue l'un des axes de mes futures activités et diverses populations recombinantes ont déjà été générées à cet effet. Pour l'ensemble de ces études, les marqueurs diagnostiques d'espèces développés durant la thèse seront très utiles et sont déjà utilisés en routine dans le cadre du partenariat Agap-IVIA.

C'est dans la perspective de valoriser les résultats de cette thèse à l'aide de l'extraordinaire diversité du CRB agrumes et des populations recombinantes créées par l'Inra et le Cirad de San Giuliano que je rejoins avec motivation l'équipe de l'UMR Agap Corse.

---

## Conclusion

---

Les agrumes cultivés sont issus d'une évolution réticulée impliquant quatre taxons ancestraux (*C. reticulata*, *C. maxima*, *C. medica* et *C. micrantha*). Leur taxonomie reste controversée du fait de leurs origines interspécifiques et de l'apomixie qui a fixé de nombreuses structures interspécifiques considérées comme de nouvelles espèces par certains taxinomistes. Une connaissance approfondie de l'origine des différentes formes cultivées et de leur structure phylogénomique est ainsi indispensable à la bonne gestion des ressources phytogénétiques et à l'optimisation des programmes d'amélioration génétique.

C'est dans cet objectif que nous avons étudié l'organisation du complexe d'espèce, avec une attention particulière au groupe des citronniers et limettiers, qui fait l'objet de moins de recherches au plan international, et que nous avons tenté de décrypter les structures interspécifiques des génomes des espèces secondaires et des variétés modernes.

Pour cela, nous avons utilisé diverses approches d'analyse des génomes. L'haplotypage par séquençage parallèle d'amplicons avec la méthodologie 454 a permis de révéler la structure phylogénomique du chromosome 2 pour 50 variétés ; la lourdeur du traitement de données n'a pas permis d'étendre l'analyse à l'ensemble des chromosomes. Afin d'analyser une large diversité au sein des collections, des SNP diagnostiques des quatre taxons ancestraux ont été recherchés à partir de données de séquençage d'amplicons (454) couvrant les neuf chromosomes du set haploïde des agrumes. Soixante-treize marqueurs SNP Kaspar ont été développés avec succès. En combinaison avec 32 marqueurs diagnostiques SNP précédemment publiés, ils ont permis de révéler la constitution interspécifique de nombreux porte-greffe et variétés actuels couvrant l'ensemble de la diversité des agrumes cultivés. Ces mêmes marqueurs couplés avec des SSR et indels nucléaires et cytoplasmiques ont permis d'analyser l'organisation et l'origine de la diversité du groupe des limettiers et citronniers dans le cadre d'un génotypage exhaustif des collections Inra-Cirad et de l'lvia. Enfin, afin de décrypter finement la structure phylogénomique de quelques variétés, une étude pangénomique des SNP a été conduite à partir de données issues de re-séquençage complet des génomes selon la méthodologie Illumina.

Ces différentes approches d'analyse des génomes ont conduit à des conclusions convergentes. Nos résultats confirment les hypothèses antérieures concernant la séquence évolutive à l'origine des bigaradiers (*C. aurantium*), des orangers (*C. sinensis*) et des pomelos (*C. paradisi*) à partir des pools géniques de *C. maxima* et *C. reticulata*. Ils mettent en évidence de fréquentes introgressions de *C. maxima* dans le génome des variétés de mandariniers, généralement considérées comme représentatives de *C. reticulata* et ont permis de quantifier les proportions relatives de ces deux taxons ancestraux dans le génome de nombreuses variétés de petits agrumes (hybrides de mandariniers, tangors, tangelos et orangelos).

Nos travaux sur les limettiers et citronniers confirment que *C. medica* est de façon très générale le parent mâle de ces types variétaux au niveau diploïde. Deux groupes de citronniers sont clairement différenciés, ceux issus d'hybridations directes *C. reticulata* x *C. medica* correspondant à trois espèces de Tanaka (*C. limonia*, *C. karna* et *C. jambhiri*) et ceux issus de croisements entre hybrides (*C. maxima* x *C. reticulata*) et *C. medica*. Le bigaradier serait ainsi le parent femelle à l'origine des citronniers à fruits jaunes traditionnels (*C. limon* de type Eureka, Lisbonne, etc.) et des limettes (*C. limetta* selon Tanaka). *C*

*limmettioides* et *C. meyeri* sont probablement également issus d'une combinaison (*C. maxima* x *C. reticulata*) x *C. medica* mais le parent femelle reste inconnu. Les limettiers type Mexicain (*C. aurantifolia*) seraient issus d'une hybridation directe *C. micrantha* x *C. medica*. Une origine similaire est proposée pour *C. weberri*, *C. aurata*, *C. exelsa* et *C. macrophylla*. Certaines variétés ont également présenté des profils n'impliquant que *C. medica* et *C. maxima* et d'autres ont des structures complexes impliquant trois ou quatre taxons ancestraux avec des combinaisons d'allèles ancestraux en homozygotie/hétérozygotie dont l'origine n'a pas pu être définie. Au total, trente-six évènements de réticulation ont été identifiés comme étant à l'origine d'une part importante de la diversité du germplasm limettier et citronnier diploïde analysé. Des variations ponctuelles sont également un moteur de diversification au sein des groupes à propagation clonale comme celui des citronniers jaunes traditionnels du Bassin Méditerranéen. Deux origines majeures ont été déterminées pour les limettiers triploïdes à gros fruits. Les types Tahiti résultent très probablement de la fécondation d'un ovule de citronnier par un gamète diploïde de limettier type Mexicain alors que l'autre grand type serait issu d'un rétrocroisement entre *C. aurantifolia* (gamète femelle diploïde) et *C. medica*. Nos résultats sur l'origine phylogénétique des principaux sous-groupes de limettiers et citronniers pourront contribuer utilement à la révision de la taxonomie des agrumes cultivés.

Nous avons développé un processus d'analyse des données SNP pangénomiques issues de re-séquençage complet Illumina qui a permis de proposer pour la première fois les caryotypes phylogénétiques de six limettiers et citronniers diploïdes et d'un limettier triploïde.

Les résultats de cette thèse laissent entrevoir des perspectives intéressantes en termes de gestion et de caractérisation des ressources génétiques agrumes et permettent d'envisager de nouvelles stratégies d'amélioration variétale. Les marqueurs diagnostiques des taxons ancestraux développés trouveront de multiples applications dans les différentes études de génétique conduites au niveau diploïde et polyploïde.

---

## Références bibliographiques

---

454 Life Sciences Corp. 2009. Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Extended MID Set. Technical Bulletin Genome Sequencer FLX System 005-2009:.

Abkenar, A.A. Isshiki, S. and Tashiro, Y. 2004. Phylogenetic relationships in the "true citrus fruit trees" revealed by PCR-RFLP analysis of cpDNA. *Scientia Horticulturae* 102:233-242.

Ablett, G. Hill, H. and Henry, R. 2006. Sequence Polymorphism Discovery in Wheat Microsatellite Flanking Regions using Pyrophosphate Sequencing. *Molecular Breeding* 17:3:281-289.

Achachi, A. Curk, F. Gaboun, F. El Fahime, E. Soulaymani, A. and Ibriz, M. 2014. Variability and genetic structure of a natural population of Citrus psorosis virus. *Annals of Microbiology* 1590-4261:.

Agüero, J. Ruiz-Ruiz, S. Vives, M.d.C. Velázquez, K. Navarro, L. Peña, L. Moreno, P. and Guerri, J. 2012. Development of Viral Vectors Based on Citrus leaf blotch virus to Express Foreign Proteins or Analyze Gene Function in Citrus Plants. *Molecular Plant-Microbe Interactions* 25:10:1326-1337.

Ahmad, R. Struss, D. and Southwick, S.M. 2003. Development and characterization of microsatellite markers in Citrus. *J.Am.Soc.Hort.Sci.* 128:4:584-590.

Albrechtsen, A. Nielsen, F.C. and Nielsen, R. 2010. Ascertainment biases in SNP chips affect measures of population divergence. *Mol.Biol.Evol.* 27:11:2534-2547.

Aleza, P. Cuenca, J. Juárez, J. Pina, J.A. and Navarro, L. 2010. 'Garbí' Mandarin: a new late-maturing triploid hybrid. *HortScience* 45:139-141.

Aleza, P. Cuenca, J. Hernández, M. Juárez, J. Navarro, L. and Ollitrault, P. Submitted. Genetic mapping of centromeres of the nine Citrus clementina chromosomes using half-tetrad analysis and recombination patterns in unreduced and haploid gametes. *BMC plant biology*.

Aleza, P. Froelicher, Y. Schwarz, S. Agusti, M. Hernandez, M. Juarez, J. Luro, F. Morillon, R. Navarro, L. and Ollitrault, P. 2011. Tetraploidization events by chromosome doubling of nucellar cells are frequent in apomictic citrus and are dependent on genotype and environment. *Ann.Bot.*

Aleza, P. Juarez, J. Cuenca, J. Ollitrault, P. and Navarro, L. 2012. Extensive citrus triploid hybrid production by 2x4x sexual hybridizations and parent-effect on the length of the juvenile phase. *Plant Cell Rep.* 31:9:1723-1735.

Aleza, P. Juarez, J. Cuenca, J. Ollitrault, P. and Navarro, L. 2010. Recovery of citrus triploid hybrids by embryo rescue and flow cytometry from 2x x 2x sexual hybridisation and its application to extensive breeding programs. *Plant Cell Rep.* 29:9:1023-1034.

Aleza, P. Juarez, J. Ollitrault, P. and Navarro, L. 2009. Production of tetraploid plants of non apomictic citrus genotypes. *Plant Cell Rep.* 28:12:1837-1846.

Allario, T. Brumos, J. Colmenero-Flores, J.M. Iglesias, D.J. Pina, J.A. Navarro, L. Talon, M. Ollitrault, P. and Morillon, R. 2013. Tetraploid Rangpur lime rootstock increases drought tolerance via enhanced constitutive root abscisic acid production. *Plant.Cell.Envirn.* 36:4:856-868.

Alonso, V. and Howell, E. 1983. Comparative-study of the hybrids of the citrus species, tangor temple and ortanique. *Fruits* 38:9:647-652.

Al-Sadi, A.M. Al-Moqbali, H.S. Al-Yahyai, R.A. and Al-Said, F.A. 2012. AFLP data suggest a potential role for the low genetic diversity of acid lime (*Citrus aurantifolia* Swingle) in Oman in the outbreak of witches' broom disease of lime. *Euphytica* 188:2:285-297.

Amar, M.H. Biswas, M.K. Zhang Zong Wen and Guo Wen Wu. 2011. Exploitation of SSR, SRAP and CAPS-SNP markers for genetic diversity of Citrus germplasm collection. *Scientia Horticulturae* 128:3:220-227.

Anses. 2012. AVIS de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail relatif à « Evaluation du Risque Simplifiée de *Xylella fastidiosa* ». Avis de l'ANSES Saisine n° 2012-SA-0121:1-5.

Arnold, M.L. 1997. *Natural hybridization and evolution*. Oxford University Press, New York.

Asíns, M.J. Juarez, J. Pina, J.A. Puchades, J. Carbonel, E.A. and Navarro, L. 2002. Una nueva clementina de baja fertilidad llamada Nulesín. *Vida Rural* 157:50-52.

Association pour le Développement des Etudes Archéologiques, Historiques, linguistiques et Naturalistes du Centre-Est de la Corse. <http://www.adecec.net/lexiques/afrutta3.html>. 2014.

Aubert, B. and Vullin, G. 1997. *Pépinières et plantations d'agrumes*. Quae, Montpellier.

Bacchi, O. 1940. Observações citológicas em citrus. I. Número de cromosômios de algumas espécies e variedades. *Jorn. Agron. [Piracicaba]* 3:249-58.

Baird, N.A. Etter, P.D. Atwood, T.S. Currey, M.C. Shiver, A.L. Lewis, Z.A. Selker, E.U. Cresko, W.A. and Johnson, E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:10:e3376.

Barbazuk, W.B. Emrich, S.J. Chen, H.D. Li, L. and Schnable, P.S. 2007. SNP discovery via 454 transcriptome sequencing. *Plant J.* 51:5:910-918.

Barkley, N.A. Krueger, R.R. Federici, C.T. and Roose, M.L. 2009. What phylogeny and gene genealogy analyses reveal about homoplasy in citrus microsatellite alleles. *Plant Syst.Evol.* 282:1/2:71-86.

Barkley, N.A. Roose, M.L. Krueger, R.R. and Federici, C.T. 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor.Appl.Genet.* 112:8:1519-1531.

Barnhart, R.K. 1988. *Barnhart Dictionary of Etymology*. H.W., Wilson Co.

Barrett, H.C. and Rhodes, A.M. 1976. A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. *Syst.Bot.* 1:105-136.

Bassene, J.B. Berti, L. Carcouet, E. Dhuique-Mayer, C. Fanciullino, A.L. Bouffin, J. Ollitrault, P. and Froelicher, Y. 2008. Influence of mitochondria origin on fruit quality in a citrus cybrid. *J.Agric.Food Chem.* 56:18:8635-8640.

Bassene, J.B. Froelicher, Y. Navarro, L. Ollitrault, P. and Ancillo, G. 2011. Influence of mitochondria on gene expression in a citrus cybrid. *Plant Cell Rep.* 30:6:1077-1085.

Bastianel, M. Schwarz, S.F. Coleta Filho, H.D. Lin, L.L. Marcos, M. and Koller, O.C. 1998. Identification of zygotic and nucellar tangerine seedlings (*Citrus* spp.) using RAPD. *Genetics and Molecular Biology* 21:123-127.

Bausher, M. Singh, N. Lee, S. Jansen, R. and Daniell, H. 2006. The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biology* 6:1:21.



Bayer, R.J. Mabblerley, D.J. Morton, C. Miller, C.H. Sharma, I.K. Pfeil, B.E. Rich, S. Hitchcock, R. and Sykes, S. 2009. A molecular phylogeny of the orange subfamily(Rutaceae: Aurantioideae) using nine cpDNA sequences. *Am J Bot.* 96:3:668-685.

Beattie, G.A.C. Holford, P. Mabblerley, D.J. Haigh, A.M. and Broadbent, P. 2008. On the Origins of Citrus, Huanglongbing, Diaphorina citri and Trioza erytreae. *IRCHLB Proceedings*23-56.

Beiko, R.G. and Hamilton, N. 2006. Phylogenetic identification of lateral genetic transfer events. *BMC Evol.Biol.* 6:15.

Belkhir, K. Borsa, P. Chikhi, L. Raufaste, N. and Bonhomme, F. 1996-2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000 Université de Montpellier II, Montpellier (France).

Benson, D.A. Clark, K. Karsch-Mizrachi, I. Lipman, D.J. Ostell, J. and Sayers, E.W. 2014. GenBank. *Nucleic Acids Res.* 42:Database issue:D32-7.

Berhow, M.A. Hasegawa, S. Kwan, K. and Bennett, R.D. 2000. Limonoids and the chemotaxonomy of Citrus and the Rutaceae family. p. 212-228. In: M.A. Berhow S. Hasegawa and G.D. Manners (eds.), *Citrus Limonoids: Functional Chemicals in Agriculture and Foods*, American Chemical Society, Washington DC.

Bernardi, J. Licciardello, C. Patrizia Russo, M. Luisa Chiusano, M. Carletti, G. Reforgiato Recupero, G. and Marocco, A. 2010. Use of a custom array to study differentially expressed genes during blood orange (*Citrus sinensis* L. Osbeck) ripening. *J.Plant Physiol.* 167:4:301-310.

Bernardi, J. Mazza, R. Caruso, P. Reforgiato Recupero, G. Marocco, A. and Licciardello, C. 2013. Molecular Breeding; Use of an expressed sequence tag-based method for single nucleotide polymorphism identification and discrimination of Citrus species and cultivars. *Molecular Breeding* 31:3:705-718.

Bernet, G.P. FernandezRibacoba, J. Carbonell, E.A. and Asins, M.J. 2010. Comparative genome-wide segregation analysis and map construction using a reciprocal cross design to facilitate citrus germplasm utilization. *Mol.Breed.* 25:4:659-673.

Bertioli, D.J. Ozias-Akins, P. Chu, Y. Dantas, K.M. Santos, S.P. Gouvea, E. Guimaraes, P.M. Leal-Bertioli, S.C. Knapp, S.J. and Moretzsohn, M.C. 2014. The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. *G3 (Bethesda)* 4:1:89-96.

Bhattacharya, S.C. and Dutta, S. 1956. Classification of citrus fruits of Assam. *Sci. Monogr.* 20.

Biraghi, A. 1935. Rilievi su alcuni Citrus a frutto acido presenti in India in relazione alla ricerca di forme resistenti al "mal secco.". *Bolletino della Reale Stazione di Patologia Vegetale, Roma Bol.* (n.s.) 15:424-441.

Biswas, M.K. Chai, L. Mayer, C. Xu, Q. Guo, W. and Deng, X. 2012. Exploiting BAC-end sequences for the mining, characterization and utility of new short sequences repeat (SSR) markers in Citrus. *Mol.Biol.Rep.* 39:5:5373-5386.

Biswas, M.K. Chai, L. Amar, M.H. Zhang, X. and Deng, X. 2011. Comparative analysis of genetic diversity in Citrus germplasm collection using AFLP, SSAP, SAMPL and SSR markers. *Scientia Horticulturae* 129:4:798-803.

Biswas, M.K. Xu, Q. and Deng, X. 2010. Utility of RAPD, ISSR, IRAP and REMAP markers for the genetic analysis of Citrus spp. *Scientia Horticulturae* 124:2:254-261.

- Blervaque, R. 2013. Séquençage haut-débit de deuxième génération: Principes et caractéristiques (2); <http://www.biorigami.com/?p=5492>. Biorigami.
- Blondel, L. 1978. Classification botanique des espèces du genre Citrus. *Fruits* 33:11:695-721.
- Bonavia, E. 1973. Oranges and lemons of India and Ceylan. Reprint Bishen and periodical Expers, Delhi.
- Bonavia, E. 1888. The cultivated oranges and lemons, etc. of India and Ceylon. W. H. Allen, London.
- Bové, J.M. and Duran-Vila, N. 2013. Quelques maladies infectieuses des agrumes transmises par greffage d'inoculation. p. 109-146. In: C. Jacquemond F. Curk and M. Heuzet (eds.), Les clémentiniers et autres petits agrumes, Quae, Versailles.
- Bradbury, I.R. Hubert, S. Higgins, B. Bowman, S. Paterson, I.G. Snelgrove, P.V. Morris, C.J. Gregory, R.S. Hardie, D.C. Borza, T. and Bentzen, P. 2011. Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. *Mol.Ecol.Resour.* 11 Suppl 1:218-225.
- Bretó, M.P. Ruiz, C. Pina, J.A. and Asins, M.J. 2001. The diversification of *Citrus clementina* Hort. ex Tan., a vegetatively propagated crop species. *Mol.Phylogenet.Evol.* 21:2:285-293.
- Britten, R.J. Rowen, L. Williams, J. and Cameron, R.A. 2003. Majority of divergence between closely related DNA samples is due to indels. *Proc.Natl.Acad.Sci.U.S.A.* 100:8:4661-4665.
- Brookes, A.J. 1999. The essence of SNPs. *Gene* 234:2:177-186.
- Bryan, G. McNicoll, J. Ramsay, G. Meyer, R. and De Jong, W. 1999. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. *Theor.Appl.Genet.* 99:5:859-867.
- Bundock, P.C. Elliott, F.G. Ablett, G. Benson, A.D. Casu, R.E. Aitken, K.S. and Henry, R.J. 2009. Targeted single nucleotide polymorphism (SNP) discovery in a highly polyploid plant species using 454 sequencing. *Plant Biotechnology Journal* 7:4:347-354.
- Bybee, S.M. Bracken-Grissom, H. Haynes, B.D. Hermansen, R.A. Byers, R.L. Clement, M.J. Udall, J.A. Wilcox, E.R. and Crandall, K.A. 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome Biol.Evol.* 3:1312-1323.
- Cai, Q. Guy, C.L. and Moore, G.A. 1994. Extension of the linkage map in Citrus using random amplified polymorphic DNA (RAPD) markers and RFLP mapping of cold-acclimation-responsive loci. *Theoretical and Applied Genetics* 89:606-614.
- Carbonell-Caballero, J. Alonso, R. Ibañez, V. Terol, J. Talon, M. and Dopazo, J. submitted. Comparative analysis of the chloroplast genome of the genus Citrus. *Genome research*.
- Carlos de Oliveira, A. Bastianel, M. Cristofani-Yaly, M. Morais do Amaral, A. and Machado, M.A. 2007. Development of genetic maps of the citrus varieties 'Murcott' tangor and 'Pera' sweet orange by using fluorescent AFLP markers. *J.Appl.Genet.* 48:3:219-231.
- Carlson, C.S. Smith, J.D. Stanaway, I.B. Rieder, M.J. and Nickerson, D.A. 2006. Direct detection of null alleles in SNP genotyping data. *Hum.Mol.Genet.* 15:12:1931-1937.
- Carneiro, M.O. Russ, C. Ross, M.G. Gabriel, S.B. Nusbaum, C. and DePristo, M.A. 2012. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* 13:375-2164-13-375.
- Caruso, M. Federici, C.T. and Roose, M.L. 2008. EST-SSR markers for asparagus genetic diversity evaluation and cultivar identification. *Mol.Breed.* 21:2:195-204.

- Carvalho, R. Soares Filho, W.S. Brasileiro-Vidal, A.C. and Guerra, M. 2005. The relationships among lemons, limes and citron: a chromosomal comparison. *Cytogenetic and Genome Research* 109:(1/3):276-282.
- Cervera, M. Esteban, O. Gil, M. Gorris, M.T. Martínez, M.C. Peña, L. and Cambra, M. 2010. Transgenic expression in citrus of single-chain antibody fragments specific to Citrus tristeza virus confers virus resistance. *Transgenic Res.* 19:6:1001-1015.
- Chai, L. Biswas, M.K. Yi, H. Guo, W. and Deng, X. 2013. Transferability, polymorphism and effectiveness for genetic mapping of the Pummelo (*Citrus grandis* Osbeck) EST-SSR markers. *Scientia Horticulturae* 155:0:85-91.
- Chapot, H. 1965. Quelques varieties d'agrumes d'Iran. *Al Awamia* 15:91-109.
- Chapot, H. 1962. Le Bergamotier. *Al Awamia* 5:1-27.
- Chauvet, M. 2001. Les voyages des agrumes. *Hommes et Plantes, revue du Conservatoire français des collections végétales spécialisées (CCVS)* 37:17-21.
- Chen, C. and Gmitter, F.G. 2013. Mining of haplotype-based expressed sequence tag single nucleotide polymorphisms in citrus. *BMC Genomics* 14:1:746.
- Chen, L.G. Omura, M. and Hidaka, T. 1991. A study on the taxonomy of citrus with GOT isozymes. *Acta Horticulturae Sinica* 18:1:27-32.
- Chen, C.X. Bowman, K.D. Choi, Y.A. Dang, P.M. Rao, M.N. Huang, S. Soneji, J.R. McCollum, T.G. and Gmitter, F.G., Jr. 2008a. EST-SSR genetic maps for *Citrus sinensis* and *Poncirus trifoliata*. *Tree Genetics and Genomes* 4:1:1-10.
- Chen, D. Ahlford, A. Schnorrer, F. Kalchhauser, I. Fellner, M. Viragh, E. Kiss, I. Syvanen, A.C. and Dickson, B.J. 2008b. High-resolution, high-throughput SNP mapping in *Drosophila melanogaster*. *Nat.Methods* 5:4:323-329.
- Cheng, Y. de Vicente, M.C. Meng, H. Guo, W. Tao, N. and Deng, X. 2005. A set of primers for analyzing chloroplast DNA diversity in Citrus and related genera. *Tree Physiol.* 25:6:661-672.
- Chin, H.F. and Roberts, E.H. 1980. Recalcitrant crop seeds. Kuala Lumpur.
- Ching, A. Caldwell, K. Jung, M. Dolan, M. Smith, O. Tingey, S. Morgante, M. and Rafalski, A. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3:1:19.
- Cipriani, G. Marrazzo, M.T. Marconi, R. Cimato, A. and Testolin, R. 2002. Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. *Theoretical and Applied Genetics* 104:2-3:223-228.
- Clark, A.G. Hubisz, M.J. Bustamante, C.D. Williamson, S.H. and Nielsen, R. 2005. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* 15:11:1496-1502.
- CNRS - ATILF. 2014. <http://www.cnrtl.fr/>. 2014.
- Cock, P.J.A. Antao, T. Chang, J.T. Chapman, B.A. Cox, C.J. Dalke, A. Friedberg, I. Hamelryck, T. Kauff, F. Wilczynski, B. and de Hoon, M.J.L. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:11:1422-1423.

Cohen, S. Jacquemond, C. Bouffin, J. Dubois, A. Venard, P. Vernière, C. and Panabières, F. 2003. Les phytophthora des agrumes en Corse : une évolution des populations pathogènes à maîtriser par la mise en oeuvre d'une lutte raisonnée = Citrus fruit phytophthora in Corsica. A population growth which requires the implementation of IPM techniques. *Phytoma - La défense des végétaux* 562:22-25.

Combrink, N.K. Labuschagne, M.T. and Bijzet, Z. 2013. Variation of fruit size and shape in Kiyomi tangor families. *Scientia Horticulturae* 162:357-364.

Cooper, W., Colby. 1982. In search of the golden apple. Vintage Press, New York.

Corazza-Nunes, M. Machado, M. Nunes, W. Cristofani, M. and Targon, M. 2002. Assessment of genetic variability in grapefruits (*Citrus paradisi* Macf.) and pummelos (*C. maxima* (Burm.) Merr.) using RAPD and SSR markers. *Euphytica* 126:2:169-176.

Cornille, A. Gladioux, P. Smulders, M.J. Roldán-Ruiz, I. Laurens, F. Le Cam, B. Nersesyan, A. Clavel, J. Olonova, M. and Feugey, L. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS genetics* 8:5:e1002703.

Couplan, F. 2012. Les plantes et leurs noms. Quae, Paris.

Crespan, M. 2004. Evidence on the evolution of polymorphism of microsatellite markers in varieties of *Vitis vinifera* L. *Theoretical and Applied Genetics* 108:2:231-237.

Cristofani, M. Machado, M.A. and Grattapaglia, D. 1999. Genetic linkage maps of *Citrus sunki* Hort. ex. Tan. and *Poncirus trifoliata* (L.)Raf. and mapping of citrus tristeza virus resistance gene. *Euphytica* 109:25-32.

Cristofani-Yaly, M. Novelli, V.M. Bastianel, M. and Machado, M.A. 2011. Transferability and level of heterozygosity of microsatellite markers in *Citrus* species. *Plant Mol.Biol.Rep.* 29:2:418-423.

Cuenca, J. Aleza, P. Navarro, L. and Ollitrault, P. 2013a. Assignment of SNP allelic configuration in polyploids using competitive allele-specific PCR: application to citrus triploid progeny. *Ann.Bot.* 111:4:731-742.

Cuenca, J. Aleza, P. Vicent, A. Brunel, D. Ollitrault, P. and Navarro, L. 2013b. Genetically based location from triploid populations and gene ontology of a 3.3-mb genome region linked to alternaria brown spot resistance in citrus reveal clusters of resistance genes. *PLoS One* 8:10:e76755.

Cuenca, J. Froelicher, Y. Aleza, P. Juárez, J. Navarro, L. and Ollitrault, P. 2011. Multilocus half-tetrad analysis and centromere mapping in citrus: evidence of SDR mechanism for 2n megagametophyte production and partial chiasma interference in mandarin cv 'Fortune'. *Heredity*.

Cuenca, J. Aleza, P. Juárez, J. Pina, J.A. and Navarro, L. 2010. 'Safor' Mandarin: A New Citrus Mid-late Triploid Hybrid. *HortScience* 45:6:977-980.

Cuppen, E. 2007. Genotyping by Allele-Specific Amplification (KASPar). *CSH Protoc.* 2007:pdb.prot4841.

Multilocus SNPs analysis allows phylogenetic assignation of DNA fragments to decipher the interspecific mosaic genome structure of cultivated Citrus.

Curk, F. Ancillo, G. Garcia-Lor, A. Luro, F. Perrier, X. Jacquemoud-Collet, J.P. Navarro, L. and Ollitrault, P. accepted. Next generation haplotyping to decipher the nuclear genomic interspecific admixture in *Citrus* species; analysis of the Chromosome 2. *BMC Genetics*.

Curk, F. Ancillo, G. Ollitrault, F. Perrier, X. Jacquemoud-Collet, J.P. Garcia-Lor, A. Navarro, L. and Ollitrault, P. submitted. Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern Citrus varieties.

Curk, F. Ancillo, G. Garcia-Lor, A. Luro, F. Navarro, L. and Ollitrault, P. 2011 Multilocus SNPs analysis allows phylogenetic assignation of DNA fragments to decipher the intraspecific mosaic genome structure of cultivated Citrus. Plant Genome Evolution Congress 2011, 4 - 6 Sep 2011, P2.21

Curk, F. Garcia-Lor, A. Snoussi, H. Froelicher, Y. Ancillo, G. Navaro, L. and Ollitrault, P. in press. New Insights on Limes and Lemons Origin from Nuclear and Cytoplasmic Markers Genotyping and Targeted Nuclear Gene Sequencing. Acta Horticulturae. Proceedings XII International Citrus Congress 2012.

Curk, F. Ollitrault, F. Garcia-Lor, A. Luro, F. Ancillo, G. Navarro, L. and Ollitrault, P. submitted. Nuclear, mitochondrial and chloroplastic SSR, Indel and SNP markers reveal the diversity and phylogenetic origin of limes and lemons.

Curk, F. Ancillo, G. Garcia-Lor, A. Luro, F. Navarro, L. and Ollitrault, P. in press. Multilocus haplotyping by parallel sequencing to decipher the interspecific mosaic genome structure of cultivated Citrus. Acta Horticulturae Proceedings XII International Citrus Congress 2012.

Dambier, D. Benyahia, H. Pensabene-Bellavia, G. Aka Kacar, Y. Froelicher, Y. Belfalah, Z. Lhou, B. Handaji, N. Printz, B. Morillon, R. Yesiloglu, T. Navarro, L. and Ollitrault, P. 2011. Somatic hybridization for citrus rootstock breeding: an effective tool to solve some important issues of the Mediterranean citrus industry. Plant Cell Rep.

Das, A. 2003. Citrus canker: a review. Journal of Applied Horticulture 5:52-60.

Davey, J. and Blaxter, M.L. 2011. RADSeq: next-generation population genetics. Brief Funct Genomics 10:2:108.

de Araújo, E.F. de Queiroz, L.P. and Machado, M.A. 2003. What is Citrus? Taxonomic implications from a study of cp-DNA evolution in the tribe Citreae (Rutaceae subfamily Aurantioideae). Organisms Diversity & Evolution 3:1:55-62.

de Bakker, P.I. Burt, N.P. Graham, R.R. Guiducci, C. Yelensky, R. Drake, J.A. Bersaglieri, T. Penney, K.L. Butler, J. Young, S. Onofrio, R.C. Lyon, H.N. Stram, D.O. Haiman, C.A. Freedman, M.L. Zhu, X. Cooper, R. Groop, L. Kolonel, L.N. Henderson, B.E. Daly, M.J. Hirschhorn, J.N. and Altshuler, D. 2006. Transferability of tag SNPs in genetic association studies in multiple populations. Nat.Genet. 38:11:1298-1303.

de Candolle, A. 1886. Origin of cultivated plants. 2nd ed. Kegan Paul, London.

de Moraes, A. dos Santos Soares Filho, W. and Guerra, M. 2007. Karyotype diversity and the origin of grapefruit. Chromosome Research 1:115-121.

de Oliveira, R.P. Aguilar-Vildoso, C.L. Cristofani, M. and Machado, M.A. 2004. Skewed RAPD markers in linkage maps of Citrus. Genetics and Molecular Biology 27:437-441.

de Simone, M. Russo, M.P. Puelo, G. Marsan, P.A. Lorenzoni, C. Marocco, A. and Recupero, G.R. 1998. Construction of genetic maps for Citrus aurantium and C. latipes based on AFLP, RAPD and RFLP markers. Fruits 53:383-390.

Delseny, M. Han, B. and Hsing, Y.I. 2010. High throughput DNA sequencing: The new sequencing revolution. Plant Sci. 179:5:407-422.

Deng, X. 2008. Citrus varieties in China. China Agriculture Press, Wuhan.

Deng, Z. La Malfa, S. Xie, Y. Xiong, X. and Gentile, A. 2007. Identification and evaluation of chloroplast uni-and trinucleotide sequence repeats in citrus. *Scientia horticulturae* 111:2:186-192.

Department of Energy's Joint Genome Institute and Center for Integrative Genomics. 2014. Phytozome; <http://www.phytozome.net/>.

Dereeper, A. Nicolas, S. Lecunff, L. Bacilieri, R. Doligez, A. Peros, J.P. Ruiz, M. and This, P. 2013. <http://sniplay.cirad.fr/cgi-bin/home.cgi>. 2014.

Dereeper, A. Nicolas, S. Lecunff, L. Bacilieri, R. Doligez, A. Peros, J.P. Ruiz, M. and This, P. 2011. SNIPlay: a web-based tool for detection, management and analysis of SNPs. Application to grapevine diversity projects. *BMC Bioinformatics* 12:1:134.

Deschamps, S. Llaca, V. and May, G.D. 2012. Genotyping-by-Sequencing in Plants. *Biology* 1:3:460-83.

Dhuique-Mayer, C. Caris-Veyrat, C. Ollitrault, P. Curk, F. and Amiot, M.J. 2005. Varietal and interspecific influence on micronutrient contents in citrus from the Mediterranean area. *J.Agric.Food Chem.* 53:6:2140-2145.

Dhuique-Mayer, C. Fanciullino, A.L. Dubois, C. and Ollitrault, P. 2009. Effect of genotype and environment on citrus juice carotenoid content. *J.Agric.Food Chem.* 57:19:9160-9168.

Ding, S.Q. Zhang, X.N. Bao, Z.R. and Liang, M.Q. 1984. *Poncirus polyandra*. *Acta Bot. Yunnan* 6:(3):292.

Distefano, G. Caruso, M. Casas, G.L. La Malfa, S. Gentile, A. Tribulato, E. Simsek, O. and Rapisarda, P. 2011. Maturation in Fruits of Clementine: Biochemical and Molecular Analysis. *Acta Horticulturae* 892:II International Symposium on Citrus Biotechnology:329-334.

Distefano, G. La Malfa, S. Gentile, A. and Wu, S. 2013. EST-SNP genotyping of citrus species using high-resolution melting curve analysis. *Tree Genetics & Genomes* 9:5:1271-1281.

DNASTAR Inc. 2014. DNASTAR Software for life Scientists; <http://www.dnastar.com/t-nextgen-seqman-ngen.aspx>.

Doolittle, W.F. 1999. Phylogenetic Classification and the Universal Tree. *Science* 284 (5423):2124-9.

Duportal, M. Jorda, E. Sanchez, C. Imbert, É. Loeillet, D. and Vannière, H. 2013. FruiTrop FOCUS Citron. *FruiTrop Cirad Hors Série* 2013:140.

Duran-Vila, N. González-Arnao, M.T. and Engelmann, F. 1997. Cryopreservation and *in vitro* culture. Proceedings of the citrus germplasm conservation workshop, Brisbane, Australia, 6-7 October, 1997..1999.22-32.44 ref.

Durham, R.E. Liou, P.C. Gmitter, F.G., Jr. and Moore, G.A. 1992. Linkage of restriction fragment length polymorphisms and isozymes in *Citrus*.. *Theor.Appl.Genet.* 84:1-2:39-48.

Duval, H. Hoerter, M. Polidori, J. Confolent, C. Masse, M. Moretti, A. Van Ghelder, C. and Esmenjaud, D. 2014. High-resolution mapping of the RMia gene for resistance to root-knot nematodes in peach. *Tree Genetics & Genomes* 10:2:297-306.

Edwards, D. and Batley, J. 2010. Plant genome sequencing: applications for crop improvement. *Plant.Biotechnol.J.* 8:1:2-9.

El-Mouei, R. Choumane, W. and Dway, F. 2011. Characterization and Estimation of Genetic Diversity in Citrus Rootstocks. *International Journal of Agriculture & Biology* 13:4.

Elshire, R.J. Glaubitz, J.C. Sun, Q. Poland, J.A. Kawamoto, K. Buckler, E.S. and Mitchell, S.E. 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6:(5):e19379. doi:10.1371/journal.pone.0019379-doi:10.1371/journal.pone.0019379.

Engelmann, F. 1997. In vitro conservation methods. p. 119-162. In: B.V. Ford-Lloyd H.J. Newbury and J.A. Callow (eds.), *Biotechnology and Plant Genetic Resources. Conservation and Use*, CABI, Wallingford, UK.

Engelmann, F. Balsemin, E. Barreneche, T. Chatelet, P. Chauvin, J. Couturon, E. Curk, F. Dantec, M. Dantec, J. Dussert, S. Feugey, L. Froelicher, Y. Fouilhaux, L. Gamiette, F. Grapin, A. Grisoni, M. Guérif, P. Guyader, A. Label, A. Luro, F. Moulin, B. Muller, M. Peyrière, A. Prigent, Y. Renard, M. Roux-Cuvelier, M. Roques, D. Rubens, S. Sapotille, J. Souchet, C. and Teyssedre, D. 2009. Cryopreservation of French plant genetic resource collections (CRYOVEG). *Cryobiology* 59:3:411.

Esen, A. and Soost, R.K. 1971. Unexpected triploids in citrus: their origin, identification and possible use. *The Journal of Heredity* 62:329-333.

Esselink, G. Nybom, H. and Vosman, B. 2004. Assignment of allelic configuration in polyploids using the MAC-PR (microsatellite DNA allele counting—peak ratios) method. *Theor.Appl.Genet.* 109:2:402-408.

Evanno, G. Regnaut, S. and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol.Ecol.* 14:8:2611-2620.

Falush, D. Stephens, M. and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:4:1567-1587.

Fanciullino, A.L. Dhuique-Mayer, C. Luro, F. Casanova, J. Morillon, R. and Ollitrault, P. 2006. Carotenoid diversity in cultivated citrus is highly influenced by genetic factors. *J.Agric.Food Chem.* 54:12:4397-4406.

Fanciullino, A.L. Dhuique-Mayer, C. Luro, F. Morillon, R. and Ollitrault, P. 2007. Carotenoid biosynthetic pathway in the citrus genus: number of copies and phylogenetic diversity of seven genes. *J.Agric.Food Chem.* 55:18:7405-7417.

Fang, D.Q. Krueger, R.R. and Roose, M.L. 1998. Phylogenetic relationships among selected *Citrus* germplasm accessions revealed by inter-simple sequence repeat (ISSR) markers. *J.Am.Soc.Hort.Sci.* 123:4:612-617.

Fang, D.Q. Roose, M.L. Krueger, R.R. and Federici, C.T. 1997. Fingerprinting trifoliolate orange germ plasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. *Theor.Appl.Genet.* 95:1/2:211-219.

Fang, J. Song, C. Qian, J. Zhang, X. Shangguan, L. Yu, H. and Wang, X. 2010. Variation of cytosine methylation in 57 sweet orange cultivars. *Acta Physiologiae Plantarum* 32:6:1023-1030.

FAO. 2014a. FAOSTAT <http://faostat3.fao.org/home/E>.

FAO. 2014b. The International Treaty on Plant Genetic Resources for Food and Agriculture; <http://www.planttreaty.org/fr>.

Federici, C.T. Fang, D.Q. Scora, R.W. and Roose, M.L. 1998. Phylogenetic relationships within the genus *Citrus* (*Rutaceae*) and related genera as revealed by RFLP and RAPD analysis. *Theor.Appl.Genet.* 96:6/7:812-822.

Federici, C.T. Roose, M.L. and Scora, R.W. 2000. RFLP analysis of the origin of *Citrus bergamia*, *Citrus jambhiri*, and *Citrus limonia*. *Acta Horticulturae* 535:55-62.

Ferrari, G.B. 1646. *Hesperides siue de malorum aureorum cultura et vsu Libri quator Io. Baptistae Ferrarii Senensis e Societate Iesu*. Réédition de Romae: Sumptibus Hermanni Scheus, 1646; Milano: F. Motta, 1992, Milano (Italy).

Fluidigm Corp. 2014. Access Array TM System; <http://www.fluidigm.com/access-array-system.html>.

Forment, J. Gadea, J. Huerta, L. Abizanda, L. Agusti, J. Alamar, S. Alos, E. Andres, F. Arribas, R. Beltran, J.P. Berbel, A. Blazquez, M.A. Brumos, J. Canas, L.A. Cercos, M. Colmenero-Flores, J.M. Conesa, A. Estables, B. Gandia, M. Garcia-Martinez, J.L. Gimeno, J. Gisbert, A. Gomez, G. Gonzalez-Candelas, L. Granell, A. Guerri, J. Lafuente, M.T. Madueno, F. Marcos, J.F. Marques, M.C. Martinez, F. Martinez-Godoy, M.A. Miralles, S. Moreno, P. Navarro, L. Pallas, V. Perez-Amador, M.A. Perez-Valle, J. Pons, C. Rodrigo, I. Rodriguez, P.L. Royo, C. Serrano, R. Soler, G. Tadeo, F. Talon, M. Terol, J. Trenor, M. Vaello, L. Vicente, O. Vidal, C. Zacarias, L. and Conejero, V. 2005. Development of a citrus genome-wide EST collection and cDNA microarray as resources for genomic studies. *Plant Mol.Biol.* 57:3:375-391.

Fortune, P. Pourtau, N. Viron, N. and Ainouche, M. 2008. Molecular phylogeny and reticulate origins of the polyploid *Bromus* species from section *Genea* (Poaceae). *Am J Bot* 95:4:454-464.

France Génomique. 2014. Mutualisation des compétences et des équipements français pour l'analyse génomique et la bio-informatique; <https://www.france-genomique.org/spip/>.

Fredon Corse. 2014. <http://www.fredon-corse.com/>.

Froelicher, Y. Dambier, D. Bassene, J.B. Costantino, G. Lotfy, S. Didout, C. Beaumont, V. Brottier, P. Risterucci, A.M. Luro, F. and Ollitrault, P. 2008. Characterization of microsatellite markers in mandarin orange (*Citrus reticulata* Blanco). *Mol.Ecol.Resour.* 8:1:119-122.

Froelicher, Y. Mouhaya, W. Bassene, J.B. Costantino, G. Kamiri, M. Luro, F. Morillon, R. and Ollitrault, P. 2011. New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. *Tree Genetics and Genomes* 7:1:49-61.

Fu, Y. Cheng, B. and Peterson, G.W. 2014. Genetic diversity analysis of yellow mustard (*Sinapis alba* L.) germplasm based on genotyping by sequencing. *Genet.Resour.Crop Evol.* 61:3:579-594.

Fu, Y. and Peterson, G.W. 2011. Genetic Diversity Analysis with 454 Pyrosequencing and Genomic Reduction Confirmed the Eastern and Western Division in the Cultivated Barley Gene Pool. *Plant Gen.* 4:3:226-237.

Fujii, H. Shimada, T. Nonaka, K. Kita, M. Kuniga, T. Endo, T. Ikoma, Y. and Omura, M. 2013. High-throughput genotyping in citrus accessions using an SNP genotyping array. *Tree Genetics & Genomes* 9:1:145-153.

Gallesio, G. 1811. *Traité du citrus*. Chez Louis Fantin Libraire, Paris.

Garcia, R. Asins, M.J. Forner, J. and Carbonell, E.A. 1999. Genetic analysis of apomixis in *Citrus* and *Poncirus* by molecular markers. *Theor.Appl.Genet.* 99:3/4:511-518.

Garcia-Lor, A. Ancillo, G. Navarro, L. and Ollitrault, P. 2013a. Citrus (Rutaceae) SNP Markers Based on Competitive Allele-Specific PCR; Transferability Across the Aurantioideae Subfamily. *Applications in Plant Sciences* 1:4:1200406.

Garcia-Lor, A. Curk, F. Snoussi-Trifa, H. Morillon, R. Ancillo, G. Luro, F. Navarro, L. and Ollitrault, P. 2013b. A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the



relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. *Ann.Bot.* 111:1:1-19.

Garcia-Lor, A. Luro, F. Navarro, L. and Ollitrault, P. 2012. Comparative use of InDel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: a perspective for genetic association studies. *Mol.Genet.Genomics* 287:1:77-94.

Garg, K. Green, P. and Nickerson, D.A. 1999. Identification of candidate coding region single nucleotide polymorphisms in 165 human genes using assembled expressed sequence tags. *Genome Res.* 9:11:1087-1092.

Gentile, F. Moretti, M. Limongi, T. Falqui, A. Bertoni, G. Scarpellini, A. Santoriello, S. Maragliano, L. Proietti Zaccaria, R. and di Fabrizio, E. 2012. Direct Imaging of DNA Fibers: The Visage of Double Helix. *Nano Letters* 12:12:6453-6458.

Ghorabaie, H.R.R. Ghazvini, R.F. Golein, B. and Nabipour, A.R. 2010. Identification of some *Citrus* accessions in a citrus germplasm utilizing simple sequence repeat markers (SSRs). *Horticulture, Environment and Biotechnology* 51:4:343-347.

Gmitter, F.G., Jr. Chen, C.X. Rao, M.N. and Soneji, J.R. 2007. Citrus fruits. (Genome Mapping and Molecular Breeding in Plants, Volume 4). *Fruits and nuts* 265-279.

Goethe, J., Wolfgang. 1795. *Wilhelm Meisters Lehrjahre*. Berlin.

Golein, B. Fifaei, R. and Ghasemi, M. 2011. Identification of zygotic and nucellar seedlings in citrus interspecific crosses by inter simple sequence repeats (ISSR) markers. *African Journal of Biotechnology* 10:82:18965-18970.

Golein, B. Talaie, A. Zamani, Z. Ebadi, A. and Behjatnia, A. 2005. Assessment of genetic variability in some Iranian sweet oranges (*Citrus sinensis* [L.] Osbeck) and mandarins (*Citrus reticulata* Blanco) using SSR markers. *International Journal of Agriculture and Biology* 7:2:167-170.

Golein, B. Bigonah, M. Azadvar, M. and Golmohammadi, M. 2012. Analysis of genetic relationship between 'Bakraee'(Citrus sp.) and some known Citrus genotypes through SSR and PCR-RFLP markers. *Scientia Horticulturae* 148:147-153.

Gomez-Cadenas, A. Tadeo, F.R. Talon, M. and Primo-Millo, E. 1996. Leaf Abscission Induced by Ethylene in Water-Stressed Intact Seedlings of Cleopatra Mandarin Requires Previous Abscisic Acid Accumulation in Roots. *Plant Physiol.* 112:1:401-408.

González-Arno, M.T. Juárez, J. Ortega, C. Navarro, L. and Duran-Vila, N. 2003. Cryopreservation of ovules and somatic embryos of citrus using the encapsulation-dehydration technique. *Cryo-letters* 24:85-94.

Grant, V. 1981. *Plant Speciation*. Colombia University Press, New York.

Green, R.M. Vardi, A. and Galun, E. 1986. The plastome of Citrus. Physical map, variation among Citrus cultivars and species and comparison with related genera. *Theor.Appl.Genet.* 72:2:170-177.

Griffin, P.C. Robin, C. and Hoffmann, A.A. 2011. A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. *BMC Biol.* 9:19-7007-9-19.

Grosser, J.W. Dutt, M. Omar, A. Orbovic, V. and Barthe, G. 2011. Progress Towards the Development of Transgenic Disease Resistance in Citrus. *Acta Hort. (ISHS)* 892:101:107.

Grosser, J.W. Deng, X.X. and Goodrich, R.M. 2007. Somaclonal variation in sweet orange: practical applications for variety improvement and possible causes. *Citrus genetics, breeding and biotechnology* 219-233.

Grosser, J.W. and Gmitter, F.G. 2011. Protoplast fusion for production of tetraploids and triploids: applications for scion and rootstock breeding in citrus. (Special Issue: In vitro ploidy manipulation in the genomics era.). *Plant Cell, Tissue and Organ Culture* 104:3:343-357.

Grosser, J.W. and Gmitter, F.G., Jr. 1990. Protoplast fusion and citrus improvement. *Plant Breed.Rev.* 8:339-374.

Guerra, M. 1993. Cytogenetics of Rutaceae. V. High chromosomal variability in *Citrus* species revealed by CMA/DAPI staining. *HEREDITY-LONDON-* 71:234-234.

Gulsen, O. Uzun, A. Canan, I. Seday, U. and Canihos, E. 2010. A new citrus linkage map based on SRAP, SSR, ISSR, POGP, RGA and RAPD markers. *Euphytica* 173:265-277.

Gulsen, O. and Roose, M.L. 2001a. Chloroplast and nuclear genome analysis of the parentage of lemons. *J.Am.Soc.Hort.Sci.* 126:2:210-215.

Gulsen, O. and Roose, M.L. 2001b. Determination of genetic diversity and phylogenetic relations to citrus ancestors in lemons by DNA markers. *Bahce* 30:1/2:53-63.

Gulsen, O. and Roose, M.L. 2001c. Lemons: diversity and relationships with selected *Citrus* genotypes as measured with nuclear genome markers. *J.Am.Soc.Hort.Sci.* 126:3:309-317.

Guo, W.W. and Grosser, J.W. 2004. Transfer of a potential canker resistance gene into *Citrus* protoplasts using GFP as the selectable marker. *Acta Horticulturae* 632:255-258.

Han, Y. Chagné, D. Gasic, K. Rikkerink, E.H. Beever, J.E. Gardiner, S.E. and Korban, S.S. 2009. BAC-end sequence-based SNPs and Bin mapping for rapid integration of physical and genetic maps in apple. *Genomics* 93:3:282-288.

Handa, T. and Oogaki, C. 1985. Numerical taxonomic study of *Citrus* L. and *Fortunella* swingle using morphological characters application of multivariate analysis. *Journal of the Japanese Society for Horticultural Science (Japan)*.

Hayashi, K. Yoshida, H. and Ashikawa, I. 2006. Development of PCR-based allele-specific and InDel marker sets for nine rice blast resistance genes. *Theor.Appl.Genet.* 113:2:251-260.

Hayden, C.E. 2012. Nanopore genome sequencer makes its debut. *Nature* doi:10.1038/nature.2012.10051.

Heckman, K.L.A. and Pease, L.R. 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature protocols* 2:924:232.

Hensz, R.A. 1977. Mutation breeding and the development of the 'Star Ruby' grapefruit. *Proceedings of the International Society of Citriculture* 2:582-585.

Hensz, R.A. 1971. Star Ruby, a new deep-red-fleshed grapefruit variety with distinct tree characteristics. *J. Rio Grande Valley Hort. Soc.* 25:54-58.

Herrero, R. Asins, M.J. Pina, J.A. Carbonell, E.A. and Navarro, L. 1996. Genetic diversity in the orange subfamily Aurantioideae. II. Genetic relationships among genera and species. *Theor.Appl.Genet.* 93:8:1327-1334.

Hillier, L.W. Marth, G.T. Quinlan, A.R. Dooling, D. Fewell, G. Barnett, D. Fox, P. Glasscock, J.I. Hickenbotham, M. Huang, W. Magrini, V.J. Richt, R.J. Sander, S.N. Stewart, D.A. Stromberg, M. Tsung, E.F. Wylie, T. Schedl, T. Wilson, R.K. and Mardis, E.R. 2008. Whole-genome sequencing and variant discovery in *C. elegans*. *Nat.Methods* 5:2:183-188.

Hodgson, R.W. 1955. Origin of citrus fruits grown in California. *Calif. Citrogr.* 49:132-36.

Hörandl, E. and Paun, O. 2007. Patterns and sources of genetic diversity in apomictic plants: implications for evolutionary potentials. p. 281-298. In: E. Hörandl U. Grossniklaus P.J. Van Dijk and T.F. Sharbel (eds.), *Apomixis: evolution, mechanisms and perspectives*, Regnum Vegetabile, Koenigstein, Germany.

Hoskins, R.A. Phan, A.C. Naeemuddin, M. Mapa, F.A. Ruddy, D.A. Ryan, J.J. Young, L.M. Wells, T. Kopczyński, C. and Ellis, M.C. 2001. Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Res.* 11:6:1100-1113.

Hou, X. Liu, S. Khan, M. Hu, C. and Zhang, J. 2014. Genome-Wide Identification, Classification, Expression Profiling, and SSR Marker Development of the MADS-Box Gene Family in Citrus. *Plant Molecular Biology Reporter* 32:1:28-41.

Huang, X. Kurata, N. Wei, X. Wang, Z. Wang, A. Zhao, Q. Zhao, Y. Liu, K. Lu, H. and Li, W. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:7421:497-501.

Huazhong Agricultural University. 2014. Citrus sinensis Annotation Project; <http://citrus.hzau.edu.cn/orange/>.

Hufford, M.B. Xu, X. Van Heerwaarden, J. Pyhäjärvi, T. Chia, J. Cartwright, R.A. Elshire, R.J. Glaubitz, J.C. Guill, K.E. and Kaeppler, S.M. 2012. Comparative population genomics of maize domestication and improvement. *Nat.Genet.* 44:7:808-811.

ICGC. 2014. International Citrus Genome / Genomics Consortium; <http://www.citrusgenome.ucr.edu/>.

Imbert, É. 2013. Production et marché des agrumes. p. 147-162. In: C. Jacquemond F. Curk and M. Heuzet (eds.), *Les Clémentiniers et autres petits agrumes*, Quae, Versailles.

Isaac, E. 1959. Influence of Religion on the Spread of Citrus: The religious practices of the Jews helped effect the introduction of citrus to Mediterranean lands. *Science* 129:3343:179-186.

Jacquemond, C. Curk, F. Froelicher, Y. and Luro, F. 2013a. Variétés et porte-greffes : création, description et sélection. p. 37-107. In: C. Jacquemond F. Curk and M. Heuzet (eds.), *Les clémentiniers et autres petits agrumes*, Quae, Versailles.

Jacquemond, C. Curk, F. and Heuzet, M. 2013b. Créer des plants : pépinière et surgreffage. p. 183-207. In: C. Jacquemond F. Curk and M. Heuzet (eds.), *Les clémentiniers et autres petits agrumes*, Quae, Versailles.

Jacquemond, C. Tison, G. Kreiter, P. Bové, J.M. Duran-Villa, N. and Curk, F. 2013c. Protéger le verger : maladies et ravageurs. p. 279-343. In: C. Jacquemond F. Curk and M. Heuzet (eds.), *Les clémentiniers et autres petits agrumes*, Quae, Versailles.

Jaeger, B. Goldbach, H. and Sommer, K.,. 2000. Release from lime induced iron chlorosis by CULTAN in fruit trees and its characterisation by analysis. *Acta Horticulturae* 531:107-113.

Jander, G. Norris, S.R. Rounsley, S.D. Bush, D.F. Levin, I.M. and Last, R.L. 2002. Arabidopsis map-based cloning in the post-genome era. *Plant Physiol.* 129:2:440-450.

Jarrell, D.C. Roose, M.L. Traugh, S.N. and Kupper, R.S. 1992. A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. *Theor.Appl.Genet.* 84:1-2:49-56.

Jena, S.N. Kumar, S. and Nair, N.K. 2009. Molecular phylogeny in Indian *Citrus* L.(Rutaceae) inferred through PCR-RFLP and trnL-trnF sequence data of chloroplast DNA. *Scientia horticulturae* 119:4:403-416.

Jiang Dong Ye QingLiang Wang FuSheng and Cao, L. 2010. The mining of citrus EST-SNP and its application in cultivar discrimination. *Agricultural Sciences in China* 9:2:179-190.

Jiang Dong Zhong GuangYan and Hong, Q. 2006. Analysis of microsatellites in citrus unigenes. *Acta Genet.Sin.* 33:4:345-353.

Jumpponen, A. and Jones, K.L. 2009. Massively parallel 454-sequencing of *Quercus macrocarpa* phyllosphere fungal communities indicates reduced richness and diversity in urban environments. *New Phytologist* 184:438-448.

Jung, Y. Kwon, H. Kang, S. Kang, J. and Kim, S. 2005. Investigation of the phylogenetic relationships within the genus *Citrus* (Rutaceae) and related species in Korea using plastid trnL-trnF sequences. *Scientia Horticulturae* 104:2:179-188.

Kacar, Y.A. Simsek, O. Mendi, Y.Y. Yildirim, B. Yesiloglu, T. Distefano, G. La Malfa, S. and Boncuk, M. 2012. Investigation of the Self Incompatibility Mechanism in Clementine (*Citrus clementina* Hort. ex Tan.) Using cDNA-AFLP. *Acta Horticulturae* 929:233-236.

Kaçar, Y.A. Şimşek, Ö. Dönmez, D. Boncuk, M. Yeşiloğlu, T. and Ollitrault, P. 2014. Genetic relationships of some *Citrus* genotypes based on the candidate iron chlorosis genes. *Turkish Journal of Agriculture and Forestry* 38:3:340-347.

Kacar, Y. Uzun, A. Polat, I. Yesiloglu, T. Yilmaz, B. Gulsen, O. Tuzcu, O. Kamiloglu, M. Kurt, S. and Seday, U. 2013. Molecular characterization and genetic diversity analysis of mandarin genotypes by SSR and SRAP markers. *International journal of food, agriculture and environment (Print)* 11:1:516-521.

Kamiri, M. Stift, M. Srairi, I. Costantino, G. Moussadik, A.E. Hmyene, A. Bakry, F. Ollitrault, P. and Froelicher, Y. 2011. Evidence for non-disomic inheritance in a *Citrus* interspecific tetraploid somatic hybrid between *C. reticulata* and *C. limon* using SSR markers and cytogenetic analysis. *Plant Cell Rep.*

Kepiro, J.L. and Roose, M.L. 2010. AFLP markers closely linked to a major gene essential for nucellar embryony (apomixis) in *Citrus maxima* x *Poncirus trifoliata*. *Tree Genetics and Genomes* 6:1:1-11.

Khalaf, A.A. Gmitter, F.G., Jr Conesa, A. Dopazo, J. and Moore, G.A. 2011. *Fortunella margarita* transcriptional reprogramming triggered by *Xanthomonas citri* subsp. *citri*. *BMC Plant.Biol.* 11:159-2229-11-159.

Kijas, J. Fowler, J. and Thomas, M. 1995. An evaluation of sequence tagged microsatellite site markers for genetic analysis within *Citrus* and related species. *Genome* 38:2:349-355.

Kijas, J.M.H. Thomas, M.R. Fowler, J.C.S. and Roose, M.L. 1997. Integration of trinucleotide microsatellites into a linkage map of *Citrus*. *Theor.Appl.Genet.* 94:5:701-706.

Kim, J.H. Handayani, E. Wakana, A. Sakai, K. Sato, M. and Han, J.H. 2013. Segregation of Self-incompatible Hybrid Seedlings in Crosses with Grapefruit and possible RAPD Markers for the S Gene Alleles. *journal of the faculty of Agriculture Kyushu University* 58:2:269-275.

- Kobayashi, M. Nagasaki, H. Garcia, V. Just, D. Bres, C. Mauxion, J.P. Le Paslier, M.C. Brunel, D. Suda, K. Minakuchi, Y. Toyoda, A. Fujiyama, A. Toyoshima, H. Suzuki, T. Igarashi, K. Rothan, C. Kaminuma, E. Nakamura, Y. Yano, K. and Aoki, K. 2014. Genome-wide analysis of intraspecific DNA polymorphism in 'Micro-Tom', a model cultivar of tomato (*Solanum lycopersicum*). *Plant Cell Physiol.* 55:2:445-454.
- Krueger, R.R. and Navarro, L. 2007. Citrus germplasm resources. p. 45-140. In: Anonymous Citrus Genetics, Breeding and Biotechnology, CAB International, Wallingford, UK.
- Kulheim, C. Yeoh, S.H. Maintz, J. Foley, W.J. and Moran, G.F. 2009. Comparative SNP diversity among four Eucalyptus species for genes from secondary metabolite biosynthetic pathways. *BMC Genomics* 10:452-2164-10-452.
- Kumamoto, J. Scora, R.W. Lawton, H.W. and Clerx W., A. 1987. Mystery of the Forbidden Fruit: Historical Epilogue on the Origin of the Grapefruit, *Citrus paradisi* (*Rutaceae*). *Economic Botany* 41:(1):97-107.
- Kumar, S. Nair, K.N. and Jena, S. 2013. Molecular differentiation in Indian Citrus L. (*Rutaceae*) inferred from nrDNA ITS sequence analysis. *Genetic Resources and Crop Evolution* 60:1:59-75.
- Kurtzleben, D. 2014. The Vox Lime Index: making sense of the Limepocalypse; <http://www.vox.com/2014/5/9/5696596/vox-tracks-the-limepocalypse>. Vox.
- Lecoq, H. and Juillet, J. 1831. Dictionnaire raisonné des termes de botanique et des familles naturelles. Librairie de l'Académie Royale de Médecine, Clermont-Ferrand.
- Li, H. Futch, S.H. and Syvertsen, J.P. 2007. Cross-correlation patterns of air and soil temperatures, rainfall and *Diaprepes abbreviatus* root weevil in citrus. *Pest Manag.Sci.* 63:11:1116-1123.
- Li, X. Xie, R. Lu, Z. and Zhou, Z. 2010. The Origin of Cultivated Citrus as Inferred from Internal Transcribed Spacer and Chloroplast DNA Sequence and Amplified Fragment Length Polymorphism Fingerprints. *Journal of the American Society for Horticultural Science* 135:4:341-350.
- Li, Y.Z. Cheng, Y.J. Yi, H.L. and Deng, X.X. 2006. Genetic diversity in mandarin landraces and wild mandarins from China based on nuclear and chloroplast simple sequence repeat markers. *Journal of Horticultural Science and Biotechnology* 81:3:371-378.
- Liang, G. Xiong, G. Guo, Q. He, Q. and Li, X. 2007. AFLP analysis and the taxonomy of Citrus. *Acta Hort.* 760:137-142.
- Licciardello, C. Russo, M.P. Recupero, G.R. Bernardi, J. Marocco, A. Butelli, E. Martin, C. Chiusano, M.L. Gentile, A. and la Malfa, S. 2011. Anthocyanins: the 'secret' of Sicilian blood oranges. *Acta Horticulturae* 892:II International Symposium on Citrus Biotechnology, Catania, Italy:359-365.
- Life Sciences Corp. 2010. 454 Sequencing System Software Manual, v 2.5.3; Part C – GS De Novo Assembler, GS Reference Mapper, SFF Tools . 454 Sequencing System Software Manual2-213.
- Life Sciences Corp. 2009a. Amplicon Fusion Primer Design Guidelines for GS FLX Titanium Series Lib-A Chemistry. Technical Bulletin Genome Sequencer FLX System 013:1-3.
- Life Sciences Corp. 2009b. Using Multiplex Identifier (MID) Adaptors for the GS FLX Titanium Chemistry - Extended Mid Set. Technical Bulletin Genome Sequencer FLX System 005:1-7.
- Lijavetzky, D. Cabezas, J.A. Ibanez, A. Rodriguez, V. and Martinez-Zapater, J.M. 2007. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8:424.

Linder, C.R. and Rieseberg, L.H. 2004. Reconstructing patterns of reticulate evolution in plants. *Am.J.Bot.* 91:1700-1708.

Ling, P. Yu, C. Deng, Z. Chen, C. Huang, S. Wendell, M.K. and Gmitter, J.F.G. 1999. Citrus genome mapping with AFLP markers. *Plant & Animal Genomes XIII Conf, San Diego, CA, USA Poster.*

Linnaeus, C. 1753. *Species Plantarum*. Vol. II. Holmiae (Stockholm), Impensis Laurentii Salvii [Printed for the Ray Society, London, in 1959.],.

Liou, P.C. Gmitter, F.G., Jr. and Moore, G.A. 1996. Characterization of the *Citrus* genome through analysis of restriction fragment length polymorphisms. *Theor.Appl.Genet.* 92:3/4:425-435.

Liu, C. Weng, J. Zhang, D. Zhang, X. Yang, X. Shi, L. Meng, Q. Yuan, J. Guo, X. and Hao, Z. 2014a. Genome-wide association study of resistance to rough dwarf disease in maize. *Eur.J.Plant Pathol.* 9:11:1087-1092.

Liu, H. Bayer, M. Druka, A. Russell, J.R. Hackett, C.A. Poland, J. Ramsay, L. Hedley, P.E. and Waugh, R. 2014b. An evaluation of genotyping by sequencing (GBS) to map the *Breviaristatum-e* (arie) locus in cultivated barley. *BMC Genomics* 15:1:104.

Liu, J. Pang, X. Cheng, Y. Meng, H. and Deng, X. 2002. Molecular characterization of the nuclear and cytoplasmic genomes of intergeneric diploid plants from cell fusion between *Microcitrus papuana* and Rough lemon. *Plant Cell Rep.* 21:4:327-332.

Liu, Y.Z. and Deng, X.X. 2007. Citrus breeding and genetics in China. *The Asian and Australasian Journal of Plant Science and Biotechnology* 1:23-28.

Liu, C. Wang, J. Huang, T. Wang, F. Yuan, F. Cheng, X. Zhang, Y. Shi, S. Wu, J. and Liu, K. 2010. A missense mutation in the VHYNP motif of a DELLA protein causes a semi-dwarf mutant phenotype in *Brassica napus*. *Theoretical and Applied Genetics* 121:2:249-258.

Liu, S. Li, W. Long, D. Hu, C. and Zhang, J. 2013. Development and Characterization of Genomic and Expressed SSRs in Citrus by Genome-Wide Analysis. *PLoS ONE* 8:10:e75149.

Lombardoa, G. Schicchia, R. Marinao, P. and Pallaa, F. 2012. Genetic analysis of *Citrus aurantium* L. (Rutaceae) cultivars by ISSR molecular markers. *Plant Biosystems* 146:1:19-26.

Loret, V. 1892. *La flore pharaonique d'après les documents hiéroglyphiques et les spécimens découverts dans les tombes.* Leroux, Ernest, Paris.

Loret, V. 1891. *Le cédratier dans l'antiquité (réédition 2004).* Connaissance et Mémoires, Paris.

Lofly, S. Luro, F. Carreel, F. Froelicher, Y. Rist, D. and Ollitrault, P. 2003. Application of cleaved amplified polymorphic sequence method for analysis of cytoplasmic genome among *Aurantioideae* intergeneric somatic hybrids. *J.Am.Soc.Hort.Sci.* 128:2:225-230.

Lu, Z. Zhou, Z. and Xie, R. 2011. Molecular Phylogeny of the "True Citrus Fruit Trees" Group (Aurantioideae, Rutaceae) as Inferred from Chloroplast DNA Sequence. *Agricultural Sciences in China* 10:1:49-57.

Luan, B. Peng, H. Polonsky, S. Rossnagel, S. Stolovitzky, G. and Martyna. 2010. Base-By-Base Ratcheting of Single Stranded DNA through a Solid-State Nanopore. *Phys. Rev. Lett.* 104:238103.

Luro, F. Gatto, J. Costantino, G. and Pailly, O. 2011. Analysis of genetic diversity in Citrus. *Plant Genetic Resources* 9:218-221.

Luro, F. Laigret, F. Lorieux, M. and Ollitrault, P. 1996. Citrus genome mapping with molecular markers: two maps obtained by segregation analysis of progeny of one intergeneric cross. *Proc Intl Soc Citricult* 2:862-866.

Luro, F. Jacquemond, C. and Curk, F. 2013. La clémentine dans la diversité génétique des agrumes. p. 17-36. In: C. Jacquemond F. Curk and M. Heuzet (eds.), *Les clémentiniers et autres petits agrumes*, Quae, Versailles.

Luro, F. Costantino, G. Terol, J. Argout, X. Allario, T. Wincker, P. Talon, M. Ollitrault, P. and Morillon, R. 2008. Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC Genomics* 9:287.

Luro, F. Laigret, F. Bove, J.M. and Ollitrault, P. 1995. DNA amplified fingerprinting, a useful tool for determination of genetic origin and diversity analysis in *Citrus*. *HortScience* 30:5:1063-1067.

Luro, F. Rist, D. and Ollitrault, P. 2001. Evaluation of genetic relationships in *Citrus* genus by means of sequence tagged microsatellites. *Acta Horticulturae* 546:237-242.

Luro, F. Venturini, N. Costantino, G. Paolini, J. Ollitrault, P. and Costa, J. 2012. Genetic and chemical diversity of citron (*Citrus medica* L.) based on nuclear and cytoplasmic markers and leaf essential oil composition. *Phytochemistry* 77:186-196.

Mabberley, D.J. 2004. Citrus (Rutaceae): a review of recent advances in etymology, systematics and medical applications. *Blumea* 49:(2/3):481-498.

Mabberley, D.J. 1997. A classification for edible Citrus (Rutaceae). *Telopea* 7:(2):167-172.

Maddison, W.P. and Knowles, L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic biology* 55:1:21-30.

Malik, M.N. Scora, R., W. and Soost, R., K. 1974. Studies on the origin of the lemon. *Hilgardia* 42:(9):361-382.

Malik, S.K. Uchoi, A. Kumar, S. Choudhary, R. Pal, D. Kole, P.R. and Bhat, K.V. 2013. Molecular characterization of *Citrus macroptera* Montr.(Satkara): An endangered wild species from northeast India. *Plant Biosystems* 147:4:857-863.

Manrique-Carpintero, N.C. Tokuhisa, J.G. Ginzberg, I. and Veilleux, R.E. 2014. Allelic variation in genes contributing to glycoalkaloid biosynthesis in a diploid interspecific population of potato. *Theor.Appl.Genet.* 127:2:391-405.

Margulies, M. Egholm, M. Altman, W.E. Attiya, S. Bader, J.S. Bemben, L.A. Berka, J. Braverman, M.S. Chen, Y.J. Chen, Z. Dewell, S.B. Du, L. Fierro, J.M. Gomes, X.V. Godwin, B.C. He, W. Helgesen, S. Ho, C.H. Irzyk, G.P. Jando, S.C. Alenquer, M.L. Jarvie, T.P. Jirage, K.B. Kim, J.B. Knight, J.R. Lanza, J.R. Leamon, J.H. Lefkowitz, S.M. Lei, M. Li, J. Lohman, K.L. Lu, H. Makhijani, V.B. McDade, K.E. McKenna, M.P. Myers, E.W. Nickerson, E. Nobile, J.R. Plant, R. Puc, B.P. Ronan, M.T. Roth, G.T. Sarkis, G.J. Simons, J.F. Simpson, J.W. Srinivasan, M. Tartaro, K.R. Tomasz, A. Vogt, K.A. Volkmer, G.A. Wang, S.H. Wang, Y. Weiner, M.P. Yu, P. Begley, R.F. and Rothberg, J.M. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:7057:376-380.

Martín, S. Alioto, D. Milne, R.G. Garnsey, S., M. García, M.L. Grau, O. Guerri, J. and Moreno, P. 2004. Detection of *Citrus psorosis virus* by ELISA, molecular hybridization, RT-PCR and immunosorbent electron microscopy and its association with citrus psorosis disease. *European Journal of Plant Pathology* 110:747-757.

Martinez-Godoy, M.A. Mauri, N. Juarez, J. Marques, M.C. Santiago, J. Forment, J. and Gadea, J. 2008. A genome-wide 20 K citrus microarray for gene expression analysis. *BMC Genomics* 9:318.

- Mascher, M. Wu, S. Amand, P.S. Stein, N. and Poland, J. 2013. Application of Genotyping-by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and Reference-Based Marker Ordering in Barley. *PLoS ONE*10:76925.
- Maughan, P.J. Yourstone, S.M. Jellen, E.N. and Udall, J.A. 2009. SNP Discovery via Genomic Reduction, Barcoding, and 454-Pyrosequencing in Amaranth. *Plant Gen.* 2:3:260-270.
- McCouch, S.R. Zhao, K. Wright, M. Tung, C. Ebana, K. Thomson, M. Reynolds, A. Wang, D. DeClerck, G. Ali, M.L. McClung, A. Eizenga, G. and Bustamante, C. 2010. Development of genome-wide SNP assays for rice. *Breed.Sci.* 60:5:524-535.
- McNally, B. Singer, A. Yu, Z. Sun, Y. Weng, Z. and Meller, A. 2010. Optical Recognition of Converted DNA Nucleotides for Single-Molecule DNA Sequencing Using Nanopore Arrays. *Nano Letters* 10:6:2237-2244.
- McNally, K.L. Bruskiewich, R. Mackill, D. Buell, C.R. Leach, J.E. and Leung, H. 2006. Sequencing multiple and diverse rice varieties. Connecting whole-genome variation with phenotypes. *Plant Physiol.* 141:1:26-31.
- Mercati, F. Riccardi, P. Leebens-Mack, J. Abenavoli, M.R. Falavigna, A. and Sunseri, F. 2013. Single nucleotide polymorphism isolated from a novel EST dataset in garden asparagus (*Asparagus officinalis* L.). *Plant Science* 203–204:0:115-123.
- Moldenke, H., N. and Moldenke, A., L. 1952. *Plants of the bible*. The Ronald Press & *Chronica Botanica*, Waltham.
- Moore, G.A. 2001. Oranges and lemons: clues to the taxonomy of Citrus from molecular markers. *TRENDS in Genetics* 17:9:536-540.
- Morales, M. Roig, E. Monforte, A.J. Arus, P. and Garcia-Mas, J. 2004. Single-nucleotide polymorphisms detected in expressed sequence tags of melon (*Cucumis melo* L.). *Genome* 47:2:352-360.
- Moreno, P. Ambros, S. Albiach-Marti, M.R. Guerri, J. and Pena, L. 2008. Citrus tristeza virus: a pathogen that changed the course of the citrus industry. *Mol.Plant.Pathol.* 9:2:251-268.
- Morton, J. 1987. Tahiti Lime. In: J.F. Morton (ed.), *Fruits of warm climates*. Miami, FL.
- Mosends, P. 2014. Here's how the the great lime shortage is affecting bars across america; <http://www.thewire.com/business/2014/05/heres-how-the-lime-shortage-is-affecting-bars-across-america/362049/>. *The Wire*.
- Muys, C. Thienpont, C. Dauchot, N. Maudoux, O. Draye, X. and Cutsem, P.V. 2014. Integration of AFLPs, SSRs and SNPs markers into a new genetic map of industrial chicory (*Cichorium intybus* L. var. *sativum*). *Plant Breeding* 133:1:130-137.
- Myles, S. Boyko, A.R. Owens, C.L. Brown, P.J. Grassi, F. Aradhya, M.K. Prins, B. Reynolds, A. Chia, J.M. Ware, D. Bustamante, C.D. and Buckler, E.S. 2011. Genetic structure and domestication history of the grape. *Proc.Natl.Acad.Sci.U.S.A.* 108:9:3530-3535.
- Nair, P.K.R. and Randhawa, G.S. 1969. Chromosome morphology of the pachytene stage with respect to differnt Citrus types. In: Chapmn (ed.), *Proc. First Int. Citrus Symp.* (Univ. California, Riverside)1:215-223.
- Nartvaranant, P. and Nartvaranant, K. 2011. Analysis based on AFLP markers of the genetic variations and their relationships for pummelo cultivars grown in the central region of Thailand. *Songklanakarin Journal of Science and Technology* 33:5:499-508.



- Nature. 2010. Human genome at ten: The sequence explosion. *Nature* 464:7289:670-671.
- Naz, S. Shahzadi, K. Rashid, S. Saleem, F. Zafarullah, A. and Ahmad, S. 2014. Molecular Characterization and Phylogenetic Relationship of Different Citrus Varieties of Pakistan. *Journal of Animal and Plant Sciences* 24:1:315-320.
- NCBI. 2014. <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>.
- Negrini, R. Mazza, R. Colli, L. Pellicchia, M. Bomba, L. Stella, A. Williams, J.L. and Ajmone-Marsan, P. 2010. Effect Of Non-Random Sampling Of SNPs On The Estimation Of Population Genetic Parameters. Leipzig. Germany August 1-6, 2010.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc.Natl.Acad.Sci.U.S.A.* 70:12:3321-3323.
- Nematollahi, A. Vahdati, K. Golein, B. Nabipour, A. and Danesh, M. 2009. Assessment of Genetic Variation in Citrus Germplasm of South of Iran Using SSR Markers. *Acta Horticulturae* 839:1 International Symposium on Biotechnology of Fruit Species: Biotechfruit2008:549-554.
- Nicolosi, E. 2007. Origin and Taxonomy. p. 19-43. In: I. Khan (ed.), *Citrus genetics, breeding and biotechnology*, Muscat, Sultanate of Oman.
- Nicolosi, E. Deng, Z.N. Gentile, A. La Malfa, S. Continella, G. and Tribulato, E. 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics* 100:1155-1166.
- Nicolosi, E. La Malfa, S. El-Otmani, M. Negbi, M. and Goldschmidt, E.E. 2005. The search for the authentic citron (*Citrus medica* L.): Historic and genetic analysis. *HortScience* 40:1963-1968.
- Novelli, V.M. Cristofani, M. and Machado, M.A. 2000. Evaluation of microsatellite markers in cultivars of sweet orange (*Citrus sinensis* [L.] Osbeck). *Acta Horticulturae* 535:47-50.
- Novelli, V.M. Cristofani, M. Souza, A.A. and Machado, M.A. 2006. Development and characterization of polymorphic microsatellite markers for the sweet orange (*Citrus sinensis* L. Osbeck). *Genetics and Molecular Biology* 29:1:90-96.
- Novelli, V.M. Takita, M.A. and Machado, M.A. 2004. Identification and analysis of single nucleotide polymorphisms (SNPs) in citrus. *Euphytica* 138:3:227-237.
- Nybom, H. 2007. Unique reproduction in dogroses (*Ros* sect. *Canina*) maintains successful and highly heterozygous genotypes. p. 169-194. In: E. Hörandl U. Grossniklaus P.J. Van Dijk and T.F. Sharbel (eds.), *Apomixis: evolution, mechanisms and perspectives*, Regnum Vegetabile, Koenigstein, Germany.
- Office québécois de la langue française. 2014. Le grand dictionnaire terminologique (GDT); <http://gdt.oqlf.gouv.qc.ca/index.aspx>. 2014.
- Olivares-Fuster, O. Hernandez-Garrido, M. Guerri, J. and Navarro, L. 2007. Plant somatic hybrid cytoplasmic DNA characterization by single-strand conformation polymorphism. *Tree Physiol.* 27:6:785-792.
- Ollitrault, P. Terol, J. Chen, C. Federici, C.T. Lofty, S. Hippolyte, I. Ollitrault, F. Berard, A. Chauveau, A. Constantino, G. Kacar, Y. Mu, L. Cuenca, J. Garcia-Lor, A. Froelicher, Y. Aleza, P. Boland, A. Billot, C. Navarro, L. Luro, F. Roose, M.L. Gmitter, F.G. Talon, M. and Brunel, D. 2011. A reference linkage map of *C. clementina* based on SNPs, SSRs and Indels.

Ollitrault, P. and De Rocca Serra, D. 1992. Les agrumes. p. 633-654. In: A. Gallais and H. Bannerot (eds.), Amélioration des espèces végétales cultivées, Inra, Paris.

Ollitrault, P. Garcia-Lor, A. Terol, J. Curk, F. Ollitrault, F. Talon, M. and Navarro, L. 2014. Comparative Values of SSRs, SNPs and InDels for Citrus Genetic Diversity Analysis. Acta Horticulturae Proceeding XII international Citrus Congress 2012.

Ollitrault, P. Jacquemond, C. Dubois, C. and Luro, F. 2003. Citrus. p. 193-217. In: Hamon Perla Seguin Marc Perrier Xavier and Glaszmann Jean-Christophe (eds.), Genetic diversity of cultivated tropical plants, Cirad, Montpellier.

Ollitrault, P. and Navarro, L. 2012. Citrus. p. 623-662. In: M. Badenes and D.e. Byrne (eds.), Fruit Breeding, Springer New York, London.

Ollitrault, F. Terol, J. Martin, A.A. Pina, J.A. Navarro, L. Talon, M. and Ollitrault, P. 2012. Development of indel markers from Citrus clementina (Rutaceae) BAC-end sequences and interspecific transferability in Citrus. Am.J.Bot. 99:7:e268-73.

Ollitrault, F. Terol, J. Pina, J.A. Navarro, L. Talon, M. and Ollitrault, P. 2010. Development of SSR markers from Citrus clementina (Rutaceae) BAC end sequences and interspecific transferability in Citrus. Am.J.Bot. 97:11:e124-9.

Ollitrault, P. Dambier, D. Luro, F. and Froelicher, Y. 2008. Ploidy manipulation for breeding seedless triploid citrus. Plant Breed.Rev. 30:323-352.

Ollitrault, P. Terol, J. Chen, C. Federici, C.T. Lotfy, S. Hippolyte, I. Ollitrault, F. Berard, A. Chauveau, A. Cuenca, J. Costantino, G. Kacar, Y. Mu, L. Garcia-Lor, A. Froelicher, Y. Aleza, P. Boland, A. Billot, C. Navarro, L. Luro, F. Roose, M.L. Gmitter, F.G. Talon, M. and Brunel, D. 2012a. A reference genetic map of C. clementina hort. ex Tan.; citrus evolution inferences from comparative mapping. BMC Genomics 13:593-2164-13-593.

Ollitrault, P. Terol, J. Garcia-Lor, A. Berard, A. Chauveau, A. Froelicher, Y. Belzile, C. Morillon, R. Navarro, L. Brunel, D. and Talon, M. 2012b. SNP mining in C. clementina BAC end sequences; transferability in the Citrus genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. BMC Genomics 13:13-2164-13-13.

Ollitrault, P. Vanel, F. Froelicher, Y. and Dambier, D. 2000. Creation of triploid citrus hybrids by electrofusion of haploid and diploid protoplasts. Acta Horticulturae535:191-197.

Omura, M. Ueda, T. Kita, M. Komatsu, A. Takanokura, Y. Shimada, T. Endo-Inagaki, T. Nesumi, H. and Yoshida, T. 2000. EST mapping of Citrus. Proceedings of the International Society of Citriculture IX congress. Orlando, USA..

Orbovic, V. Soria, P. Moore, G.A. and Grosser, J.W. 2011. The use of citrus tristeza virus (ctv) containing a green fluorescent protein gene as a tool to evaluate resistance/tolerance of transgenic citrus plants. Crop Protection 30:5:572-576.

Organisation des Nations Unies pour l'Alimentation et l'Agriculture. 2009. Traité international sur les ressources phytogénétiques pour l'alimentation et l'agriculture. Rome.

Otto, S.P. and Whitton, J. 2000. Polyploid incidence and evolution. Annu.Rev.Genet. 34:401-437.

Oxford Nanopore Technologies. 2014. The MinION™ device: a miniaturised sensing system; <https://www.nanoporetech.com/technology/the-minion-device-a-miniaturised-sensing-system/the-minion-device-a-miniaturised-sensing-system>.

Pagnoux, C. Celant, A. Coubray, S. Fiorentino, G. and Zech-Matterne, V. 2013. The introduction of Citrus to Italy, with reference to the identification problems of seed remains. *Vegetation History and Archaeobotany* 22:5:421-438.

Pamilo, P. and Nei, M. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol* 5:568-583.

Pang, X. Hu, C.G. and Deng, X.X. 2007. Phylogenetic relationships within Citrus and its related genera as inferred from AFLP markers. *Genetic Resources and Crop Evolution* 54:429-436.

Pasquali, G. Orbovic, V. and Grosser, J.W. 2009. Transgenic grapefruit plants expressing the  $P_{APETALA3}$ - $IPT_{GP}$  gene exhibit altered expression of PR genes. *Plant Cell, Tissue and Organ Culture* 97:2:215-223.

Penjor, T. Yamamoto, M. Uehara, M. Ide, M. Matsumoto, N. Matsumoto, R. and Nagano, Y. 2013. Phylogenetic relationships of Citrus and its relatives based on matK gene sequences. *PloS one* 8:4:e62574.

Perrier, X. and Jacquemoud-Collet, J.P. 2006. DARwin software. <http://darwin.cirad.fr/>.

Pessina, D. Gentili, R. Barcaccia, G. Nicolè, S. Rossi, G. Barbesti, S. and Sgorbati, S. 2011. DNA content, morphometric and molecular marker analyses of Citrus limonimediterranea, C. limon and C. medica for the determination of their variability and genetic relationships within the genus Citrus. *Scientia Horticulturae* 129:4:663-673.

Peterson, B.K. Weber, J.N. Kay, E.H. Fisher, H.S. and Hoekstra, H.E. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* 7:5:e37135.

Polat, I. Kacar, Y. Yesiloglu, T. Uzun, A. Tuzcu, O. Gulsen, O. Incesu, M. Kafa, G. Turgutoglu, E. and Anil, S. 2012. Molecular characterization of sour orange (Citrus aurantium) accessions and their relatives using SSR and SRAP markers. *Genetics and Molecular Research* 11:3:3267-3276.

Polonsky, S. Rossnagel, S. and Stolovitzky, G. 2007. Nanopore in metal-dielectric sandwich for DNA position control. *Appl. Phys. Lett.* 91:153103.

Pons, E. Peris, J.E. and Peña, L. 2012. Field performance of transgenic citrus trees: Assessment of the long-term expression of uidA and nptII transgenes and its impact on relevant agronomic and phenotypic characteristics. *BMC biotechnology* 12:1:41.

Pootakham, W. Chanprasert, J. Jomchai, N. Sangsrakru, D. Yoocha, T. Therawattanasuk, K. and Tangphatsornruang, S. 2011. Single nucleotide polymorphism marker development in the rubber tree, Hevea brasiliensis (Euphorbiaceae). *American Journal of Botany* 98:11:e337-e338.

Pootakham, W. Uthapaisanwong, P. Sangsrakru, D. Yoocha, T. Tragoonrung, S. and Tangphatsornruang, S. 2013. Development and characterization of single-nucleotide polymorphism markers from 454 transcriptome sequences in oil palm (Elaeis guineensis). *Plant Breeding* 132:6:711-717.

Potvin, C. Bergeron, Y. and Simon, J. 1983. A numerical taxonomic study of selected Citrus species (Rutaceae) based on biochemical characters. *Syst.Bot.* 127-133.

Praloran, J.C. 1971. Les agrumes. G.-P. Maisonneuve et Larose, Paris.

PRINSEQ. 2012. PReprocessing and INformation of SEquence data: Easy and rapid quality control and data preprocessing; <http://prinseq.sourceforge.net/index.html>.

Pritchard Lab, S.U. 2014. Structure Software; <http://pritchardlab.stanford.edu/structure.html>.  
<http://pritchardlab.stanford.edu/structure.html>.

Pritchard, J.K. Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:2:945-959.

Rabelais, F. 1546. Tiers livre des faits et dits Héroïques du noble Pantagruel . Wechel, Chrétien, Paris.

Raghuvanshi, S.S. 1969. Cytological evidence bearing on evolution in Citrus. In: Chapman (ed.), Proc. First Int. Citrus Symp. (Univ. California, Riverside)1:207-214.

Ramadugu, C. Pfeil, B.E. Keremane, M.L. Lee, R.F. Maureira-Butler, I.J. and Roose, M.L. 2013. A six nuclear gene phylogeny of Citrus (Rutaceae) taking into account hybridization and lineage sorting. *PLoS One* 8:7:e68410.

Raman, H. Raman, R. Wood, R. and Martin, P. 2006. Repetitive Indel Markers within the ALMT1 Gene Conditioning Aluminium Tolerance in Wheat (*Triticum aestivum* L.). *Molecular Breeding* 18 (2):171-183.

Rao, N.K. Hanson, J. Dulloo, M.E. Ghosh, K. Nowell, A. and Iarinde, M. 2006. Manual of seed handling in genebanks. Bioversity International, Rome (Italy).

Rao, M.N. Soneji, J.R. Chen, C.X. Choi, Y.A. Huang, S. and Gmitter, F.G.,Jr. 2007. Identification of zygotic and nucellar seedlings from citrus rootstock candidates using RAPD. *Acta Horticulturae*738:255-260.

Rao, M.N. Soneji, J.R. Chen, C.X. Huang, S. and Gmitter, F.G.,Jr. 2008. Characterization of zygotic and nucellar seedlings from sour orange-like citrus rootstock candidates using RAPD and EST-SSR markers. *Tree Genetics and Genomes* 4:1:113-124.

Ravel, C. Praud, S. Murigneux, A. Canaguier, A. Sapet, F. Samson, D. Balfourier, F. Dufour, P. Chalhoub, B. Brunel, D. Beckert, M. and Charmet, G. 2006. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). *Genome* 49:9:1131-1139.

Rebour, H. 1966. Les agrumes : manuel de culture des citrus pour le Bassin méditerranéen. 5e Edition, Bayer, Paris.

Recupero, G.R. Russo, G. Recupero, S. and Di Vaio, C. 2010. The culture of lemon in Italy. VI Simposio Internacional Citrícola. Colima, Mexico November 2010. p.207-216.

Recupero, G.R. Russo, G. and Recupero, S. 2005. New Promising Citrus Triploid Hybrids Selected from Crosses between Monoembryonic Diploid Female and Tetraploid Male Parents. *HortScience* 40:3:516-520.

Reece, P.C. and Childs, J.F.L. 1962. Character differences among seedlings of the Persian lime. *Proc. Fla. State Hort. Soc.* 75:110-16.

Rickert, A.M. Kim, J.H. Meyer, S. Nagel, A. Ballvora, A. Oefner, P.J. and Gebhardt, C. 2003. First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. *Plant Biotechnology Journal* 1:6:399-410.

Rieseberg, L.H. 1996. Homology among RAPD fragments in interspecific comparisons. *Molecular Ecology* 5:99-105.

Rieseberg, L.H. Raymond, O. Rosenthal, D.M. Lai, Z. Livingstone, K. Nakazato, T. Durphy, J.L. Schwarzbach, A.E. Donovan, L.A. and Lexer, C. 2003. Major ecological transitions in annual sunflowers facilitated by hybridization. *Science* 301:1211-1216.

Rieseberg, L.H. and Soltis, D.E. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Trends in Plants* 5:65-84.

Risso, J.A. 1813. Essai sur l'histoire naturelle des oranges, bigaradiers, limettiers, cédratiers, limoniers ou citroniers, cultivés dans le Département des Alpes Maritimes. *Ann. Mus. Paris* 20:169-212.

Roose, M.L. Federici, C.T. Mu, L. Kwok, K. and Vu, C. 2009. Map-based ancestry of sweet orange and other citrus variety groups. Gentile A, Tribulato E. eds. Second International Citrus Biotechnology Symposium , 28 . Tremestieri Etneo, Italy: Emme Erre Grafica.

Roose, M.L. and Close, T.J. 2008. Genomics of citrus, a major fruit crop of tropical and subtropical regions. (Plant Genetics and Genomics: Crops and Models, Volume 1). *Genomics of tropical crop plants* 187-201.

Roose, M.L. Feng, D. Cheng, F.S. Tayyar, R.I. Federici, C.T. and Kupper, R.S. 2000. Mapping the *Citrus* genome. *Acta Horticulturae* 535:25-32.

Roose, M.L. and Williams, T.E. 2007. Mutation breeding. *Citrus genetics, breeding and biotechnology* 345-352.

Rosenblum, E.B. and Novembre, J. 2007. Ascertainment bias in spatially structured populations: a case study in the eastern fence lizard. *J.Hered.* 98:4:331-336.

Rostand, J. 1941. Les idées nouvelles de la génétique. P.U.F., Paris.

Rostoks, N. Mudie, S. Cardle, L. Russell, J. Ramsay, L. Booth, A. Svensson, J.T. Wanamaker, S. Walia, H. Rodriguez, E.M. Hedley, P.E. Liu, H. Morris, J. Close, T.J. Marshall, D.F. and Waugh, R. 2005. Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Molecular Genetics and Genomics* 274:1617-4615.

Rousseau-Gueutin, M. Gaston, A. Aïnouche, A. Aïnouche, M.L. Olbricht, K. Staudt, G. Richard, L. and Denoyes-Rothan, B. 2009. Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): new insights from phylogenetic analyses of low-copy nuclear genes. *Mol Phylogenet Evol* 51:3:515-530.

Royal Botanic Gardens, Kew and Missouri Botanical Garden. 2013. <http://www.theplantlist.org/>. 2014.

Rozas, J. 2009. DNA sequence polymorphism analysis using DnaSP. *Methods Mol.Biol.* 537:337-350.

Ruiz, C. and Asins, M.J. 2003. Comparison between *Poncirus* and *Citrus* genetic linkage maps. *Theor.Appl.Genet.* 106:5:826-836.

Rumphius, G.E. 1999 (1661). *The Ambonese Curiosity Cabinet*. Yale University Press, New Haven, CT.

Russo, G. Recupero, S. Puglisi, A. and Recupero, G. 2004. New triploid citrus hybrids by Italian genetic improvement. *Rivista di Frutticoltura e di Ortofloricoltura* 66:14-18.

Sachidanandam, R. Weissman, D. Schmidt, S.C. Kakol, J.M. Stein, L.D. Marth, G. Sherry, S. Mullikin, J.C. Mortimore, B.J. Willey, D.L. Hunt, S.E. Cole, C.G. Coggill, P.C. Rice, C.M. Ning, Z. Rogers, J. Bentley, D.R. Kwok, P.Y. Mardis, E.R. Yeh, R.T. Schultz, B. Cook, L. Davenport, R. Dante, M. Fulton, L. Hillier, L. Waterston, R.H. McPherson, J.D. Gilman, B. Schaffner, S. Van Etten, W. Reich, D. Higgins, J. Daly, M.J. Blumenstiel, B. Baldwin, J. Stange-Thomann, N. Zody, M.C. Linton, L. Lander,

- E.S. and Atshuler, D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928-933.
- Şahin-Çevik, M. and Moore, G.A. 2012. Quantitative trait loci analysis of morphological traits in Citrus. *Plant Biotechnology Reports* 6:1:47-57.
- Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* 4:4:406-425.
- Salette, J. 2000. Les Agrumes. *Compte Rendu de l'Académie d'Agriculture de France* 8:Séance du 13 décembre 2000.
- Sanger, F. and Coulson, A.R. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J.Mol.Biol.* 94:3:441-448.
- Sanger, F. Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* 74:12:5463-5467.
- Sankar, A.A. and Moore, G.A. 2001. Evaluation of inter-simple sequence repeat analysis for mapping in Citrus and extension of the genetic linkage map. *Theoretical & Applied Genetics* 102:206-214.
- Santos, M.G. Passos, O.S. Soares Filho, W.S. Girardi, E.A. Gesteira, A.S. Ferreira, C. and Fagoaga, C. 2013. Variability analysis of 'Persian' acid lime tree selections using agronomic and molecular markers. *Genet Mol Res* 12:4:4604-4614.
- Sardos, J. McKey, D. Duval, M.F. Malapa, R. Noyer, J.L. and Lebot, V. 2008. Evolution of cassava (*Manihot esculenta* Crantz) after recent introduction into a South Pacific Island system: the contribution of sex to the diversification of a clonally propagated crop. *Genome* 51:11:912-921.
- Savita Bhagat, A. Pati, P.K. Virk, G.S. and Nagpal, A. 2012. An efficient micropropagation protocol for Citrus jambhiri Lush. and assessment of clonal fidelity employing anatomical studies and RAPD markers. *In Vitro Cellular & Developmental Biology-Plant* 48:5:512-520.
- Scarano, M. Tusa, N. Abbate, L. Lucretti, S. Nardi, L. and Ferrante, S. 2003. Flow cytometry, SSR and modified AFLP markers for the identification of zygotic plantlets in backcrosses between 'Femminello' lemon cybrids (2n and 4n) and a diploid clone of 'Femminello' lemon (*Citrus limon* L. Burm. F.) tolerant to mal secco disease. *Plant Science* 164:6:1009-1017.
- Scarcelli, N. 2005. Structure et dynamique de la diversité d'une plante cultivée à multiplication végétative: le cas des ignames au Bénin (*Dioscorea* sp.). PhD thesis Montpellier II University, France.
- Scora, R.W. 1988. Biochemistry, taxonomy and evolution of modern cultivated citrus. *Proceedings of the Sixth International Citrus Congress*(1):277-289.
- Scora, R.W. 1975. On the history and origin of *Citrus*. *Bull. Torrey Bot. Club* 102:369-375.
- Scora, R.W. Kumamoto, J. Soost, R.K. and Nauer, E.M. 1982. Contribution to the origin of the grapefruit *Citrus paradisi* (Rutaceae). *Syst.Bot.* 7:170-177.
- Shen, Y.J. Jiang, H. Jin, J.P. Zhang, Z.B. Xi, B. He, Y.Y. Wang, G. Wang, C. Qian, L. Li, X. Yu, Q.B. Liu, H.J. Chen, D.H. Gao, J.H. Huang, H. Shi, T.L. and Yang, Z.N. 2004. Development of genome-wide DNA polymorphism database for map-based cloning of rice genes. *Plant Physiol.* 135:3:1198-1205.
- Shendure, J. Porreca, G.J. Reppas, N.B. Lin, X. McCutcheon, J.P. Rosenbaum, A.M. Wang, M.D. Zhang, K. Mitra, R.D. and Church, G.M. 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* 309:1728-1760.

- Shimada, T. Hirabayashi, T. Endo, T. Fujii, H. Kita, M. and Omura, M. 2005. Isolation and characterization of the somatic embryogenesis receptor-like kinase gene homologue (CitSERK1) from *Citrus unshiu* Marc. *Scientia Horticulturae* 103:2:233-238.
- Shimada, T. Fujii, H. Endo, T. Ueda, T. Sugiyama, A. Nakano, M. Kita, M. Yoshioka, T. Shimizu, T. Nesumi, H. Ikoma, Y. Moriguchi, T. and Omura, M. 2014. Construction of a citrus framework genetic map anchored by 708 gene-based markers. *Tree Genetics & Genomes* 1-13.
- Shimizu, T. Fujii, H. Kotoda, N. Yano, K. and Endo, T. 2011. Data mining of citrus expression sequence data sets and application for functional genomic study. *Acta Horticulturae (ISHS)* 892:29-36.
- Shimizu, T. Fujii, H. Nishikawa, F. Shimada, T. Kotoda, N. Yano, K. and Endo, T. 2009. Data mining of citrus sequence data sets to develop microarrays for expression and genomic analysis.
- Shimizu, T. Yoshioka, T. Nagasaki, H. Kaminuma, E. Toyoda, A. Fujiyama, A. and Nakamura, Y. 2012. Whole genome sequencing and mapping analysis for identifying polymorphism among 11 citrus varieties. XII International Citrus Congress - Valencia, Spain S03O03:62.
- Simsek, O. Aka Kacar, Y. Yesiloglu, T. and Ollitrault, P. 2011. Determination by SSCP markers of allelic diversity of candidate genes for tolerance to iron chlorosis in citrus germplasm. *Acta Horticulturae (ISHS)* 892:85-91.
- Singh, A. Sinha, P. Dubey, A. and Srivastav, M. 2011. Characterization of citrus germplasm using simple sequence repeat (SSR) markers. *The Indian Journal of Genetics and Plant Breeding* 71:4:349-355.
- Snoussi, H. Duval, M.F. Garcia-Lor, A. Belfalah, Z. Froelicher, Y. Risterucci, A.M. Perrier, X. Jacquemoud-Collet, J.P. Navarro, L. Harrabi, M. and Ollitrault, P. 2012. Assessment of the genetic diversity of the Tunisian citrus rootstock germplasm. *BMC Genet.* 13:16-2156-13-16.
- Sønstebo, J.H. Gielly, L. Brysting, A.K. Elven, R. Edwards, M. Haile, J. Willerslev, E. Coissac, E. Rioux, D. Sannier, J. Taberlet, P. and Brochmann, C. 2010. Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources* 10:6:1009-1018.
- Soost, R.K. and Cameron, J.W. 1975. Citrus. p. 507-540. In: J. Janick and J.N. Moore (eds.), *Advances in Fruit Breeding*, Purdue University Press, West Lafayette, Indiana.
- Soost, R.K. and Cameron, J.W. 1969. Tree and fruit characters of Citrus triploid from tetraploid by diploid crosses. *Hilgardia* 20:569-575.
- SouthGreen. 2014. <http://www.southgreen.fr/>.
- Stebbins, G. 1950. *Variation and evolution in plants*. Columbia University Press, New York.
- Stein, L.D. 2010. The case for cloud computing in genome informatics. *Genome Biol.* 11:5:207.
- Stevens, P.F. 2013. <http://www.mobot.org/MOBOT/research/APweb/>. 2014.
- Storey, R. and Walker, R.R. 1999. Citrus and salinity. *Scientia Horticulturae* 78:39-81.
- Stover, E. Stange, R.R. McCollum, T.G. Jaynes, J. Irey, M. and Mirkov, E. 2013. Screening Antimicrobial Peptides In Vitro for Use in Developing Transgenic Citrus Resistant to Huanglongbing and Citrus Canker. *Journal of the American Society for Horticultural Science* 138:2:142-148.

Swingle, W.T. 1943. The botany of citrus and its wild relatives of the orange family. p. 129-474. In: H.J. Webber and L.D. Batchelor (eds.), *The Citrus Industry*, Vol. I. Berkeley and Los Angeles : Univ. California Press.

Swingle, W.T. 1914. Citrus and related genera. p. 779-785. In: L.H. Bailey (ed.), *Standard Cyclopedia of Horticulture*, The Macmillan CO., New York.

Swingle, W.T. and Reece, P.C. 1967. The botany of Citrus and its wild relatives. p. 190-430. In: W. Reuther H.J. Webber and L.D. Batchelor (eds.), *The citrus industry*, Vol. 1. University of California Press, Berkeley, California, USA.

Tadeo, F.R. Cercós, M. Colmenero-Flores, J.M. Iglesias, D.J. Naranjo, M.A. Ríos, G. Carrera, E. Ruiz-Rivero, O. Lliso, I. Morillon, R. Ollitrault, P. and Talon, M. 2008. Molecular Physiology of Development and Quality of Citrus. p. 147-223. In: *Anonymous Advances in Botanical Research*, Vol. Volume 47. Academic Press.

Talon, M. and Gmitter, F.G.,Jr. 2008. Citrus genomics. *Int.J.Plant.Genomics* 2008:528361.

Tamura, K. Stecher, G. Peterson, D. Filipowski, A. and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol.Biol.Evol.* 30:12:2725-2729.

Tanaka, T. 1977. Fundamental discussion of Citrus classification. *Study in Citrologia*, Osaka 14:1-6.

Tanaka, T. 1969. Taxonomic problem of Citrus fruits in the Orient. *Bull. Univ. Osaka Pref. Ser. B.* 21:133-138.

Tanaka, T. 1961. *Citologia: Semi-centennial Commemoration Papers on Citrus Studies*. Citologia Supporting Foundation, Osaka.

Tanaka, T. 1954. Species problem in Citrus (Revisio Aurantiacearum IX). Japanese Society for Promotion of Science, Tokyo, Japan.

Tanaka, T. 1929. Remarks on citrus and citrus relatives in China. *Lingnan Science Journal* 7:337-348.

Tatum, J.H. Berry, R.E. and Hearn, C.J. 1974. Characterization of citrus cultivars and separation of nucellar and zygotic seedlings by thin layer chromatography. *Florida State Horticultural Society* 87:75-81.

Taudien, S. Groth, M. Huse, K. Petzold, A. Szafranski, K. Hampe, J. Rosenstiel, P. Schreiber, S. and Platzer, M. 2010. Haplotyping and copy number estimation of the highly polymorphic human beta-defensin locus on 8p23 by 454 amplicon sequencing. *BMC Genomics* 11:1:252.

Taylor, J.F. 2013. Implementation and accuracy of genomic selection. *Aquaculture*.

Terol, J. Carbonell, J. Alonso, R. Tadeo, F.R. Herrero, A. Ibáñez, V. Muñoz, J.V. López-García, A. Hueso, L. Colmenero-Flores, J.M. Conesa, A. Dopazo, J. and Talon, M. 2012. Sequencing of 150 citrus varieties: linking genotypes to phenotypes. XII International Citrus Congress - Valencia, Spain S03O02:61.

Terol, J. Ibáñez, V. Carbonell, J. Alonso, R. Estornell, L.H. Licciardello, C. Gut, I.G. Dopazo, J. and Talon, M. submitted. Involvement of a citrus meiotic recombination TTC-repeat motif in the formation of gross deletions generated by ionizing radiation and MULE activation. *BMC genomics*.

Terol, J. Conesa, A. Colmenero, J.M. Cercos, M. Tadeo, F. Agusti, J. Alos, E. Andres, F. Soler, G. Brumos, J. Iglesias, D.J. Gotz, S. Legaz, F. Argout, X. Courtois, B. Ollitrault, P. Dossat, C. Wincker, P. Morillon, R. and Talon, M. 2007. Analysis of 13000 unique Citrus clusters associated with fruit quality, production and salinity tolerance. *BMC Genomics* 8:31.



Terol, J. Naranjo, M.A. Ollitrault, P. and Talon, M. 2008. Development of genomic resources for *Citrus clementina*: characterization of three deep-coverage BAC libraries and analysis of 46,000 BAC end sequences. *BMC Genomics* 9:423.

Texeira, D.C. Ayres, J. Kitajima, E.W. Danet, L. Jagoueix-Eveillard, S. Saillard, C. and Bové, J.M. 2005. First Report of a Huanglongbing-Like Disease of Citrus in Sao Paulo State, Brazil and Association of a New *Liberibacter* Species, "*Candidatus Liberibacter americanus*", with the Disease. *Plant Disease* 89:1:107-107.

Thudi, M. Li, Y. Jackson, S.A. May, G.D. and Varshney, R.K. 2012. Current state-of-art of sequencing technologies for plant genomics research. *Brief Funct.Genomics* 11:1:3-11.

Tokunaga, T. Nii, M. Tsumura, T. and Yamao, M. 2005. Production of triploids and breeding seedless cultivar 'Tokushima 3X No.1' from tetraploid x diploid crosses in sudachi (*Citrus sodachi* Shirai). *Horticultural Research (Japan)* 4:11-15.

Tolkowsky, S. 1938. *Hesperides: a history of the culture and use of citrus fruits*. J. Bale, Sons & Curnow, Ltd., London.

Trabut, L. 1902. L'hybridation des citrus : une nouvelle tangerine, la clémentine. *Revue d'Horticulture* 74:232-234.

Trask, J.A. Malhi, R.S. Kanthaswamy, S. Johnson, J. Garnica, W.T. Malladi, V.S. and Smith, D.G. 2011. The effect of SNP discovery method and sample size on estimation of population genetic data for Chinese and Indian rhesus macaques (*Macaca mulatta*). *Primates* 52:2:129-138.

Tripolitsiotis, C. Nikoloudakis, N. Linos, A. and Hagidimitriou, M. 2013. Molecular Characterization and Analysis of the Greek Citrus Germplasm. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 41:2.

Troggio, M. Malacarne, G. Coppola, G. Segala, C. Cartwright, D.A. Pindo, M. Stefanini, M. Mank, R. Moroldo, M. Morgante, M. Grando, M.S. and Velasco, R. 2007. A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir bacterial artificial chromosome contigs. *Genetics* 176:4:2637-2650.

Tudela, D. and Primo-Millo, E. 1992. 1-Aminocyclopropane-1-Carboxylic Acid Transported from Roots to Shoots Promotes Leaf Abscission in Cleopatra Mandarin (*Citrus reshni* Hort. ex Tan.) Seedlings Rehydrated after Water Stress. *Plant Physiol.* 100:1:131-137.

United States Department of Agriculture Agricultural Research Service, Beltsville Area. 2013. GRIN Taxonomy for Plants; <http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?10679>; 2014.

USDA. 2014. (2009) Citrus Genome Database; <http://www.citrusgenomedb.org>. USDA NIFA Specialty Crop Research Initiative Program.

Uzun, A. Gulsen, O. Seday, U. Yesiloglu, T. Aka-Kacar, Y. and Tuzcu, O. 2011a. Investigation of genetic relationships among trifoliata oranges and their hybrid relatives based on ISSR markers. *Romanian Biotechnological Letters* 16:4:6431.

Uzun, A. Yesiloglu, T. Polat, I. Aka-Kacar, Y. Gulsen, O. Yildirim, B. Tuzcu, O. Tepe, S. Canan, I. and Anil, S. 2011b. Evaluation of Genetic Diversity in Lemons and Some of Their Relatives Based on SRAP and SSR Markers. *Plant Molecular Biology Reporter* 29:3:693-701.

Uzun, A. Yesiloglu, T. Aka-Kacar, Y. Tuzcu, O. and Gulsen, O. 2009. Genetic diversity and relationships within Citrus and related genera based on sequence related amplified polymorphism markers (SRAPs). *Scientia Horticulturae* 121:3:306-312.

Van Berloo, R. 2008. GGT 2.0: versatile software for visualization and analysis of genetic data. *J.Hered.* 99:2:232-236.

Van Os, H. Andrzejewski, S. Bakker, E. Barrena, I. Bryan, G.J. Caromel, B. Ghareeb, B. Isidore, E. de Jong, W. van Koert, P. Lefebvre, V. Milbourne, D. Ritter, E. van der Voort, J.N. Rousselle-Bourgeois, F. van Vliet, J. Waugh, R. Visser, R.G. Bakker, J. and van Eck, H.J. 2006. Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. *Genetics* 173:2:1075-1087.

Van Tassell, C.P. Smith, T.P. Matukumalli, L.K. Taylor, J.F. Schnabel, R.D. Lawley, C.T. Haudenschild, C.D. Moore, S.S. Warren, W.C. and Sonstegard, T.S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat.Methods* 5:3:247-252.

Vasemagi, A. Gross, R. Palm, D. Paaver, T. and Primmer, C.R. 2010. Discovery and application of insertion-deletion (INDEL) polymorphisms for QTL mapping of early life-history traits in Atlantic salmon. *BMC Genomics* 11:156-2164-11-156.

Vezzulli, S. Micheletti, D. Riaz, S. Pindo, M. Viola, R. This, P. Walker, M.A. Troggio, M. and Velasco, R. 2008. A SNP transferability survey within the genus *Vitis*. *BMC Plant.Biol.* 8:128-2229-8-128.

Viloria, Z. and Grosser, J.W. 2005. Acid citrus fruit improvement via interploid hybridization using allotetraploid somatic hybrid and autotetraploid breeding parents. *J.Am.Soc.Hort.Sci.* 130:3:392-402.

Volk, G. Bonnart, R. Shepherd, A. Krueger, R.R. and Lee, R. 2012. Cryopreservation of citrus for long-term conservation. Valencia, Spain.

Vos, P. Hogers, R. Bleeker, M. Reijans, M. Lee, T. Homes, M. Frijters, A. Pot, J. Peleman, J. Kuiper, M. and Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23 (21):4407-4414.

Wali, S. Munir, F. and Mahmood, T. 2013. Phylogenetic studies of selected Citrus species based on chloroplast gene, *rps14*. *International Journal of Agriculture and Biology* 15:2:357-361.

Wang, N. and Trivedi, P. 2013. Citrus Huanglongbing: a newly relevant disease presents unprecedented challenges. *Phytopathology* 103:7:652-665.

Wang, J. Chen, D. Lei, Y. Chang, J. Hao, B. Xing, F. Li, S. Xu, Q. Deng, X. and Chen, L. 2014. *Citrus sinensis* Annotation Project (CAP): A Comprehensive Database for Sweet Orange Genome. *PLoS ONE* 9:1:e87723.

Webber, H.J. Reuther, W. and Lawton, H.W. 1967. History and development of the Citrus industry. p. 1-39. In: W. Reuther and et al. (eds.), *The Citrus Industry*, Vol. 1. University of California, Division of Agricultural Sciences, Berkeley.

Webber, H.J. 1943. Cultivated varieties of citrus. p. 475-668. In: Anonymous *The Citrus Industry. History, World Distribution, Botany and Varieties*, Vol. 1. California.

Weising, K. and Gardner, R.C. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42:1:9-19.

Williams, T. and Roose, M. 2004. 'TDE2' Mandarin hybrid (Shasta Gold Mandarin), 'TDE3' Mandarin hybrid (Tahoe Gold Mandarin) and 'TDE4' Mandarin hybrid (Yosemite Gold Mandarin): Three New Mid

and Late-Season Triploid Seedless Mandarin Hybrids from California. *Proceedings International Society Citriculture* 1:394-398.

Wright, S. 1978. Variability Within and Among Natural Populations. In: *Anonymous Evolution and the Genetics of Populations*, Vol. 4. The University of Chicago Press, Chicago, IL.

Wu, G.A. Prochnik, S. Jenkins, J. Salse, J. Hellsten, U. Murat, F. Perrier, X. Ruiz, M. Scalabrin, S. Terol, J. Takita, M.A. Labadie, K. Poulain, J. Couloux, A. Jabbari, K. Cattonaro, F. Del Fabbro, C. Pinosio, S. Zuccolo, A. Chapman, J. Grimwood, J. Tadeo, F.R. Estornell, L.H. Munoz-Sanz, J.V. Ibanez, V. Herrero-Ortega, A. Aleza, P. Perez-Perez, J. Ramon, D. Brunel, D. Luro, F. Chen, C. Farmerie, W.G. Desany, B. Kodira, C. Mohiuddin, M. Harkins, T. Fredrikson, K. Burns, P. Lomsadze, A. Borodovsky, M. Reforgiato, G. Freitas-Astua, J. Quetier, F. Navarro, L. Roose, M. Wincker, P. Schmutz, J. Morgante, M. Machado, M.A. Talon, M. Jaillon, O. Ollitrault, P. Gmitter, F. and Rokhsar, D. 2014. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat.Biotechnol.* 32:7:656-662.

Wu, Z. Raven, P.H. and Hong, D. 1994. *Flora of China* (English edition). Science Press, Beijing.

Xiao, J. Chen, J. Zhang, H. Xu, H. Wang, H. and Xie, M. 2011. Gene expression profiling in response to drought stress in citrus leaves by cDNA-AFLP. *Acta* 38:3:417-424.

Xing, C. Schumacher, F.R. Xing, G. Lu, Q. Wang, T. and Elston, R.C. 2005. Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. *BMC Genet.* 6 Suppl 1:S29.

Xu, Q. Chen, L.L. Ruan, X. Chen, D. Zhu, A. Chen, C. Bertrand, D. Jiao, W.B. Hao, B.H. Lyon, M.P. Chen, J. Gao, S. Xing, F. Lan, H. Chang, J.W. Ge, X. Lei, Y. Hu, Q. Miao, Y. Wang, L. Xiao, S. Biswas, M.K. Zeng, W. Guo, F. Cao, H. Yang, X. Xu, X.W. Cheng, Y.J. Xu, J. Liu, J.H. Luo, O.J. Tang, Z. Guo, W.W. Kuang, H. Zhang, H.Y. Roose, M.L. Nagarajan, N. Deng, X.X. and Ruan, Y. 2013. The draft genome of sweet orange (*Citrus sinensis*). *Nat.Genet.* 45:1:59-66.

Yakushiji, H. Morinaga, K. and Nonami, H.,. 1998. Sugar accumulation and partitioning in Satsuma mandarin tree tissues and fruit in response to drought stress. *Journal Of The American Society For Horticultural Science* 123:(4):719-726.

Yamamoto, M. Kobayashi, S. Nakamura, Y. and Yamada, Y. 1993. Phylogenetic relationships of citrus revealed by RFLP analysis of mitochondrial and chloroplast DNA. *Japanese Journal of Breeding* 43:(3):355-365.

Yamamoto, M. Kouno, R. Nakagawa, T. Usui, T. Kubo, T. and Tominaga, S. 2011. Isozyme and DNA analyses of local Citrus germplasm on Amami Islands. *Journal of the Japanese Society for Horticultural Science* 80:268-273.

Yamamoto, M. Abkenar, A.A. Matsumoto, R. Nesumi, H. Yoshida, T. Kuniga, T. Kubo, T. and Tominaga, S. 2007. CMA banding patterns of chromosomes in major Citrus species. *JOURNAL-JAPANESE SOCIETY FOR HORTICULTURAL SCIENCE* 76:1:36.

Yamamoto, M. Tsuchimochi, Y. Ninomiya, T. Koga, T. Kitajima, A. Yamasaki, A. Inafuku-Teramoto, S. Yang, X. Yang, X. Zhong, G. Nasril, N. Kubo, T. and Tominaga, S. 2013. Diversity of Chloroplast DNA in Various Mandarins (*Citrus* spp.) and Other Citrus Demonstrated by CAPS Analysis. *Journal of the Japanese Society for Horticultural Science* 82:2:106-113.

Yang, Y. YueZhi, P. Xun, G. and MouTian, F. 2010. Genetic variation in the endangered Rutaceae species *Citrus hongheensis* based on ISSR fingerprint. *Genetic Resources and Crop Evolution* 57:1239-1248.

- Ye JunLi Zhu AnDan Tao NengGuo Xu Qiang Xu Juan and Deng, X. 2010. Comprehensive analysis of expressed sequence tags from the pulp of the red mutant 'Cara Cara' navel orange (*Citrus sinensis* Osbeck). *Journal of Integrative Plant Biology* 52:10:856-867.
- Yildiz, E. Kaplankiran, M. Demirkeser, T. Uzun, A. and Toplu, C. 2013. Identification of Zygotic and Nucellar Individuals Produced from Several Citrus Crosses Using SSRs Markers. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 41:2.
- Yuhui, H. Makusu, T. Ralph, H.S. Fan, B. Masateru, T. and Tomoji, K. 2013. Thermophoretic Manipulation of DNA Translocation through Nanopores. *ACS Nano* 7 (1):538-546.
- Zabeau, M. and Vos, P. 1993. Selective restriction fragment amplification: a general method for DNA fingerprinting. *European Patent Office* 93/13:534-858.
- Zaragoza, S. 2007. Aproximación a la historia de los cítricos. Origen, dispersión y evolución de su uso y cultivo.
- Zaragoza, S. 1991. Historia de la Naranja. Levante, Valencia.
- Zhang, Q. Zhao, H. Lin, L. and Wen, H. 2011. Comparative studies on the quality of different cultivars of *Citrus grandis* var. *tamentosa* Hart. *Medicinal Plant* 2:6:5-8.
- Zhou, C. Qi, Y. You, X. Yang, L. Guo, P. Ye, X. Zhou, X. Ke, F. and Chen, L. 2013. Leaf cDNA-AFLP analysis of two citrus species differing in manganese tolerance in response to long-term manganese-toxicity. *BMC Genomics* 14:1:621.
- Zhu, S. Wu, B. Ma, Y. Chen, J. and Zhong, G. 2013. Obtaining citrus hybrids by in vitro culture of embryos from mature seeds and early identification of hybrid seedlings by allele-specific PCR. *Scientia Horticulturae* 161:300-305.
- Zhu, Y.L. Song, Q.J. Hyten, D.L. Van Tassell, C.P. Matukumalli, L.K. Grimm, D.R. Hyatt, S.M. Fickus, E.W. Young, N.D. and Cregan, P.B. 2003. Single-nucleotide polymorphisms in soybean. *Genetics* 163:3:1123-1134.

---

# Annexes

---

## Annexe I

---

### Articles additionnels et contributions à Congrès en relation avec la thèse

- Annexe I.1. Multilocus Haplotyping by Parallel Sequencing to Decipher the Interspecific Mosaic Genome Structure of Cultivated *Citrus* (Curk *et al.*; in press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012)
- Annexe I.2. Multilocus SNPs analysis allows phylogenetic assignation of DNA fragments to decipher the intraspecific mosaic genome structure of cultivated Citrus (Curk *et al.*; 2011; Plant Genome Evolution Congress 2011, 4 - 6 Sep 2011, P2.21)
- Annexe I.3. Two-step PCR targeted amplicon library for next generation haplotyping; an example in *Citrus* (Curk *et al.*)
- Annexe I.4. New Insights on Limes and Lemons Origin from Nuclear and Cytoplasmic Markers Genotyping and Targeted Nuclear Gene Sequencing (Curk *et al.*; in press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012 and poster)
- Annexe I.5. Comparative values of SSRs, SNPs and InDels for citrus genetic diversity analysis (Ollitrault *et al.*; in press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012; S02P05)
- Annexe I.6. A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species (Garcia-Lor *et al.* 2013; Annals of Botany; 111 (1): 1-19.; doi: 10.1093/aob/mcs227; First published online: October 26, 2012)

## Annexe I.1

**Multilocus Haplotyping by Parallel Sequencing to Decipher the Interspecific Mosaic Genome Structure of Cultivated *Citrus***

Franck Curk<sup>1,2</sup>, Gema Ancillo<sup>2</sup>, Andres Garcia-Lor<sup>2</sup>, François Luro<sup>1</sup>, Xavier Perrier<sup>3</sup>, Jean-Pierre Jacquemoud-Collet<sup>3</sup>, Luis Navarro<sup>2\*</sup> and Patrick Ollitrault<sup>2,3\*</sup>

<sup>1</sup>UMR AGAP, Institut National de la Recherche Agronomique (INRA), Centre INRA de Corse, San Giuliano, France; <sup>2</sup>Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, Valencia, Spain; <sup>3</sup>UMR AGAP, Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France.

**Abstract**

The most important economic *Citrus* species originated from natural interspecific hybridization between four ancestral taxa (*C. reticulata*, *C. maxima*, *C. medica* and *C. micrantha*) with limited further interspecific recombination due to apomixis and vegetative propagation. Such reticulate evolution coupled with vegetative propagation results in genomes that are mosaics of large chromosome fragments of the basic taxa, in frequent interspecific heterozygosity. Breeding of these species is hampered by their complex heterozygous genomic structures. Haplotyping of multiple gene fragments along the genome should be a powerful approach to resolve the evolutionary history of the gene pools, to reveal the admixture genomic structure of current species and to develop innovative breeding schemes. We have analysed the efficiency of parallel sequencing with 454 methodology to decipher the hybrid structure of modern citrus species and cultivars along chromosome 2. 454 amplicon libraries were established with the fluidigm array system for 48 genotypes and 16 gene fragments of chromosome 2. Haplotypes were established from the reads of each accession and phylogenetic analyses were performed from the haplotypic data of each gene fragment. The length of 454 reads and the level of differentiation between the ancestral taxa of modern citrus allowed efficient haplotype phylogenetic assignments for 12 of the 16 gene fragments. The analysis of the mixed genomic structure of modern species and cultivars (i) revealed *C. maxima* introgressions in modern mandarins; (ii) was consistent with previous hypothesis regarding the origin of secondary species; and (iii) provided a new picture of the evolution of chromosome 2. Perspectives to rebuild the main secondary species from the basic taxa are discussed.

**Keywords:** Phylogeny, haplotype, evolution, SNP, NGS, genome admixture

Acta Horticulturae, in press (12 pages)

# Multilocus SNPs Analysis Allows Phylogenetic Assignment of DNA Fragments to Decipher the Interspecific Mosaic Genome Structure of Cultivated Citrus



Curk F. <sup>(1,2)</sup>, Ancillo G. <sup>(2)</sup>, Garcia-Lor A. <sup>(2)</sup>, Luro F. <sup>(1)</sup>, Navarro L. <sup>(2)</sup> and Ollitrault P. <sup>(3,2)</sup>  
 (1): INRA, (French National Institute for Agricultural Research) UR1103 Génétique et Ecophysiologie de la Qualité des Agrumes, F-20230 San Giuliano, France;  
 (2): IMA, (Valencian Institute for Agricultural Research), Centro de Protección Vegetal y Biotecnología, Ctra. Moncada-Náquera km 4.5, E-46113 Moncada, Valencia, Spain;  
 (3): CIRAD, (International Center of Agricultural Research for Development), Département BUD, 1001 Ngazun, Avenue Agropolis, TA-A75/02Z, F-34398 Montpellier Cedex 5, France;



## Introduction

All current studies seem to support the theory that four basic taxa (*C. medica*, *C. maxima*, *C. reticulata* and *C. micrantha*) have generated all cultivated Citrus species (Figure 1). It is supposed that the genomes of most of the modern Citrus cultivars, vegetatively propagated, are interspecific mosaic of large DNA fragments issued from a limited number of inter-specific meiosis. In the present work we analyzed how multilocus study of closely linked SNPs allows a phylogenetic assignment of DNA fragments of the main cultivated species.

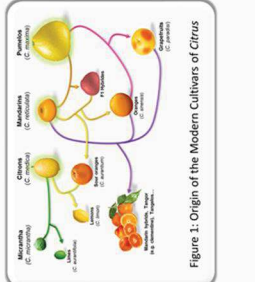


Figure 1: Origin of the Modern Cultivars of Citrus

## Genotype Assignment (Sanger)

The average SNP rate was about 20 SNPs/Kb. Citrus *reticulata* was the most polymorphic basic taxa with an average of 4.2 SNPs/Kb. For each gene an ACP was performed from SNP data to analyze the genetic organization (e.g. Figure 3). It was assumed that intermediary position between two species cluster corresponded to interspecific heterozygosity

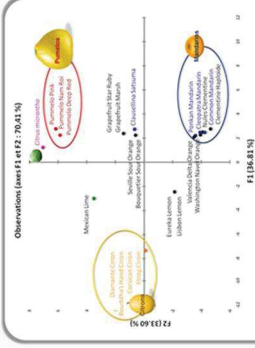


Figure 3: ACP, Phylogenetic Assignment of the Sequenced Fragments of the Locus 5.5

## Haplotype Assignment (454)

Consensus haplotype sequences were successfully obtained from 454 sequencing with genotype sequence in total agreement with Sanger control. Each haplotype sequence of Clementine was univocally assigned to one of the haplotype clusters of the basic taxa. As an example, we can see with the Figure 5 (NJ tree of the locus 5.5) that both haplotypes belong to the Mandarin group.

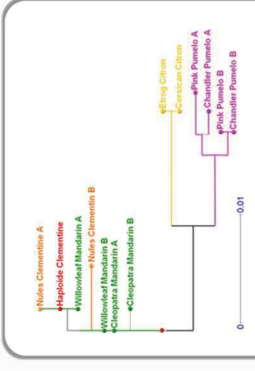


Figure 5: Neighbor Joining Tree of Locus 5.5 Haplotypes

## Material and Methods

Genomic fragments of 25 genes dispersed in the different chromosomes covering more than 12.5 Kb were amplified by PCR and sequenced [Sanger] for 24 accessions representative of 10 species of Citrus. Moreover we checked the potential of parallel pyrosequencing (454 Roche) for direct multilocus haplotyping of heterozygous genotypes. Amplified fragments from 7 genes in 8 genotypes of Citrus were obtained by using an original new method based on universal primers (Figure 2).

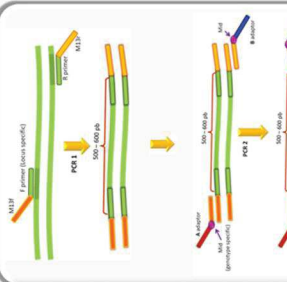


Figure 2: Fragment Amplification Using Universal M13 Primers

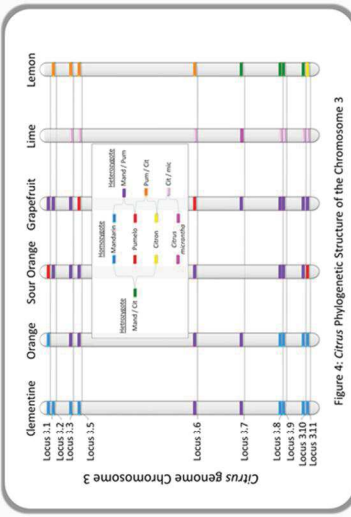


Figure 4: Citrus Phylogenetic Structure of the Chromosome 3

For each amplified fragment, the polymorphism was enough for multilocus differentiation of the basic species and assignment of a phylogenetic origin for the secondary species. A preliminary reconstruction of phylogenetic structure of the chromosome 3 is proposed for clementine, sweet orange, sour-orange, grapefruit, lime and lemon (Figure 4). However only haplotype analysis would allow definitive assignment. A Parallel pyrosequencing (454 Roche) for multilocus haplotyping of heterozygous genotypes has been done...

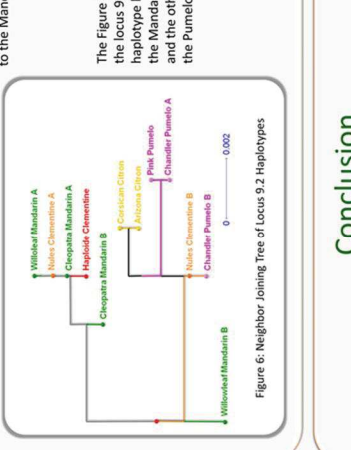


Figure 6: NJ tree of the locus 9.2

## Conclusion

Phylogenetic origin of specific DNA fragments can be assigned from multilocus analysis of closely linked SNPs. Multilocus haplotyping by parallel sequencing of individual DNA molecule will be a very powerful tool to decipher the interspecific mosaic genome structure of cultivated Citrus.



### Annexe I.3

## Two-step PCR targeted amplicon library for next generation haplotyping; an example in *Citrus*

Franck Curk<sup>a,b</sup>, Ancillo Gema<sup>b</sup>, Luis Navarro<sup>b</sup>, Patrick Ollitrault<sup>c,b</sup>

<sup>a</sup> UMR AGAP, Institut National de la Recherche Agronomique (Inra), Centre Inra de Corse, F-20230 San Giuliano, France; [curk@corse.inra.fr](mailto:curk@corse.inra.fr) (F. Curk)

<sup>b</sup> Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada (Valencia), Spain; [ancillo@ivia.es](mailto:ancillo@ivia.es), [lnavarro@ivia.es](mailto:lnavarro@ivia.es)

<sup>c</sup> UMR AGAP, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), TA A-108/02, 34398 Montpellier, Cedex 5, France; [ollitrault@cirad.fr](mailto:ollitrault@cirad.fr)

### Abstract

Phylogenetic of gene pools resulting from reticulate evolution, such as the *Citrus* genus, should be greatly improved by haplotyping of multiple gene fragments covering the genome in large germplasm samples. WGS remain costly for such analysis and haplotyping from short reads is a challenge for heterozygous genotypes. Targeted haplotyping of long enough DNA fragments by NGS should allow such multilocus and multitaxa haplotyping in homolog and ortholog genes. We have tested in *Citrus* the efficiency of a two-step PCR approach with the aim of drastically reducing the cost for establishing a multi-genotypes tagged library of amplicons and analysing the potential of 454 sequencing for haplotyping. The two-step PCR method was efficient enough to establish tagged libraries of amplicons and 454 sequencing provided genotyping data identical to Sanger ones. The length of 454 reads and level of differentiation between the ancestral taxa of modern citrus allowed efficient phylogenetic assignments of most haplotypes.

### Key words

Two-step PCR, haplotyping, phylogenetic assignments, SNP, NGS, Citrus.

## 1. Introduction

Reticulate evolution results in incongruence with gene phylogenies and makes difficult the establishment of phylogenetic trees. Haplotyping of multiple gene fragments along the whole genome should be a powerful approach to reveal the admixture of genomic structure of actual species and to resolve the evolutionary history of such gene pools (Linder and Rieseberg, 2004). Traditional Sanger sequencing after cloning is too much time consuming and expensive for efficient application in phylogeography and phylogenetic studies in large populations.

In the last years, massively parallel sequencing enabled rapid and relatively inexpensive DNA sequence data production and therefore facilitated genome-wide sequence variant discovery. Both Roche 454 pyrosequencing (Barbazuk *et al.*, 2007; Maughan and Redfield, 2009) and Illumina Genome Analyzer (Hillier *et al.*, 2008; Van Tassel *et al.*, 2008) methodology have evolved rapidly providing longer reads and impressively decreasing cost. However, despite this evolution WGS in large populations remains expensive and requires heavy bioinformatics treatments. McCormack *et al.* (McCormack *et al.*, 2012) reviewed the actual promise and challenge for the application of next generation sequencing to phylogeography and phylogenetics. One major challenge is to reduce the complexity of the genome and to generate and manage data of orthologous sequences for many individuals. Several approaches have been currently developed including reduction of genome complexity by restriction enzyme approaches (Baird *et al.*, 2008; Davey *et al.*, 2011; Elshire *et al.*, 2011; Peterson *et al.*, 2012). However when orthologous fragment sequences are researched these approaches appear even more adapted for narrow diversity scale and studies of progenies. When larger gene pool diversity is concerned, targeted capture (Crosby and Criddle, 2007; Saintenac *et al.*, 2011; Carneiro *et al.*, 2012) or targeted amplicon sequencing (Binladen *et al.*, 2007; Meyer *et al.*, 2007; Meyer *et al.*, 2008; Bybee *et al.*, 2011a; Griffin *et al.*, 2011) appear more adapted. With such approaches, a decisive advantage for haplotyping in heterozygous structure is that sequencing data come from single DNA molecule without cloning requirement. Therefore, it can be expected that with methods allowing long enough reads (over 500 bp), such as the actual 454 Titanium, it should be possible to establish multilocus haplotypes that should be phylogenetically significant when working at a sufficiently differentiated level.

Other challenge for cost effective NGS application to phylogeography and phylogenetics is that many individuals can be combined in the same sequencing run to divide the run costs among many samples (Glenn, 2011). It implies the use of short identifying DNA sequences (“tags” or “MIDs”) that are incorporated into the DNA fragments either by PCR (Binladen *et al.*, 2007) or ligation (Meyer *et al.*, 2008). Sequences are later sorted using bioinformatics tools. Such parallel tagged sequencing is well adapted for small to medium-sized projects with few loci that amplify well across individuals (Griffin *et al.*, 2011) and has been successfully applied to whole mitochondrial sequencing (Chan *et al.*, 2010; Morin *et al.*, 2010; Gunnarsdottir *et al.*, 2011) and whole chloroplast sequencing (Parks *et al.*, 2009). NGS also requires platform-specific, proprietary adaptor sequences to be incorporated into DNA fragments and this step of library preparation is often performed in conjunction with tagging. These libraries can be expensive if proprietary kits are used; even more expensive than the sequencing itself (Glenn, 2011). In the framework of the targeted amplicon approach, with

the objective of parallel sequencing of multiple targeted sequences in many individuals, the available proprietary Roche amplicon technology imply to generate a lot of primer pairs for establishing tagged sequencing libraries. Indeed if  $n$  genome fragments are targeted for  $m$  individuals,  $n \times m$  primer pairs should be necessary.

The Access Array System (Fluidigm, San Francisco, CA) was implemented to simultaneously prepare tagged amplicons of 48 (or 96) samples for 48 (or 96) targeted sequences (Moonsamy *et al.*, 2013). However it is relatively costly and its format would not be adapted for specific studies. Moreover in previous experiment with Citrus (Curk *et al.*, accepted) a very high heterogeneity of number of reads was observed for the different amplicons illustrated the limit of this approach when working in non-model species with a primer set not fully optimized for the method. As argued by Bybee *et al.* (Bybee *et al.*, 2011a), this approach also does not offer researchers full control of PCR protocols and reagents, an essential part of doing any phylogenetics project, especially at higher taxonomic levels where differential PCR efficiency can be expected between individuals.

Therefore, the objectives of the present work were:

(i) to implement a “homemade” method to establish 454 sequencing amplicon library adapted for multiple DNA targets and individuals parallel sequencing. It is based on a two-step PCR amplification similar to the one proposed by Bybee *et al.* (Bybee *et al.*, 2011a) using a first PCR round with primers specific for the targeted sequences combined with M13 universal primer and a second run with primers combining the M13 universal sequences, the ‘tag’ specific for each individual and the proprietary adaptors.

(ii) to analyse the potential of the 454 amplicon sequencing method for efficient targeted parallel haplotyping to decipher complex interspecific genomic structure resulting from reticulate evolution.

*Citrus* genus was chosen as example of reticulate gene pool. Indeed, the modern citrus cultivars are considered to result from interspecific hybridization between four ancestral taxa (*C. reticulata*, *C. maxima*, *C. medica* and *C. micrantha*) with limited further inter-specific hybridization due to apomixis and vegetative propagation (Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2013). Among the four ancestral taxa, the first three ones have generated most of the economically important secondary species (*C. sinensis*, sweet orange; *C. aurantium*, sour orange; *C. paradisi*, grapefruit; *C. limon*, lemon). 454 library was established by the two step PCR method applied to seven gene fragments for height citrus accessions. Haplotypes were established from 454 reads and the corresponding genotype compared with Sanger sequencing of the two step PCR products.

## 2. Material and methods

### 2.1. Plant material

Leaf material, from eight accessions of *Citrus* genus was collected from the IVIA Citrus Germplasm Bank of pathogen-free plants (Valencia, Spain; accessions with IVIA identification number). We adopted the Swingle and Reece (Swingle and Reece, 1967) botanical classification for scientific names (table 1). Two representatives of the three main

ancestral taxa (*C. reticulata*, *C. maxima* and *C. medica*) were selected as well as the haploid clementine used by the International Citrus Genome consortium to establish the reference citrus whole genome sequence available at <http://www.phytozome.net/clementine.php> and the diploid ‘Clemenules’ clementine, a natural hybrid between ‘Willowleaf’ mandarin and sweet orange (Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b).

**Table 1.** Varieties by common horticultural group and species (classification of Swingle and Reece 1967)

Common horticultural group name	Species	Common names	Identification number
Pummelos	<i>Citrus maxima</i> (Burm.) Merr.	‘Chandler’ pummelo	IVIA207
	<i>Citrus maxima</i> (Burm.) Merr.	‘Pink’ pummelo	IVIA275
Mandarins	<i>Citrus reticulata</i> Blanco	‘Willowleaf’ mandarin	IVIA154
	<i>Citrus reticulata</i> Blanco	‘Cleopatra’ mandarin	IVIA385
Citrons	<i>Citrus medica</i> L.	‘Corsican’ citron	IVIA567
	<i>Citrus medica</i> L.	‘Etrog’ citron	IVIA169
Clementines	<i>Citrus reticulata</i> Blanco	‘Clemenules’ clementine	IVIA22
	<i>Citrus reticulata</i> Blanco	‘Clemenules’ haploid clementine	IVIA638

## 2.2. DNA extraction

High molecular weight genomic DNA was extracted from leaf samples using the DNeasy Plant Mini Kit (Qiagen S.A.; Madrid, Spain) according to the manufacturer’s instructions.

## 2.3. Target genomic fragment selection

Seven clementine EST sequences, previously used by Luro *et al.* (Luro *et al.*, 2008) to develop SSR markers and located by Ollitrault *et al.* (Ollitrault *et al.*, 2012a) in four linkage groups of the clementine genetic map, were selected. [Supplementary information – Table A.1].

## 2.4. Amplicon library preparation and sequencing

454 parallel sequencing is done using a mixture of all the amplicons for all the genotypes, so it requires a different “tag” of DNA for each genotype. We used a set of Multiplex genotype Identifiers, or MIDs, defined by Roche:

([http://www.liv.ac.uk/media/livacuk/centreforgenomicresearch/The\\_GS\\_FLX\\_Titanium\\_Chemistry\\_Extended\\_MID\\_Set.pdf](http://www.liv.ac.uk/media/livacuk/centreforgenomicresearch/The_GS_FLX_Titanium_Chemistry_Extended_MID_Set.pdf)).

The sequences of used MID for the eight genotypes and Titanium primers are given in [Supplementary information – Table A.2].

454 sequencing technique requires amplicon primers containing a directional GS FLX Titanium primer sequence (which includes a four base library “key” sequence) at the 5-prime

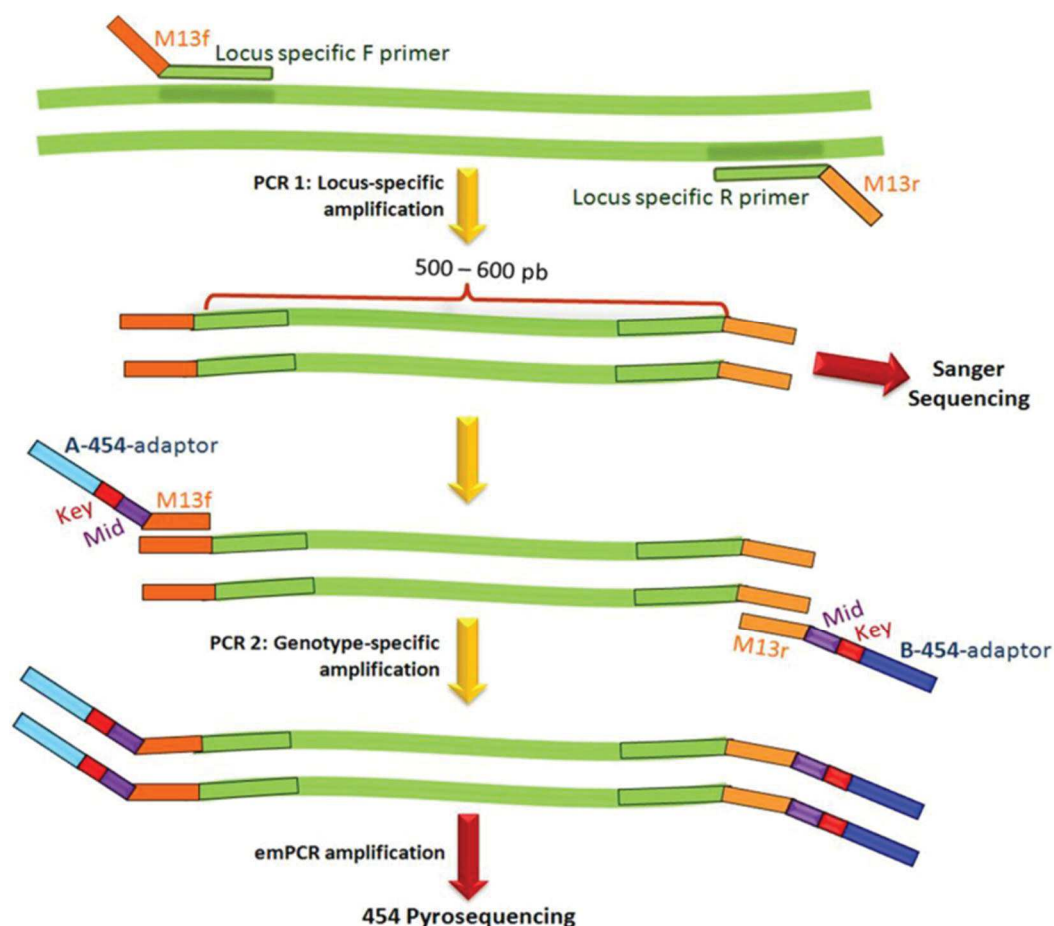
portion of the oligonucleotide in addition to the gen specific sequence at the 3-prime end. Multiplex genotype Identifier (MID) sequence is added between the primer A (or B) and gene specific sequence to allow automated software identification of samples after pooling and sequencing ([http://www.igsb.org/uploads/pdf/TCB-09013\\_AmpliconFusionPrimerDesignGuidelines.pdf](http://www.igsb.org/uploads/pdf/TCB-09013_AmpliconFusionPrimerDesignGuidelines.pdf)).

With this standard library preparation method, it is necessary to generate  $n \times m$  primers where  $n$  is the number of targeted sequences and  $m$  the number of individuals to be studied. For example if 20 sequences are targeted for 48 Individuals, 960 primer pairs would be necessary. Such method becomes rapidly highly costly if numerous individual and targeted sequences are concerned. Therefore we have developed a two-step PCR “tagged library preparation method to limit the number of necessary primers. The first PCR run is specific to the targeted fragments while the second one allows adding genotype specific tag and adaptors for the 454 platform. For the first PCR (PCR1, Figure 1) a M13 extension was added in 5' to the primer specific for each tagged DNA fragment and the amplification was performed with individualized genotypes. After purification, a second amplification has been performed (PCR2, Figure 1) genotype by genotype, using as primers the Titanium adaptor, the key, Multiplex Identifiers (MID) of each genotype and the same M13 extension (in 3') for the PCR1 (Figure 1). With such approach only  $n+m$  primer pairs are needed (for example 68 in case of 20 targeted fragments and 48 individuals).

The specific primers for the targeted genome fragments (selected from the seven selected clementine EST sequences) were designed with Primer3Plus website (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/>) to generate 500 to 600 bp amplicons [**Supplementary information - Table A.1**].

Seven universal sequencing forward and reverse primers [**Supplementary information –Table A.3**] associated to the specific targeted genome fragment primers previously designed, were tested with the Oligo Analyser 3.1 of IDT (Integrated DNA Technologies <http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/>) in order to discard any primers with hairpin, self-dimer or hetero-dimer problems. M13 (-40) forward and reverse primers were finally been selected [**Supplementary information – Table A.3**].

Amplicons of the 7 gene fragments were produced for the eight accessions. The locus specific PCR amplifications (PCR1, Figure 1) were performed using a Mastercycler Ep Gradient S thermocycler (Eppendorf) in a final volume of 20  $\mu$ l including 13.6  $\mu$ l of dH<sub>2</sub>O, 2  $\mu$ l of PCR buffer (10x), 1  $\mu$ l of magnesium chloride (50 mM), 0.4  $\mu$ l of dNTPs (10 mM), 0.2  $\mu$ l of Forwards Primer (20  $\mu$ M) and 0.2  $\mu$ l of Reverse Primer (20  $\mu$ M), 2  $\mu$ l of genomic DNA (20 ng/ $\mu$ l) and 0.6  $\mu$ l of 1U/ $\mu$ l Taq DNA polymerase. The thermocycling conditions included a 4 min denaturing step at 94°C for 1 cycle; 30 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 50 s; and a final elongation cycle of 72°C for 7 min. The PCR products were separated on a 1% agarose gel and scored visually for the presence of good sized amplicons.



**Figure 1.** Two step PCR for tagged amplicon library preparation for 454 pyrosequencing

With the objective to compare 454 and Sanger sequencing data, aliquots of amplicons obtained from PCR 1 (Figure 1) were sequenced using the Sanger method from the 5' end and 3' end using fluorescently labelled dideoxynucleotides (Big Dye Terminator Cycle Sequencing Kit v3.1).

The Genotype specific amplification PCR (PCR 2 Figure 1) was performed with 1 µl of PCR 1 purified product. The PCR 2 consisted of 16 µl of dH<sub>2</sub>O, 1.5 µl of PCR buffer (10x), 0.6 µl of magnesium chloride (50 mM), 0.2 µl of dNTPs (25 mM), 0.2 µl of Forwards Primer (25 µM) and 0.2 µl of Reverse Primer (25 µM), 1 µl of PCR 1 purified product (1 ng/µl) and 0.6 µl of 1U/µl Taq DNA polymerase. The thermocycling conditions included a 4 min denaturing step at 94°C for 1 cycle; 30 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 55 s; and a final elongation cycle of 72°C for 7 min. The PCR products were separated on a 1% agarose gel and scored visually for the presence of good sized amplicons. The spliced amplicons of this second PCR were purified with a Qiagen QIAquick Gel Extraction Kit in order to isolate the 500-600 bp inserts and avoid presence of Titanium primers, primer dimers, hairpins or unspecific products susceptible to reduce the number of 454 useful reads.

The concentration of all samples was measured using a Biorad fluorometer VersaFluor and each sample adjusted to the same concentration. Sample were then mixed to perform a sequencing run on a GS FLX Titanium system from Roche 454.

## 2.5. Sequence data analysis for SNP calling

### 2.5.1. Sanger

For each targeted fragment of each accession a consensus sequence was established after forward and reverse sequence alignment using BioEdit:

<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>.

Heterozygosity or homozygosity of all genotypes was verified visually in the chromatogram for all SNP positions. Each nucleotide polymorphism was coded according to the International Union of Pure and Applied Chemistry (IUPAC). The sequences of each genotype for the same fragment were aligned and compared with Bioedit software for SNPs mining.

### 2.5.2. 454

454 pyrosequencing reads were automatically identified and sorted by MID and specific gene primers using SFF Tool commands of Newbler software (Roche's "GS de novo assembler"):

[http://bioinfo.genotoul.fr/fileadmin/Documents/GS-FLX-System-Software-Manual-v2.5.3\\_PartC-Assembler-Mapper-SFF-Tools.pdf](http://bioinfo.genotoul.fr/fileadmin/Documents/GS-FLX-System-Software-Manual-v2.5.3_PartC-Assembler-Mapper-SFF-Tools.pdf).)

For each variety and for each sequenced gene, 454 pyrosequencing reads were aligned using SeqMan NGen software version 7.0 (<http://www.dnastar.com>). From this alignment one (for homozygote sequences) or two (for heterozygote sequences) consensus sequences were generated. Each consensus sequence represented one haplotype.

### 2.5.3. Genetic analysis of the SNP data

Unbiased expected heterozygosity, observed heterozygosity, fixation index value ( $F_w$ ; (Wright, 1978) and the  $F_{st}$  parameter ( $F_{st}$ ) were calculated using GENETIX v. 4.03 software (<http://kimura.univ-montp2.fr/genetix/intro.htm>) (Belkhir *et al.*, 1996-2004).

Haplotype and genotypic phylogenetic relationships for each fragment were studied using MEGA version 6 (Tamura *et al.*, 2013). Phenograms were created using the UPGMA tree option test using 1,000 Bootstraps.

Neighbour-Joining (NJ) analysis, based on the SNP data of all concatenated fragment, was done using DARwin software (Perrier and Jacquemoud-Collet, 2006), using the simple matching dissimilarity index. This index was also used to infer the intra and inter taxa average differentiation. For NJA, indel data were included and treated as missing data and pairwise site deletion with a minimal proportion of valid sites required for each unit pair of 80%. The robustness of branches was tested using 500 Bootstraps.

## 3. Results

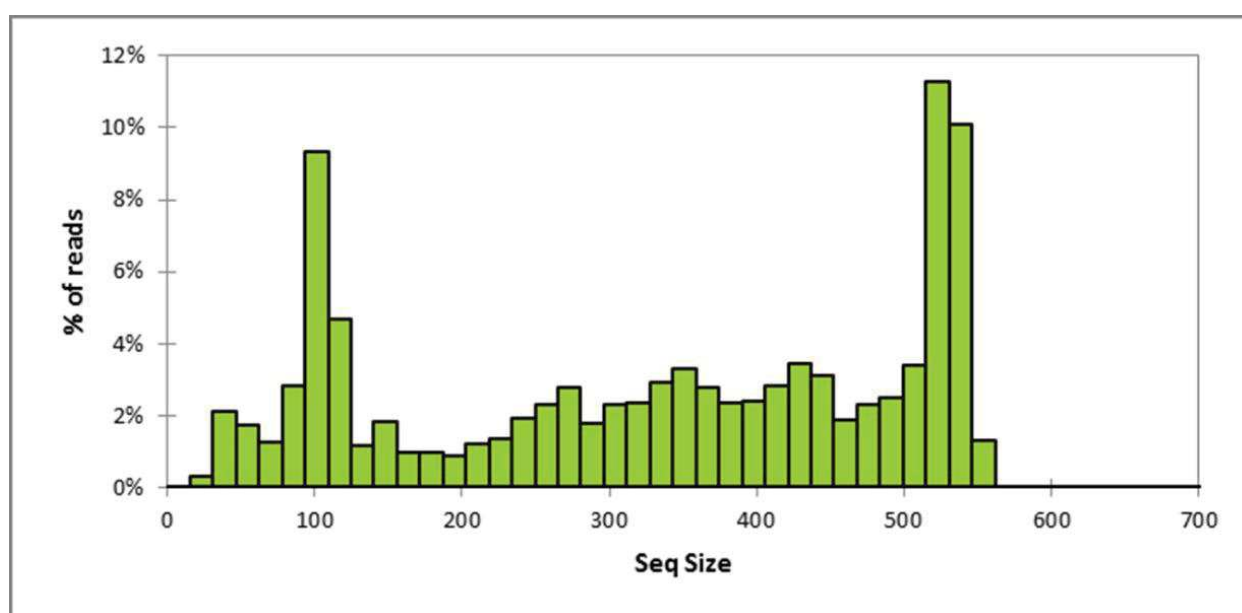
### 3.1. Library preparation, sequencing and read distribution

Eight citrus accessions combined with seven gene fragments were used to test an amplicon library method combining two PCR step to produce amplicons with varietal tags and necessary "Titanium" sequences for 454 sequencing processing. The first PCR with gene specific primers and M13 extension in 5' was successful for all gene fragments and varieties with the amplification of expected DNA fragment of the expected size (between 524 and 630 bp; [Supplementary information – Table A.1]. PCR products were purified and

1 µl was used to perform the second PCR with primers combining the titanium sequence, the MID specific to each genotype and the M13 extension in 3'. After this second PCR run, amplicons were purified by splicing DNA bands. Among the 56 amplicons, five were judged to be of too low quality (high percentage of low size fragments) and 51 amplicon tagged libraries were mixed according to their relative DNA concentration and were sequenced. A total of 69,425 reads were obtained for the 51 amplicons. Reads were classified according to their MID and Titanium and MID sequences were removed using the 454 software tools. The number of reads for each variety (MID) varied between 6,387 for Corsican citron and 11,901 for 'Pink' pummelo (table 2). Average length of reads was 333. However, this average covered a heterogeneous distribution with 25% of short reads (<150) and a peak of long reads between 510 and 540 bp representing 21% of the amplicons (Figure 2).

**Table 2.** Number and length of reads obtained for the different citrus varieties

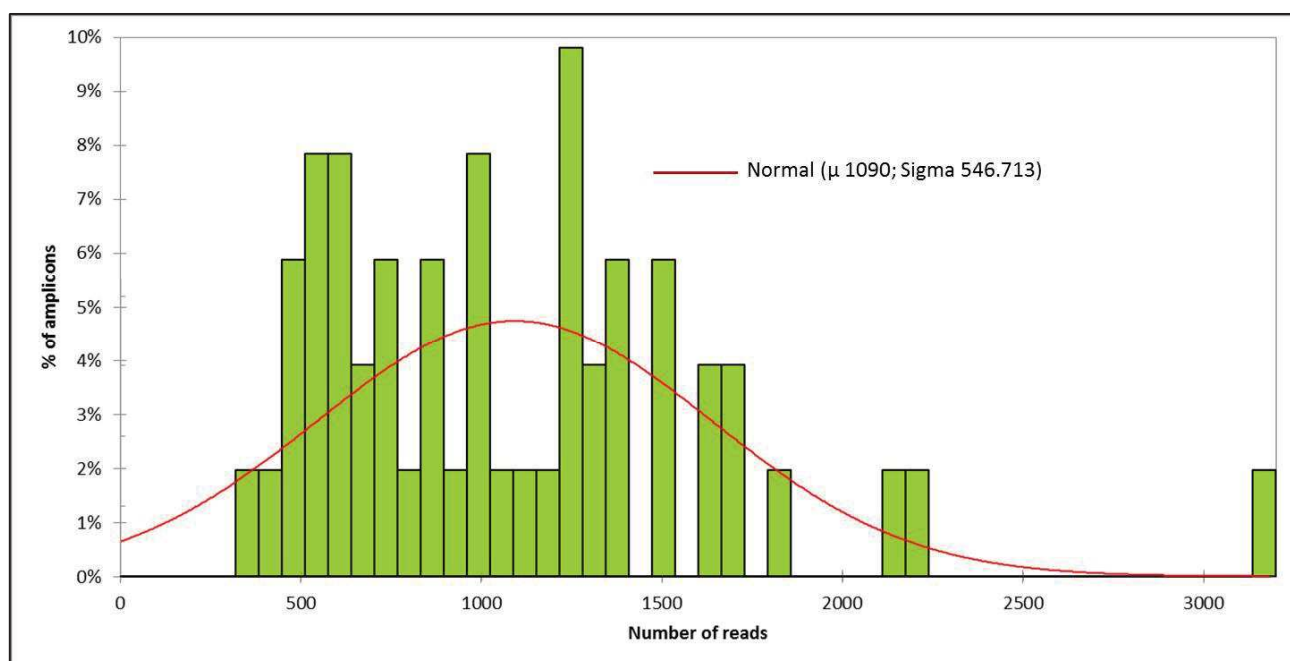
Varieties	Total reads	Usefull reads	% usefull reads	Medium read size (bp)
'Etrog' citron	9 505	8504	89%	332
'Corsican' citron	6 387	5516	86%	377
'Chandler' pummelo	7 730	6199	80%	313
'Pink' pummelo	11 901	10645	89%	344
'Cleopatra' mandarin	6 632	5389	81%	337
'Willowleaf' mandarin	11 203	7487	67%	306
'Clemenules' haploid clementine	8 632	6656	77%	347
'Clemenules' clementine	7 435	5198	70%	259
Total reads	69 425	55 594	80%	333



**Figure 2.** Size distribution of the amplicons



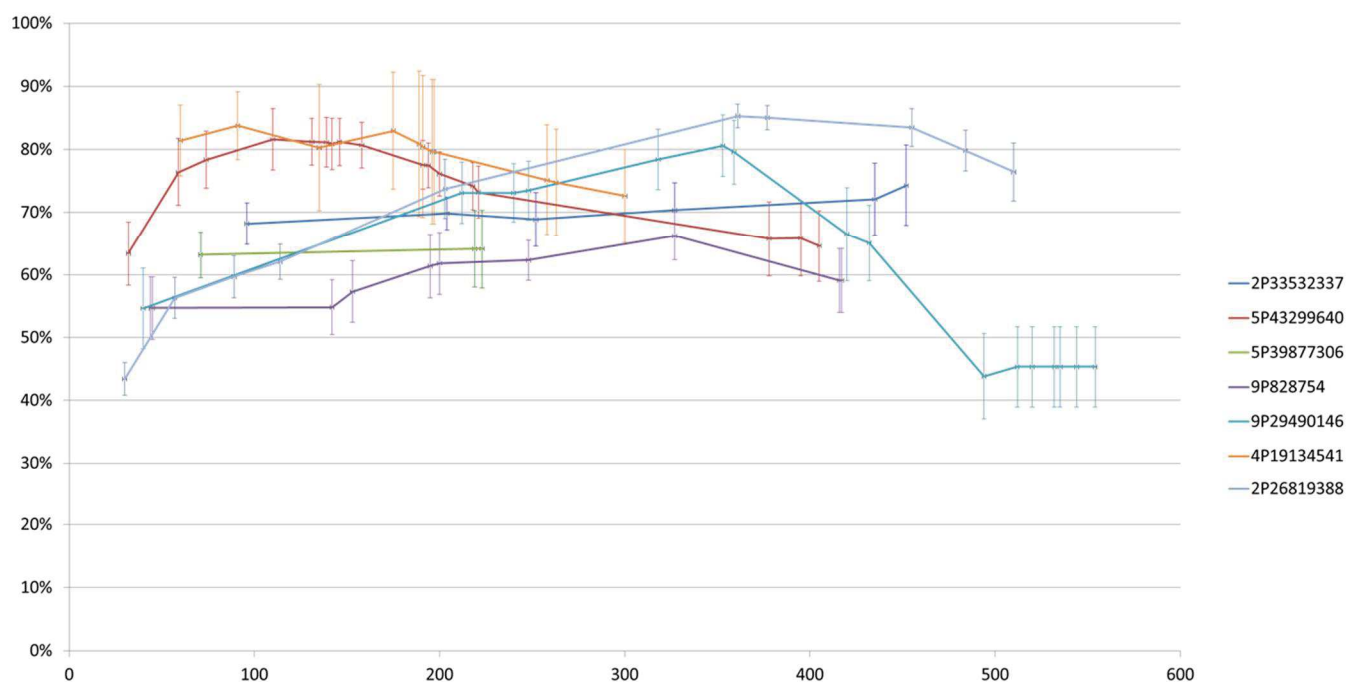
We used the SeqMan NGen software to analyse the reads. 80% of the reads (55,594) were found useful for genotype calling (table 2). The remaining 20% were short reads without the targeted sequences. Among the useful sequence the average sequence size was 440 bp. The useful reads were classified according to their MID (varieties) and fragment gene sequences. Considering a homogenous distribution of reads among the amplicons, the expected average number of read per amplicon was 1,090. The distribution of the number of reads is centered on this value but is far to follow a normal law (Figure 3). One amplicon (4P19134541/'Pink' pummelo) is highly over-represented with more than 3,000 reads. 5P39877306 is over-represented comparatively to other gene fragments [**supplementary information - Table A.4**]. Seven amplicons (13%) displayed less than half of the average expected read value.



**Figure 3. Reads distribution for 51 amplicons**

### ***3.2. Genotype calling and average distribution of reads/SNP loci in relation with position in the amplicon***

For the seven genes a total of 73 SNPs were discovered. When analyzing the coverage of each SNP relative to the useful number of reads of the corresponding gene fragments (Figure 4) a global tendency was observed for a better coverage in the central part compared with distal SNPs. This is a logical result considering that 454 amplicon sequencing was performed both from 3' and 5' ends and that the average size of useful sequences was 440, with longer amplicons. Coverage under 50% of the read number of the amplicons were found for the very distal SNPs of 9P29490146 (630 bp expected sequence size) and 2P26819388 (535 bp expected sequence size) amplicons.



**Figure 4.** Coverage of each SNP relatively to the number of useful read of the considered amplicons (standard deviations were estimated from the value obtained for different varieties analysed)

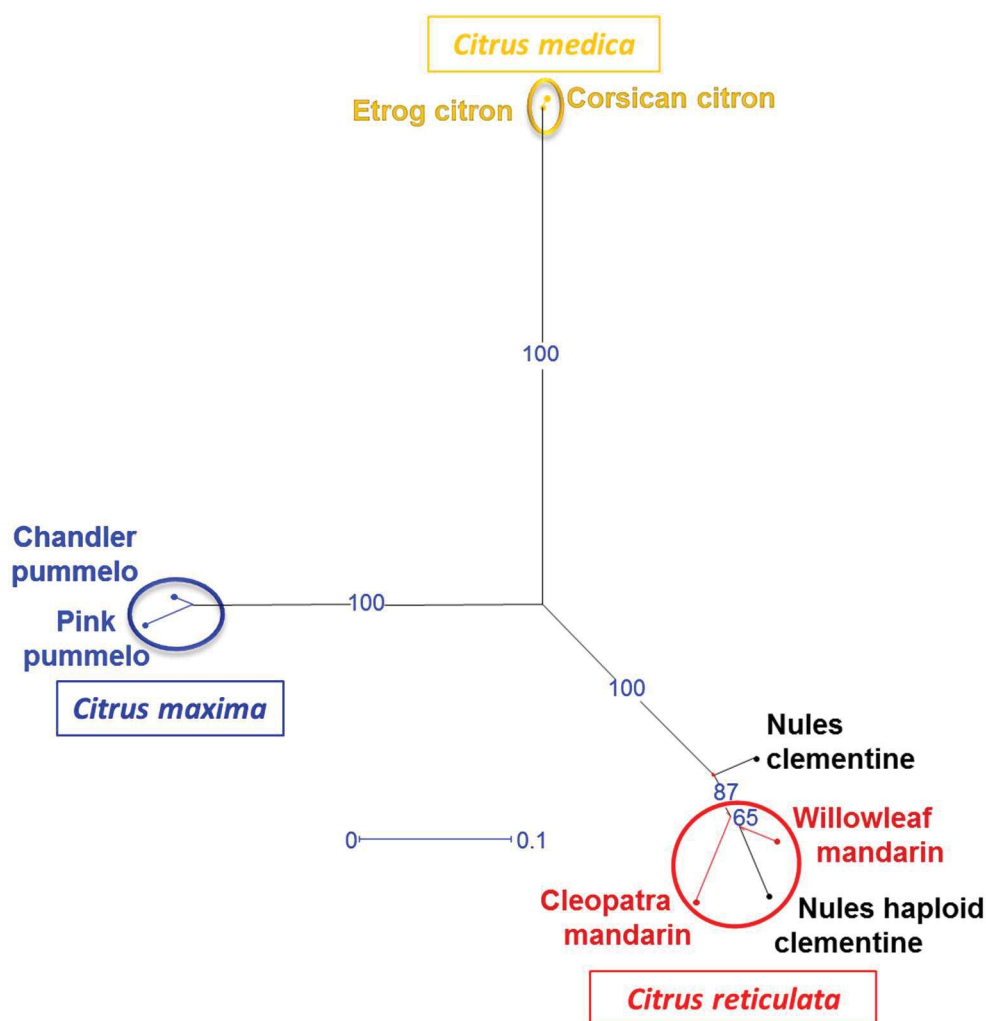
### 3.3. Comparative genotype calling of 454 sequencing from “two step amplicon library” and direct Sanger amplicon sequencing

All amplicons of the first run of PCR produced for 454 sequencing were also sequenced by Sanger sequencing method. Indels were observed in two gene fragments (4P19134541 and 5P39877306) producing unreadable chromatograms from the indel position to the end of the sequence. However, combining the forward and reverse sequences we successfully established consensus Sanger sequences. The same SNP loci were identified by the two methods and SNP genotype callings were the same with the two methods for all amplicons (table 3). Moreover the genotype calling of the haploid clementine with the two methods totally conform to the sequences extracted from the reference whole genome sequence (phytozome.net) developed from the same haploid line. NJ trees done from all SNPs of all gene fragments (Figure 5) clearly separated *C. medica* from *C. maxima* and from *C. reticulata* and showed strong bootstrap support in all branches of the three different clusters.

**Table 3.** SNP calling from 454 and Sanger sequencing methods

Gene fragment	Theoretical amplicon size	Useful Sanger sequence size	Nb SNPs Sanger	Nb SNPs 454	SNP/kb	Intron Size	SNP Intron	Exon size	SNP exon	5'UTR Size	SNP 5'UTR	3'UTR size	SNP 3'UTR
9P828754	589	589	10	10	16.9	–	–	311	3	–	–	278	7
4P19134541*	534	534	11	11	20.6	–	–	129	2	405	9	–	–
2P26819388	535	535	10	10	18.7	–	–	535	10	–	–	–	–
9P29490146	630	610	16	16	25.4	321	12	289	4	–	–	–	–
2P33532337	593	410	6	6	10.1	–	–	459	6	–	–	–	–
5P39877306*	524	439	3	3	5.7	–	–	439	3	–	–	–	–
5P42299640	552	490	17	17	30.8	–	–	490	17	–	–	–	–
Total	3957	3607	73	73	18.4	321	12	2652	45	405	9	278	7

\*presence of Indels

**Figure 5.** Neighbour joining tree based on all SNPs of all gene fragments.

### 3.4. Differentiation between and within the three basic taxa

The average SNP diversity is similar ( $H_e$  close to 0.40; table 4) for each gene fragments except for 5P39877306 that displayed a lower diversity ( $H_e=0.15$ ). The genetic structuration is high (significant heterozygosity deficit) for most gene fragments with  $F_w$  value between 0.6 and 0.8 except for 5P39877306 ( $F_w=-0.1$ ). Most part of this genetic structuration is associated to the phylogenetic origin as testified by the similar  $F_w$  and  $F_{st}$  values calculated considering *C. maxima*, *C. reticulata* and *C. medica* as three subpopulations. The relative levels of differentiation between the representatives of the three basic taxa varied from 10.95 to 12.83 SNPs/kb and was in average 10 times higher than the within taxa diversity (from 0.15 for *C. medica* to 2.11 for *C. reticulata*, table 5).

**Table 4.** Genetic parameters of SNP polymorphism for each gene fragment

	He	Ho	Fw	Fst
2P26849388	0.41	0.12	0.65	0.82
2P33532337	0.38	0.06	0.74	0.78
4P19134541	0.39	0.03	0.80	0.82
5P39877306	0.15	0.17	-0.09	0.00
5P42299640	0.39	0.08	0.68	0.76
9P828754	0.37	0.12	0.59	0.72
9P29490146	0.39	0.05	0.76	0.82

He: Unbiased expected heterozygosity, Ho observed heterozygosity, Fw: fixation index value; Fst: Fstat parameter.

**Table 5.** Intra and inter varietal group dissimilarities (Average number of SNP/kb between two varieties within –diagonal- and between varietal groups).

	<i>Mandarin</i>	<i>Pummelo</i>	<i>Citron</i>
<i>Mandarin</i>	2.11		
<i>Pummelo</i>	10.95	1.06	
<i>Citron</i>	12.30	12.83	0.15

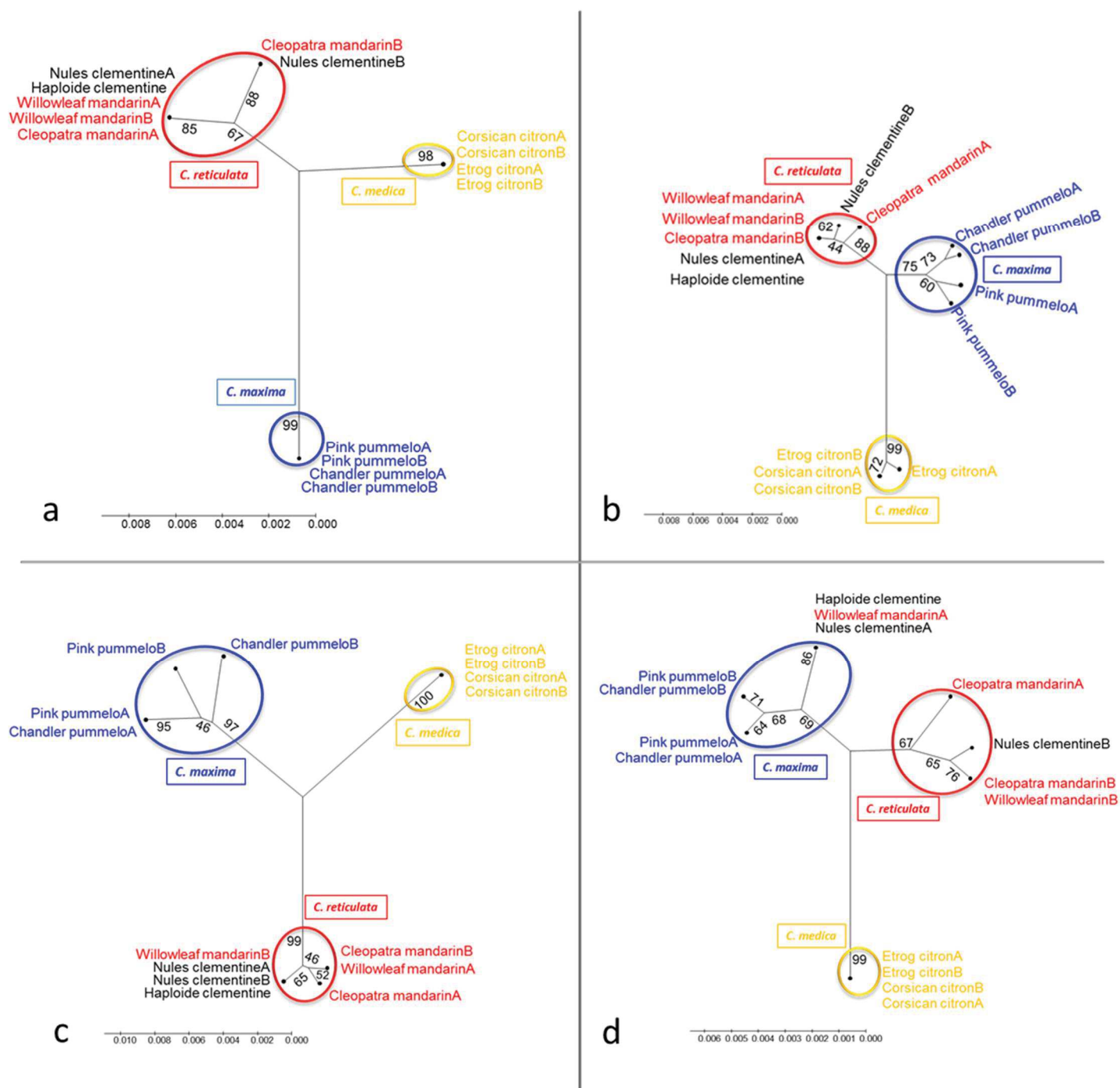
### 3.5. Haplotype inference

For each variety, the allelic phase between successive SNPs of a same gene fragment was determined step by step from the 5' end to the 3' by identifying the allele association of each marker pair in the set of reads covering both markers. A mean number of 5.71 haplotypes were identified per gene fragment. 9P29490146 was the most polymorphic fragment with eight haplotypes, and 2P33532337, 4P19134541 and 5P39877306 the least one with only four haplotypes identified (table 6).

**Table 6.** Haplotypes per gene fragment and varieties

	2P26819388	2P33532337	4P19134541	5P39877306	5P42299640	9P828754	9P29490146	Average
'Cleopatra' mandarin A	H6	H3	H1	H4	H5	H3	H4	
'Cleopatra' mandarin B	H5	H3	H2	H2	H6	H6	H3	
'Willowleaf' mandarin A	H1	H3	H2	H1	H6	H5	H3	
'Willowleaf' mandarin B	H5	H4	H2	H3	H7	H6	H3	
Mandarin haplotypes	3	2	2	4	3	3	2	2.71
'Chandler' pummelo A	H3	H2	H4	H4	H3	H2	H7	
'Chandler' pummelo B	H2	H1	H4	H4	H2	H1	H6	
'Pink' pummelo A	H3	H1	H4	H4	H3	H1	H7	
'Pink' pummelo B	H2	H1	H4	H4	H1	H1	H8	
Pummelo haplotypes	2	2	1	1	3	2	3	2
'Corsican' citron A	H7	H2	H3	H4	H4	H3	H1	
'Corsican' citron B	H7	H2	H3	H4	H4	H3	H1	
'Etrog' citron A	H7	H2	H3	H4	H4	H3	H2	
'Etrog' citron B	H7	H2	H3	H4	H4	H3	H1	
Citron haplotypes	1	1	1	1	1	1	2	1.14
'Clemenules' clementine A	H1	H3	H2	H3	H7	H6	H3	
'Clemenules' clementine B	H4	H2	H1	H4	H7	H2	H5	
Haploid clementine	H1	H3	H2	H3	H7	H6	H3	
Clementine haplotypes	2	2	2	2	1	2	2	1.86
Total diferent haplotypes	7	4	4	4	7	6	8	5.71

For four gene fragments the inter-taxa differentiation was enough to clearly differentiate clusters of haplotype of each taxon (Figure 6a, b, c, and d).



**Figure 6.** UPGMA of the haplotypes of each gene fragments with enough inter-taxa differentiation to clearly differentiate clusters of haplotype of each taxon: a. 4P19134541; b. 9P29490146; c. 5P42299640; d. 2P26819388.

For 5P39877306 fragment the diversity structure was not enough for taxa differentiation (Figure 7), while 2P33532337 and 9P828754 differentiate only *C. reticulata* from the two other taxa (Figure 8a and b). 2P33532337 and 5P39877306 were the fragments with the lowest number of SNPs identified and 5P39877306 was the smallest fragment sequenced. Clementine, known to be a hybrid between sweet orange and ‘Willowleaf’ mandarin (Ollitrault *et al.*, 2012a; Ollitrault *et al.*, 2012b), displays two haplotypes for six of the seven gene fragments. In three of the four fragments with clear taxa differentiation, the two haplotypes of clementine have *C. reticulata* origin. For the 2P26819388 gene fragment

one of the clementine haplotypes is in the *C. maxima* cluster as one of the 'Willowleaf' mandarin. Considering that sweet orange results from hybridization between the *C. reticulata* and *C. maxima* gene pools (Nicolosi *et al.*, 2000; Ollitrault *et al.*, 2012b; Garcia-Lor *et al.*, 2013) and is not related with *C. medica*, it can be assumed that for fragments 2P33532337 and 9P828754 clementine displays a *C. reticulata/C. maxima* interspecific heterozygosity, with the *C. reticulata* fragment inherited from 'Willowleaf' mandarin and the *C. maxima* one from sweet orange.

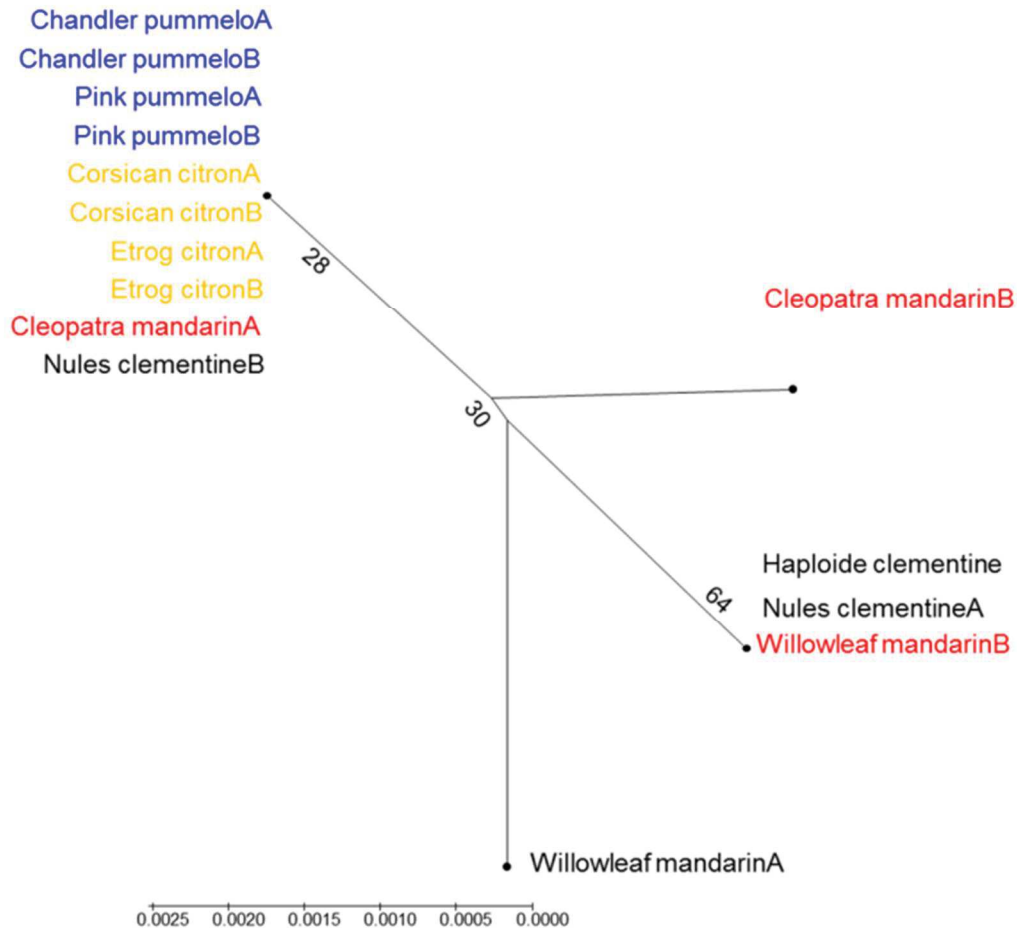
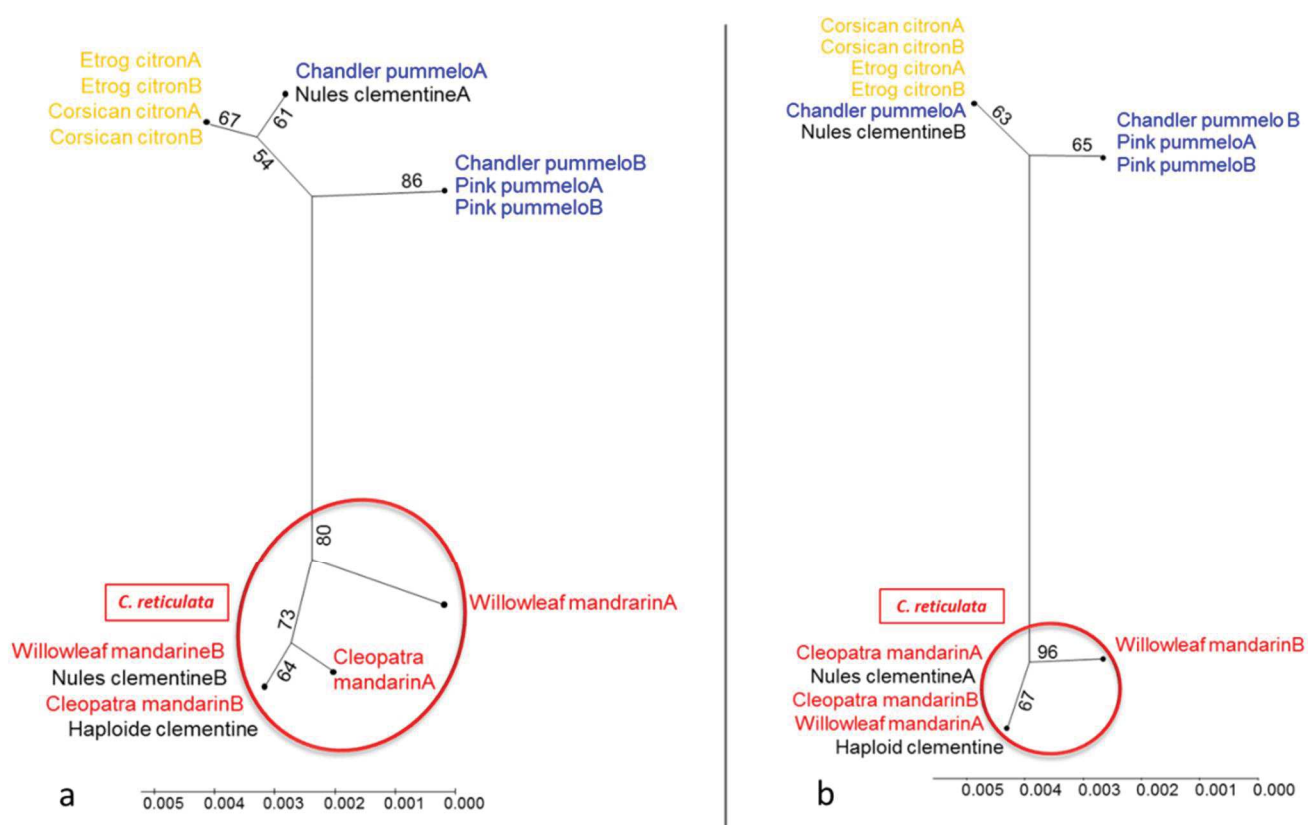


Figure 7. UPGMA of 5P39877306 gene fragment haplotypes.



**Figure 8. UPGMA of the haplotypes 9P828754 (a) and 2P33532337 (b) gene fragments.**

#### 4. Discussion

Bar coding is an essential step for parallel sequencing of reduced genome representation with the objective of pooling multiple individuals in a same sequencing run. For sample bar coding with PCR enrichment based strategies, Meyer *et al.* (Meyer *et al.*, 2007; Meyer *et al.*, 2008) (Meyer *et al.*, 2007; Meyer *et al.*, 2008) proposed a ligation method of the MID to the end of the amplicon, improved by Griffin *et al.* (Griffin *et al.*, 2011). However, it is still costly and time consuming while the insertion of the MID in the PCR primer between the gene specific sequence and the Titanium extension (Binladen *et al.*, 2007) adopted by Roche on its standardized protocols for amplicon sequencing rapidly becomes very expensive because it requires the purchase of long primers for every combination of targeted sequence/samples used in the study. In the present work we have tested a two-step PCR similar to that of Bybee *et al.* (Bybee *et al.*, 2011a) to prepare bar coded library for 454 sequencing for seven gene fragment targets and eight individuals. It was efficient and provided 454 SNP calling data identical to Sanger sequencing of the same amplicons. The high coverage level of each amplicon is clearly a key element for this high conformity of 454 SNP mining data. 80% of the total reads were useful for the genotype calling. Those not useful were short sequences of primer dimers that remain a problem despite the cleaning process of the targeted amplicons after the two-step PCR, probably due to short fragment bias during emPCR. Primer dimers were also found to be a difficulty by Bybee *et al.* (Bybee *et al.*, 2011b). The rate of primer dimers should probably be improved by using a Taq polymerase with proofreading capability (AccuPrime Taq) during the two-step PCR as



proposed by Bybee *et al.* (Bybee *et al.*, 2011a). The distribution of the number of reads by amplicon was centered on the theoretical average value however heterogeneity was observed between the 56 sequenced amplicons. One amplicon was highly over-represented probably due to a bad estimation of its DNA concentration. The smaller targeted fragment was over-represented compared to other gene fragments. Seven amplicons (13%) displayed less than half of the average expected read value. Bybee *et al.* (Bybee *et al.*, 2011a) also observed important differences of the number of reads per amplicons and a lower number for the longer amplicons. The heterogeneity of coverage obtained with our two step PCR approach was much lower than the one recently reported using the Fluidigm Access Array System with citrus (Curk *et al.*, accepted).

A total of 73 SNPs were found in the 7 gene fragments. The analysis of the relative number of reads for a specific SNP within one amplicon displayed variations between 43% and 85% of the numbers of reads of the amplicons. Higher values were obtained in the middle of the amplicons where both forward and reverse reads provided data for the SNP. Such one to two variation may be taken into account when planning the reading coverage for efficient haplotype inference. Over all gene fragments 18.4 SNPs/kb were identified. This rate was variable between gene fragments (from 5.7 to 30.8) and was close to be two-fold higher in introns than in exons as already observed by Garcia-Lor *et al.* (Garcia-Lor *et al.*, 2013). Most of the genetic structuration of the sample was associated to the phylogenetic origin (*C. maxima*, *C. reticulata* and *C. medica* differentiation) as testified by the similar *F<sub>w</sub>* and *F<sub>st</sub>* values and the NJ tree analysis with all gene fragments concatenated. The relative levels of differentiation between the representatives of these taxa were in average 10 times higher than the within taxa diversity. This is in agreement with previous molecular studies (Herrero *et al.*, 1996; Nicolosi *et al.*, 2000; Barkley *et al.*, 2006; Garcia-Lor *et al.*, 2013) revealing the high structuration of the diversity of the cultivated citrus around *C. maxima*, *C. medica*, and *C. reticulata* and recognizing these taxa as the ancestors of the modern cultivated *Citrus* species.

For four out of the seven genes, the differentiation level between taxa was enough to infer phylogenetic origin of the haplotypes. The three remaining gene fragments displayed the lowest rate of SNPs/Kb and one of them was the shortest analysed fragment. For two of these fragments *C. medica* and *C. maxima* were not significantly differentiated but were clearly differentiated from the *C. reticulata* cluster. The 5P39877306 fragment displayed no genetic structuration (*F<sub>w</sub>* and *F<sub>st</sub>* close to 0) and therefore no phylogenetic information. The targeted amplification of intronic regions displaying more polymorphism, should probably improve the resolution of the phylogeny of these three genes.

Haplotype inference is an intensive research area particularly in human genetics (Browning and Browning, 2011). A lot of algorithms have been developed to infer the phase between heterozygous loci but the underlying hypothesis of the models adapted to human genetics should be far from plant genetics reproductive biology and evolutionary patterns. Therefore haplotyping through laboratory-based experimental methods remain an essential approach in plants. Sanger sequencing of direct amplicons of heterozygous genotypes gives double peaks at heterozygous sites sometime difficult to reliably determine. Moreover allelic phase between heterozygous loci cannot be determined. Sanger sequencing of cloned gene provide haplotype (Fortune *et al.*, 2008; Rousseau-Gueutin *et al.*, 2009; Ramadugu *et al.*, 2013); However it is time-consuming and expensive limiting greatly the number of individuals

and genes to be studied.. High-throughput ‘next-generation sequencing’ of barcoded DNA mixtures is, comparatively with other approaches, an affordable and successful solution to establish haplotype data. Deep amplicon sequencing using the 454 technology allows, in human research, to obtain thousands of haplotypes calls per amplicon for the beta-defensin locus. Taudien *et al.* (Taudien *et al.*, 2010) considered this method as efficient enough for haplotyping and copy number estimation in small to medium sized cohorts. In the present work it was successfully applied to heterozygous genotypes to determine haplotypes, and revealed the admixture structure of Clementine with complete *C. reticulata* origin for 3 gene fragments and *C. reticulata/C. maxima* heterozygosity for three fragments. The fact that sometimes primers do not amplify some alleles in certain taxa due to variation in the primer area, can limit the efficiency of PCR-based sequence analysis in admixture genotypes. For Ramadugu *et al.* (Ramadugu *et al.*, 2013) such PCR bias probably leads to underestimate the heterozygosity in some citrus species. Such bias should be limited using re-sequencing data of the different taxa to define primers in conserved gene fragments. Another key should be to preferentially amplify intronic region, to optimize the SNPs rate and avoid selective effects for more efficient haplotype phylogenetic studies.

## 5. Conclusion

The two-step PCR is a simple approach allowing producing tagged amplicons that are ready to be directly sequenced using a next-generation platform. It is based on a first PCR to amplify a targeted genome region with universal primer extension [M13 (-40)], followed by a second PCR that attaches a known barcode to identify amplicons from different samples with the necessary proprietary sequence for parallel sequencing platforms. It avoids purchasing every possible combination of MID-locus specific primer. Another advantage is the complete control of PCR protocol components and that it is fully scalable. Difficulties are the accurate quantification and equimolar pooling of PCR amplicons from multiple reactions and the production of primer-dimers diminishing the proportion of useful reads. The two-step PCR method was efficient to establish tagged libraries of amplicons to be used for 454 sequencing and the 454 sequencing data provided genotyping calling identical to that obtained by Sanger sequencing. Haplotypes were successfully established and the level of differentiation between taxa coupled with the 500 bp reads of 454 was enough for phylogenetic differentiation of haplotypes of four out of the seven targeted gene fragments. It is therefore an interesting alternative for targeted haplotyping by NGS approach and new generation phylogeny.

## ACKNOWLEDGEMENTS

This work was supported by a grant (AGL2011-26490) from the Ministry of ‘Economía y Competitividad’–‘Fondo Europeo de Desarrollo Regional’ (FEDER) and a grant (Prometeo II/2013/008) from the Generalitat Valenciana, Spain. G Ancillo is recipient of a contract from Conselleria de Agricultura, Pesca y Alimentación (Generalitat Valenciana, Spain) under Proy\_IVIA13/10.

## Author Contributions Statement

FC and GA performed the experiment and data analysis, PO and LN conceived the experiment; FC and PO wrote the manuscript; All authors reviewed the manuscript.

## References

### Literature Cited

- Baird NA, Etter PD, Atwood TS, et al. 2008.** Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS one* **3**: e3376.
- Barbazuk WB, Emrich SJ, Chen HD, Li L, Schnable PS. 2007.** SNP discovery via 454 transcriptome sequencing. *The Plant Journal : for cell and molecular biology* **51**: 910-918.
- Barkley NA, Roose ML, Krueger RR, Federici CT. 2006.** Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *TAG.Theoretical and applied genetics.Theoretische und angewandte Genetik* **112**: 1519-1531.
- Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 1996-2004.** GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire Génome, Populations, Interactions, CNRS UMR 5000* .
- Binladen J, Gilbert MT, Bollback JP, et al. 2007.** The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS one* **2**: e197.
- Browning SR, Browning BL. 2011.** Haplotype phasing: existing methods and new developments. *Nature reviews Genetics* **16**; **12**: 703-14.
- Bybee SM, Bracken-Grissom H, Haynes BD, et al. 2011a.** Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. *Genome biology and evolution* **3**: 1312-1323.
- Bybee SM, Bracken-Grissom HD, Hermansen RA, Clement MJ, Crandall KA, Felder DL. 2011b.** Directed next generation sequencing for phylogenetics: An example using Decapoda (Crustacea). *Zoologischer Anzeiger - A Journal of Comparative Zoology* **250**: 497-506.
- Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, DePristo MA. 2012.** Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC genomics* **13**: 375-2164-13-375.
- Chan YC, Roos C, Inoue-Murayama M, et al. 2010.** Mitochondrial genome sequences effectively reveal the phylogeny of *Hylobates gibbons*. *PLoS one* **5**: e14419.
- Crosby LD, Criddle CS. 2007.** Gene capture and random amplification for quantitative recovery of homologous genes. *Molecular and cellular probes* **21**: 140-147.
- Curk F, Ancillo G, Garcia-Lor A, et al. accepted.** Next generation haplotyping to decipher the nuclear genomic interspecific admixture in *Citrus* species; analysis of the Chromosome 2. *BMC Genetics* .
- Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011.** Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature reviews Genetics* **12**: 499-510.
- Elshire RJ, Glaubitz JC, Sun Q, et al. 2011.** A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* **6**: e19379. doi:10.1371/journal.pone.0019379-doi:10.1371/journal.pone.0019379.
- Fortune P, Pourtau N, Viron N, Ainouche M. 2008.** Molecular phylogeny and reticulate origins of the polyploid *Bromus* species from section *Genea* (Poaceae). *Am J Bot* **95**: 454-464.
- Garcia-Lor A, Curk F, Snoussi-Trifa H, et al. 2013.** A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. *Annals of botany* **111**: 1-19.

- Glenn TC. 2011.** Field guide to next-generation DNA sequencers. *Molecular ecology resources* **11**: 759-769.
- Griffin PC, Robin C, Hoffmann AA. 2011.** A next-generation sequencing method for overcoming the multiple gene copy problem in polyploid phylogenetics, applied to Poa grasses. *BMC biology* **9**: 19-7007-9-19.
- Gunnarsdottir ED, Li M, Bauchet M, Finstermeier K, Stoneking M. 2011.** High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome research* **21**: 1-11.
- Herrero R, Asins MJ, Pina JA, Carbonell EA, Navarro L. 1996.** Genetic diversity in the orange subfamily Aurantioideae. II. Genetic relationships among genera and species. *Theoretical and Applied Genetics* **93**: 1327-1334.
- Hillier LW, Marth GT, Quinlan AR, et al. 2008.** Whole-genome sequencing and variant discovery in *C. elegans*. *Nature methods* **5**: 183-188.
- Linder CR, Rieseberg LH. 2004.** Reconstructing patterns of reticulate evolution in plants. *American Journal of Botany* **91**: 1700-1708.
- Luro F, Costantino G, Terol J, et al. 2008.** Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other Citrus species and their effectiveness for genetic mapping. *BMC genomics* **9**: 287.
- Maughan H, Redfield RJ. 2009.** Extensive variation in natural competence in *Haemophilus influenzae*. *Evolution; international journal of organic evolution* **63**: 1852-1866.
- McCormack JE, Maley JM, Hird SM, Derryberry EP, Graves GR, Brumfield RT. 2012.** Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. *Molecular phylogenetics and evolution* **62**: 397-406.
- Meyer M, Stenzel U, Myles S, Prufer K, Hofreiter M. 2007.** Targeted high-throughput sequencing of tagged nucleic acid samples. *Nucleic Acids Res* **35**: 97.
- Meyer M, Stenzel U, Hofreiter M. 2008.** Parallel tagged sequencing on the 454 platform. *Nature protocols* **3**: 267-278.
- Moonsamy PV, Williams T, Bonella P, et al. 2013.** High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ system for simplified amplicon library. *Tissue Antigens* **81**: 1399-0039.
- Morin PA, Archer FI, Foote AD, et al. 2010.** Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome research* **20**: 908-916.
- Nicolosi E, Deng ZN, Gentile A, La Malfa S, Continella G, Tribulato E. 2000.** Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics* **100**: 1155-1166.
- Ollitrault P, Terol J, Chen C, et al. 2012a.** A reference genetic map of *C. clementina* hort. ex Tan.; citrus evolution inferences from comparative mapping. *BMC genomics* **13**: 593-2164-13-593.
- Ollitrault P, Terol J, Garcia-Lor A, et al. 2012b.** SNP mining in *C. clementina* BAC end sequences; transferability in the Citrus genus (Rutaceae), phylogenetic inferences and perspectives for genetic mapping. *BMC genomics* **13**: 13-2164-13-13.
- Parks M, Cronn R, Liston A. 2009.** Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC biology* **7**: 84-7007-7-84.
- Perrier X, Jacquemoud-Collet JP. 2006.** DARwin software. .

- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012.** Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS one* **7**: e37135.
- Ramadugu C, Pfeil BE, Keremane ML, Lee RF, Maureira-Butler IJ, Roose ML. 2013.** A six nuclear gene phylogeny of Citrus (Rutaceae) taking into account hybridization and lineage sorting. *PloS one* **8**: e68410.
- Rousseau-Gueutin M, Gaston A, Aïnouche A, et al. 2009.** Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): new insights from phylogenetic analyses of low-copy nuclear genes. *Mol Phylogenet Evol* **51**: 515-530.
- Saintenac C, Jiang D, Akhunov ED. 2011.** Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. *Genome biology* **12**: R88-2011-12-9-r88.
- Swingle WT, Reece PC. 1967.** The botany of Citrus and its wild relatives. In: W Reuther, HJ Webber, LD Batchelor, eds. *The citrus industry*, 2 nd edn. University of California Press: Berkeley, California, USA, 190-430.
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013.** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular biology and evolution* **30**: 2725-2729.
- Taudien S, Groth M, Huse K, et al. 2010.** Haplotyping and copy number estimation of the highly polymorphic human beta-defensin locus on 8p23 by 454 amplicon sequencing. *BMC Genomics* **11**: 252.
- Van Tassell CP, Smith TP, Matukumalli LK, et al. 2008.** SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature methods* **5**: 247-252.
- Wright S. 1978.** Variability Within and Among Natural Populations. In: Anonymous *Evolution and the Genetics of Populations*. The University of Chicago Press: Chicago, IL.

## Supplementary data

Table A.1. Gene specific primers for Fluidigm amplification

Mest	Seq. name	Phytozome* reference	Scaffold	Primer seq.	Expected Size
MEST 495	2P26819388	scaffold_2:26,819,389..26,819,923	2	Forwar d ACTTTGGGGCTTTTGTGCT Reverse AAAATACCACTCCTCGCTGTTT	535
MEST 46	2P33532337	scaffold_2:33,532,339..33,532,930	2	Forwar d CCAAAAACGAGAACCAGAAAC Reverse ATTCTCGCCCTTTGCTTG	593
MEST 425	4P19134541	scaffold_4:19,134,541..19,135,074	4	Forwar d CCCGTCTCTCCTTGAGTTTG Reverse ATCAGTTTGTTCGGCTCCA	534
MEST 56	5P43299640	scaffold_5:42,299,640..42,300,191	5	Forwar d CGCACTCCATCTCTCTCC Reverse CTAATCCCGCCTTCTGTG	552
MEST 115	5P39877306	scaffold_5:31,410,378..31,410,901	5	Forwar d GCATCAACTCCAACACTCTCTC Reverse CCAATCCACGGCTTA TCTCTC	524
MEST 330	9P828754	scaffold_9:828,754..829,342	9	Forwar d GATGATGTTGCTGAGGGAAAG Reverse CCAAGTTACATAGAGAGTCAAGCAC	589
HYB	9P29490146	scaffold_9:29,490,146..29,490,775	9	Forwar d CAGATCGAGACTTTCACGGA Reverse AAAGAAGCATGCCACAGAGC	630

\*[http://www.phytozome.net/search.php?show=text&method=Org\\_Clementina](http://www.phytozome.net/search.php?show=text&method=Org_Clementina)

Table A.2. Multiplex genotype Identifiers (MID), related genotypes and Titanium primers

ID	Seq	Genotypes
MID 01	ACGAGTGCCT	Etrog citron IVIA169
MID 02	ACGCTCGACA	Corsican citron IVIA567
MID 03	AGACGCACTC	Chandler pummelo IVIA207
MID 04	AGCACTGTAG	Pink pummelo IVIA275
MID 05	ATCAGACACG	Cleopatra mandarin IVIA385
MID 06	ATATCGCGAG	Willowleaf mandarin IVIA154
MID 07	CGTGCTCTA	Clemenules haploid clementine IVIA638
MID 08	CTCGCGTGTC	Clemenules clementine IVIA22
Forward Titanium primers A (including the Key)	CGTATCGCCTCCCTCGCGCCATCAG	
Reverse Titanium primers B (including the Key)	CTATGCGCCTTGCCAGCCCCTCAG	
Final complete forward primer	5'-(Primer A-Key)-(MID)-(Gene-specific-primer)-3'	
Final complete reverse primer	5'-(Primer B-Key)-(MID)-(Gene-specific-primer)-3'	

**Table A.3.** List of the universal sequencing primers tested

20 mer	GTTGTAAAACGACGGCCAGT	CACAGGAAACAGCTATGACC
ABI 20 mer	GACGTTGTAAAACGACGGCC	CACAGGAAACAGCTATGACC
ABI 18 mer	TGTAAAACGACGGCCAGT	CAGGAAACAGCTATGACC
M13 (-21)	TGTAAAACGACGGCCAGT	
M13 (-20)	GTAAAACGACGGCCAGTG	
<b>M13 (-40)</b>	<b>GTTTTCCAGTCACGAC</b>	<b>CAGGAAACAGCTATGAC</b>
M13 (-41)	GGTTTTCCAGTCACGAC	
M13/pUC		CAGGAAACAGCTATGACC
M13 (-27)		GGAAACAGCTATGACCATG
M13 (-48)		AGCGGATAACAATTCACAC

**Table A.4.** Number of reads per amplicon (Variety/gene fragment)

Gene fragment	9P29490146	2P33532337	5P43299640	5P39877306	9P828754	4P19134541	2P26819388
Expected Size	630	593	552	524	589	534	535
Etrog citron	972	662	1267	1637	1254	1395	1317
Corsican citron	715	330	0	1283	1169	891	1128
Chandler pummelo	853	464	0	2168	607	1224	883
Pink pummelo	1821	400	706	1696	1502	3165	1355
Cleopatra mandarin	512	575	554	1717	604	502	925
Willowleaf mandarin	523	994	970	1520	1269	1620	591
Haploid Clementine	664	1528	511	1253	1021	632	1047
Nules Clementine	765	0	0	2210	1407	816	0
Total	6 825	4 953	4 008	13 484	8 833	10 245	7 246

## Annexe I.4

## New Insights on Limes and Lemons Origin from Nuclear and Cytoplasmic Markers Genotyping and Targeted Nuclear Gene Sequencing

Franck Curk<sup>1,2</sup>, Andres Garcia-Lor<sup>2</sup>, Hager Snoussi<sup>3</sup>, Yann Froelicher<sup>4</sup>, Gema Ancillo<sup>2</sup>, Luis Navarro<sup>2</sup> and Patrick Ollitrault<sup>2,4</sup>

<sup>1</sup>Institut National de la Recherche Agronomique (INRA UR GEQA), UR1103, DGAP, France.; <sup>2</sup>Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, Valencia, Spain; <sup>3</sup>Institut National de la Recherche Agronomique de Tunisie (INRAT), Tunisia; <sup>4</sup>Département Scientifique Systèmes Biologiques (BIOS), Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD), France.

### Abstract

It is believed that *Citrus medica*, *C. maxima*, *C. reticulata* and *C. micrantha* have generated all cultivated citrus species. Depending on the classification, lemons and limes are classified either into two species, *C. limon* and *C. aurantifolia* (Swingle and Reece, 1967) or into more than 30 (Tanaka, 1977). In order to study the molecular phylogeny of this citrus group, we analyzed 23 targeted sequenced nuclear genes and used three mitochondrial and five chloroplastic markers for 15 lemons and limes compared with representatives of the four basic taxa. We observed three main groups, each one derived from direct interspecific hybridizations: (1) the Mexican lime group (*C. aurantifolia*), including *C. macrophylla*, arising from hybridizations between papeda (*C. micrantha*) and citron (*C. medica*); (2) the yellow lemon group (*C. limon*) that are hybrids between sour orange (*C. aurantium*, which is believed to be a hybrid between *C. maxima* and *C. reticulata*) and citron; and (3) a rootstock lemon/lime group (Rough lemon and Rangpur lime) that are hybrids between an acid mandarin and a citron. We also identified different probable backcrosses and genotypes with more complex origin. None of the analyzed limes and lemons shared the *C. medica* cytoplasm, while this taxon is the common nuclear contributor of all limes and lemons. Limes and lemons appear to be a very complex citrus varietal group with the contribution of the 4 basic taxa. Neither the Swingle and Reece classification nor the Tanaka fit with the genetic evidence.

**Keywords:** citrus, molecular phylogeny, SSR, Indel, SNP

In press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012



Annexe I.4

# New Insights on Lime and Lemon Origins from Targeted Nuclear Gene Sequencing, SSRs, InDels and Cytoplasmic Markers Genotyping



Franck CURK<sup>1,2</sup>, Andrés GARCIA-LOR<sup>2</sup>, Hager SNOUSS<sup>3</sup>, Yann FROELICHER<sup>4</sup>, Gema ANCILLO<sup>2</sup>, Luis NAVARRO<sup>2</sup>, and Patrick OLLITRAULT<sup>4,2</sup>

<sup>1</sup>INRA, UR1103 Génétique et Ecophysiologie de la Qualité des Agrumes (INRA UR GEQA), DGAP, France; <sup>2</sup>Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; <sup>3</sup>INRAT (Institut National de la Recherche Agronomique de Tunisie), Tunisia; and <sup>4</sup>Centre de Coopération International en Recherche Agronomique pour le Développement (CIRAD), BIOS, France. Contact email: curk@corse.inra.fr; Inavarro@ivia.es; patrick.ollitault@cirad.fr

## Introduction

It is believed that four basic taxa, citron (*C. medica*), pummelo (*C. maxima*), mandarin (*C. reticulata*) and *C. micrantha* have generated all cultivated *Citrus* species. Depending on the botanical classifications, lemons and limes are classified either into two species, *C. limon* and *C. aurantiifolia* (Swingle and Reece) or into more than 30 (Tanaka). In order to clarify the phylogeny of this *Citrus* group, we used nuclear SNPs from 15 targeted DNA fragments mined by sanger sequencing, 10 nuclear Indels (García-Lor *et al.*, 2012a, b), 23 nuclear SSRs (Froelicher *et al.*, 2008; Luro *et al.*, 2008), 3 mitochondrial Indels (Froelicher *et al.*, 2011) and 5 chloroplastic SSR markers (Lotfy *et al.*, 2003) for 15 lemons and limes compared with 19 representatives of the basic taxa and 8 secondary species.



## Maternal phylogeny

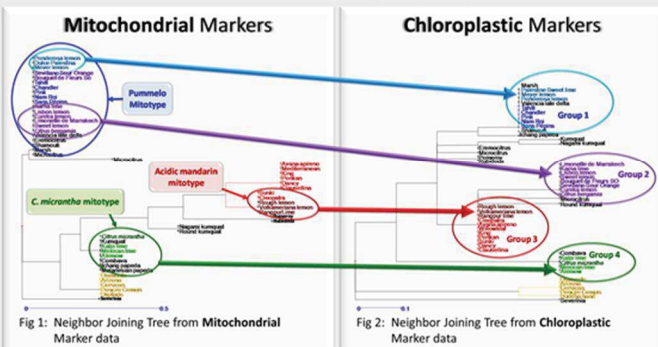


Fig 1: Neighbor Joining Tree from Mitochondrial Marker data

Fig 2: Neighbor Joining Tree from Chloroplastic Marker data

(Darwin (Dissimilarity Analysis and Representation for Windows): <http://darwin.cirad.fr/home.php>)

Four groups of lemons-and-limes were differentiated (Fig 2): Group 1 [Palestine sweet lime (*C. limetoides* Tan.), Meyer lemon (*C. meyeri* Y. Tan.), and Ponderosa (*C. pyriformis* Hassk.)] and Group 2 [Limonette de Marrakech (*C. limetta* Risso), Karina lime (*C. karina* Raf.), Lisbon, Sweet and Eureka lemons (*C. Limon* (L) Burm.) and *Citrus bergamia* Risso & Poit.] have a pummelo type cytoplasm (Fig 1). Group 2 share its cytoplasm with Sour Orange (*C. aurantium* L.) while group 1 share it with Sweet orange and Grapefruit (Fig 2); Group 3 [Rough lemon (*C. jambhiri* Lush.), Volkamer lemon (*C. limonia* Osb.), and Rangpur lime (*C. limonia* Osb.)] has an acidic Mandarin type cytoplasm and Group 4 [Mexican lime (*C. aurantiifolia* (Christm.) Swing), Alemow (*C. macrophylla* Wester) and Kalpi lime (*C. webberii* Wester)] has a *Citrus micrantha* type cytoplasm. None of the 4 groups shows any citron, *Microcitrus*, *Eremocitrus*, *Fortunella* or *Poncirus* maternal contribution.

## Nuclear Phylogeny

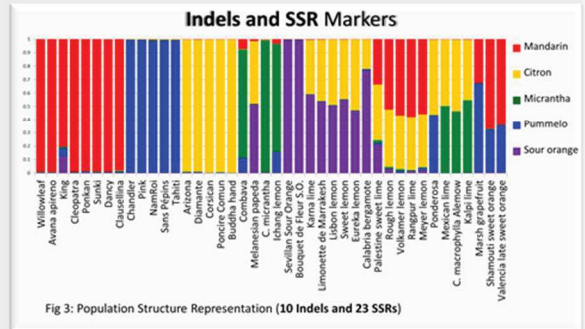


Fig 3: Population Structure Representation (10 Indels and 23 SSRs)

(Structure Software: <http://pritch.bsd.uchicago.edu/structure.html>)

All analyzed lemons and limes are related to citron (Fig 3). Sour oranges appear as a parental group of some lemons and limes like Karina lime, Limonette de Marrakech, Lisbon, Eureka and Sweet lemons, and *Citrus bergamia*. An another group constituted by Palestine sweet lime, Rough lemon, Volkamer lemon, Rangpur lime and Meyer lemon seems to result mainly from citron and mandarin contributions. Ponderosa seems to be directly related to pummelo and citron. Mexican lime, Alemow and Kalpi lime share equilibrated contributions of *C. micrantha* and citron. They are very clearly differentiated from the others limes and lemons in the Cluster analysis (Fig 6). The sub-clustering of the other limes and lemons (Fig 6) is globally in agreement with the structure analysis.

## Multilocus SNPs Analysis

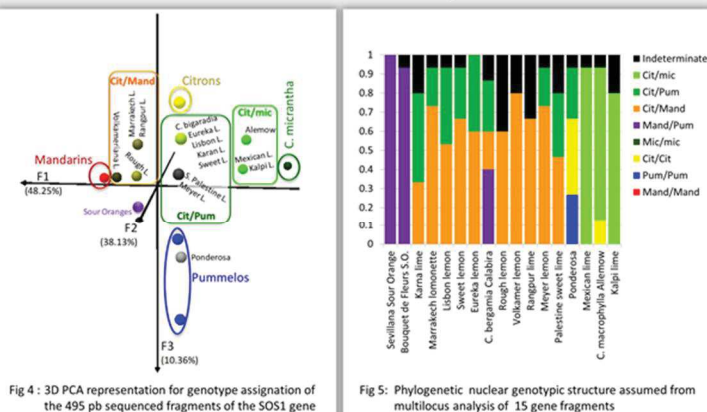


Fig 4: 3D PCA representation for genotype assignment of the 495 pb sequenced fragments of the S051 gene

Fig 5: Phylogenetic nuclear genotypic structure assumed from multilocus analysis of 15 gene fragments

575 SNPs were identified in 10849 pb of the 15 targeted sequenced fragments. For each gene a Principal Component Analysis (PCA) was performed from SNP data to analyze the phylogenetic genotypic structure (e.g. Gene S051, Figure 4). It was supposed that intermediary position between two basic taxa cluster corresponded to interspecific heterozygosity; this hypothesis was confirmed by checking the heterozygosity for phylogenetic diagnostic SNPs. For most of the amplified fragment, the polymorphism was enough for unambiguous multilocus differentiation of the basic species and assignment of a phylogenetic status in many limes and lemons confirming that citron was a direct parent of these genotypes. The other parents appear to be (i) interspecific hybrids M x P for Karina, Marrakech, Lisbon, Eureka, sweet, Meyer, and Palestine, (ii) mandarin parents for Rough lemon, Volkamer lemon and Rangpur lime, (iii) *C. micrantha* for Kalpi and Mexican lime. Mandarin/Pummelo heterozygosity was observed for *C. bergamia*, homozygosity citron/citron in Alemow and Ponderosa (as well as pummelo/pummelo for Ponderosa) suggesting more complex origins such as ((Pxm)xC) x (PxM) for *C. bergamia*, (Mic x C) x C for Alemow and (Px)(Cx) for Ponderosa.

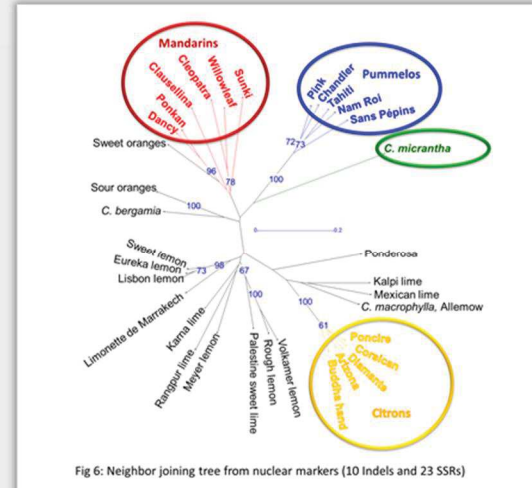


Fig 6: Neighbor joining tree from nuclear markers (10 Indels and 23 SSRs)

## References

Froelicher Y, Dambar D, Bassano JB, Costantino G, Lotfy S, Dubou C, Beaumont V, Brevier F, Riberuau AM, Luro F, Ollitault P. 2008. Characterization of microsatellite markers in mandarin orange (*Citrus reticulata* Bl.). Molecular breeding resources. 8 (1): 219-232.  
 Froelicher Y, Mouloua M, Bassano JB, Costantino G, Karam M, Luro F, Morillon X, Ollitault P. 2011. New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. Hort Science and Genetics 7: 48-51.  
 García-Lor A, Luro F, Navarro L, Ollitault P. 2012. Comparative use of indel and SSR markers in deciphering the interspecific structure of cultivated citrus genetic diversity: A perspective for genetic association studies. Molecular genetics and genomics. 287(1): 79-86.  
 García-Lor A, Luro F, Morillon X, Ančić G, Luro F, Navarro L, Ollitault P. 2012. Nuclear phylogeny within Citrus (Rutaceae) and four related genera. Annals of Botany. doi:10.1093/aob/abv227.  
 Luro F, Costantino G, Vercellotti L, Ančić G, Lotfy S, Winkler R, Spina M, Ollitault P, Morillon X. 2010. Transferability of the EST-SSRs developed on hybrid clementine (*Citrus clementina* Hort ex Tan.) to other Citrus species and their effectiveness for genetic mapping. BMC Genomics. 9 (287): 1-13.  
 Lotfy S, Froelicher Y, Riberuau AM, Winkler R, Ollitault P. 2005. Development of microsatellite markers of chloroplast genome studies in Citrus (Rutaceae abstract 529-10). 7th International Congress of Plant Molecular Biology, SPAIN 2005, Barcelona Spain, June 23-28.  
 Neelam S, Ding ZH, García A, Morón A, Costantino G, Vercellotti L. 2008. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. Theoretical and Applied Genetics 100: 1355-1366.

## Conclusion

Nuclear and cytoplasmic data confirm previous hypothesis (Nicolosi *et al.*, 2000) for some limes and lemons and allow proposing new origin hypothesis for others :

- Eureka, Lisbon and sweet lemons as well as Marrakech lime result from sour orange x citron hybridization,
- *Citrus bergamia* is probably an hybrid between sour orange and yellow lemon (Eureka type) and Ponderosa a cross between two (pummelo x citron) hybrids,
- Volkamer lemon, Rangpur lime and Rough lemon should be direct hybrids between a mandarin from the acidic group and citron,
- Mexican and Kalpi lime should result from *C. micrantha* x citron hybridization,
- Alemow should be a (*C. micrantha* x citron) x citron backcross,
- The origin of Karina and Palestine sweet lime and Meyer lemon appears more complex,
- Future multilocus SNPs analysis of a larger number of sites for phylogenetic assignment of DNA fragments will allow to decipher the interspecific mosaic genome structure of this particular *Citrus* group and will give more answers about his origins.



## Annexe I.5

### Comparative values of SSRs, SNPs and InDels for citrus genetic diversity analysis

Ollitrault P.<sup>1</sup>, Garcia-Lor A.<sup>2</sup>, Terol J.<sup>3</sup>, Curk F.<sup>4</sup>, Ollitrault F.<sup>2</sup>, Talon M.<sup>3</sup>, and Navarro L.<sup>2</sup>

<sup>1</sup>Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), BIOS, France; <sup>2</sup>Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Protección Vegetal y Biotecnología, Spain; <sup>3</sup>Instituto Valenciano de Investigaciones Agrarias (IVIA), Centro de Genómica, Spain; and <sup>4</sup>Institut National de la Recherche Agronomique (INRA), GAP, France.

#### Abstract

SSRs have long been considered as almost ideal markers for genetic diversity analysis. With the increasing availability of sequencing data, SNPs and Indels become major classes of codominant markers with genome wide coverage. We have analyzed the respective values of SSRs, Indels, and SNPs for intra and interspecific Citrus genetic diversity analysis. Moreover, we have compared the diversity structure revealed by markers mined in a single heterozygous genotype (the clementine) and markers mined in a large interspecific survey. A random set of 25 markers was selected for each marker class to genotype 48 citrus accessions. SSRs were the most polymorphic markers at the intraspecific level allowing complete varietal differentiation within basic taxa (*Citrus reticulata*, *Citrus maxima*, *Citrus medica*). However, SSRs gave the lowest values for interspecific differentiation, followed by SNPs and InDels, that displayed low intraspecific variability but high interspecific differentiation. A clear effect of the discovery panel was observed for SNPs and Indels. The ascertainment biases associated with the clementine heterozygosity mining resulted mainly in an over estimation of within *C. reticulata* diversity and an underestimation of the interspecific differentiation. Therefore SSRs are very useful for intraspecific structure analysis while SNPs and InDels mined in large discovery panel will be more powerful to decipher the interspecific mosaic structure of secondary cultivated species.

In press, Acta Horticulturae, Proceedings XII International Citrus Congress 2012

## Annexe I.6

## A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the ‘true citrus fruit trees’ group (Citrinae, Rutaceae) and the origin of cultivated species

Andres Garcia-Lor<sup>1</sup>, Franck Curk<sup>1,2</sup>, Hager Snoussi-Trifa<sup>3</sup>, Raphael Morillon<sup>4</sup>, Gema Ancillo<sup>1</sup>, François Luro<sup>2</sup>, Luis Navarro<sup>1</sup>, and Patrick Ollitrault<sup>1,4</sup>

<sup>1</sup>Centro de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), 46113 Moncada (Valencia), Spain; <sup>2</sup>INRA, UR1103 Génétique et Ecophysiologie de la Qualité des Agrumes, F-20230 San Giuliano, France; <sup>3</sup>Horticultural Laboratory, Tunisian National Agronomic Research Institute (INRAT), Rue Hedi Karray, 2049 Ariana, Tunisia; <sup>4</sup>UMR AGAP, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), TA A-108/02, 34398 Montpellier, Cedex 5, France

### Abstract

**Background and Aims** Despite differences in morphology, the genera representing ‘true citrus fruit trees’ are sexually compatible, and their phylogenetic relationships remain unclear. Most of the important commercial ‘species’ of *Citrus* are believed to be of interspecific origin. By studying polymorphisms of 27 nuclear genes, the average molecular differentiation between species was estimated and some phylogenetic relationships between ‘true citrus fruit trees’ were clarified.

**Methods** Sanger sequencing of PCR-amplified fragments from 18 genes involved in metabolite biosynthesis pathways and nine putative genes for salt tolerance was performed for 45 genotypes of *Citrus* and relatives of *Citrus* to mine single nucleotide polymorphisms (SNPs) and indel polymorphisms. Fifty nuclear simple sequence repeats (SSRs) were also analysed.

**Key Results** A total of 16 238 kb of DNA was sequenced for each genotype, and 1097 single nucleotide polymorphisms (SNPs) and 50 indels were identified. These polymorphisms were more valuable than SSRs for inter-taxon differentiation. Nuclear phylogenetic analysis revealed that *Citrus reticulata* and *Fortunella* form a cluster that is differentiated from the clade that includes three other basic taxa of cultivated citrus (*C. maxima*, *C. medica* and *C. micrantha*). These results confirm the taxonomic subdivision between the subgenera *Metacitrus* and *Archicitrus*. A few genes displayed positive selection patterns within or between species, but most of them displayed neutral patterns. The phylogenetic inheritance patterns of the analysed genes were inferred for commercial *Citrus* spp.

**Conclusions** Numerous molecular polymorphisms (SNPs and indels), which are potentially useful for the analysis of interspecific genetic structures, have been identified. The nuclear phylogenetic network for *Citrus* and its sexually compatible relatives was consistent with the geographical origins of these genera. The positive selection observed for a few genes will help further works to analyse the molecular basis of the variability of the associated traits. This study presents new insights into the origin of *C. sinensis*.

**Key words:** Phylogeny, evolution, SNP, indel, SSR, *Rutaceae*, *Citrus*, *Fortunella*, *Microcitrus*, *Eremocitrus*, *Poncirus*

Annals of Botany (2013) 111 (1): 1-19.; doi: 10.1093/aob/mcs227; First published online: October 26, 2012

# Annexe II

---

Supplementary information Chapitre I

Table S.I.1. Varieties by common horticultural group and scientific names

Common horticultural group name	Swingle scientific name	Common names	ID #	Amplicon library test		
Pummelos	<i>Citrus maxima</i> (Burm.) Merr.	Chandler pummelo	IVIA207	X		
	<i>Citrus maxima</i> (Burm.) Merr.	Da Xanh pummelo	IVIA589			
	<i>Citrus maxima</i> (Burm.) Merr.	Deep Red pummelo	IVIA277			
	<i>Citrus maxima</i> (Burm.) Merr.	Eingedi pummelo	SRA610			
	<i>Citrus maxima</i> (Burm.) Merr.	Kao Pan pummelo	SRA321			
	<i>Citrus maxima</i> (Burm.) Merr.	Nam Roi pummelo	IVIA590			
	<i>Citrus maxima</i> (Burm.) Merr.	Pink pummelo	IVIA275	X		
	<i>Citrus maxima</i> (Burm.) Merr.	Sans pépins pummelo	SRA710			
	<i>Citrus maxima</i> (Burm.) Merr.	Tahitian pummelo	SRA727			
	<i>Citrus maxima</i> (Burm.) Merr.	Timor pummelo	SRA707			
Ancestral groups		<i>Citrus reticulata</i> Blanco	Dancy mandarin	IVIA434		
		<i>Citrus reticulata</i> Blanco	Imperial mandarin	IVIA576		
		<i>Citrus reticulata</i> Blanco	Ponkan mandarin	IVIA482		
		<i>Citrus reticulata</i> Blanco	Sun Chu Cha mandarin	IVIA483		
		<i>Citrus reticulata</i> Blanco	Temple mandarin	IVIA81		
		<i>Citrus reticulata</i> Blanco	Willowleaf mandarin	IVIA154	X	
	Mandarins		<i>Citrus reticulata</i> Blanco	Beauty of Glen Retreat mandarin	SRA261	
			<i>Citrus reticulata</i> Blanco	Clausellina satsuma mandarin	IVIA19	
			<i>Citrus reticulata</i> Blanco	Shekwasha mandarin	IVIA238	
			<i>Citrus reticulata</i> Blanco	King mandarin	IVIA477	
			<i>Citrus reticulata</i> Blanco	Cleopatra mandarin	IVIA385	X
			<i>Citrus reticulata</i> Blanco	Sunki mandarin	IVIA239	
			<i>Citrus tachibana</i> (Mak.) Tan.	Fuzhu mandarin	IVIA571	
	<i>Citrus tachibana</i> (Mak.) Tan.	Nanfengmijuu mandarin	SRA839			
Citrons		<i>Citrus medica</i> L.	Buddha's Hand citron	IVIA202		
		<i>Citrus medica</i> L.	Corsican citron	IVIA567	X	
		<i>Citrus medica</i> L.	Diamante citron	IVIA560		
		<i>Citrus medica</i> L.	Etrog citron	IVIA169	X	
		<i>Citrus medica</i> L.	Humpang citron	SRA722		
		<i>Citrus medica</i> L.	Poncire commun citron	SRA701		
Papedas	<i>Citrus micrantha</i> Wester	Small flower papeda	IVIA626			
Bergamot	<i>Citrus aurantium</i> L.	Castagnaro Bergamot	SRA612			
Lime	<i>Citrus aurantifolia</i> (Christm.) Swing.	Mexican lime	IVIA164			
Alemow	<i>Citrus macrophylla</i> Wester	Alemow	IVIA288			
Sour oranges		<i>Citrus aurantium</i> L.	Bouquet de Fleurs sour orange	IVIA139		
		<i>Citrus aurantium</i> L.	Sevillano sour orange	IVIA117		
Lemons		<i>Citrus limon</i> (L.) Burm.	Eureka lemon	IVIA297		
		<i>Citrus limon</i> (L.) Burm.	Lisbon lemon	IVIA219		
		<i>Citrus limon</i> (L.) Burm.	Sweet lemon	IVIA145		
Hybrids of Lemon		<i>Citrus limon</i> X <i>C. sinensis</i> ?	Meyer lemon	IVIA443		
		<i>Citrus</i> Sp.	Volckamer lemon	IVIA432		
Grapefruits		<i>Citrus paradisi</i> Macf.	Marsh grapefruit	IVIA176		
		<i>Citrus paradisi</i> Macf.	Star Ruby grapefruit	IVIA197		
Clementines		<i>Citrus reticulata</i> Blanco	Caffin clementine	IVIA132		
		<i>Citrus reticulata</i> Blanco	Clemenules clementine	IVIA22	X	
		<i>Citrus reticulata</i> Blanco	Clemenules haploid clementine	IVIA638	X	
Sweet oranges		<i>Citrus sinensis</i> (L.) Osb.	Sanguinelli sweet orange	IVIA34		
		<i>Citrus sinensis</i> (L.) Osb.	Valencia late sweet orange	IVIA363		
		<i>Citrus sinensis</i> (L.) Osb.	Washington navel sweet orange	IVIA222		
Outgroup	<i>Severinia buxifolia</i> (Poir.) Ten.	Chinese box orange	IVIA147			

Table S.I.2. Information of amplicon location (physical and genetic), annotation of genes, and specific primers for Fluidigm amplification

Scaffold	Seq. name	Fragment location in the reference sequence*	Genetic map estimated fragment position within chromosome 2 (cM)**	Gene name*	Functional annotations for this locus*	Primer seq.	Expected Size
2	2P737170	737,169 to 737,620	3	Ciclev10016030m.g	Hsp20/alpha crystallin family	F TCCGTTGCTATTCCCTTCAC R AAAGCAGCCGAACATAAGGA	452
2	2P3068140	3,068,139 to 3,068,628	7	Ciclev10017346m.g	Mitochondrial ATPase inhibitor, IATP	F GCTCCTGTTGTGGATTGCT R AAACCAATGACCTCCCACAG	490
2	2P4517048	4,517,047 to 4,517,551	9	Ciclev10014206m.g	E3 ubiquitin ligase involved in syntaxin degradation	F TCCAACACTCAATGCAAACC R GGCTGCTTCATCAACAGTGA	505
2	2P8108334	8,108,333 to 8,108,880	23	Ciclev10017459m.g	there are no functional annotations for this locus	F GAACACTCCTTTGGGCTCAG R ATGATCAGCCAACAACACCA	548
2	2P11442721	11,442,720 to 11,443,266	49	Ciclev10018363m.g	Plant organelle RNA recognition domain	F CCTCTGCCATTCTTCACTC R CTCACGACTCTCCCAGAAGC	547
2	2P13928427	13,928,426 to 13,928,927	51	Ciclev10015532m.g	Mitochondrial transcription termination factor, mTERF	F CTCTGCCAAAGTCGTGTCAA R TGGCCTCAGCAAAGAAGATT	502
2	2P21022460	21,022,459 to 21,022,996	57	Ciclev10018135m.g	F-box domain	F GGGAAACGATGAATCCGTAA R CCTTCTTCCCATCAAATCCA	538
2	2P25198627	25,198,626 to 25,199,079	68	Ciclev10015267m.g	Regulator of chromosome condensation (RCC1) repeat	F CGATTATGAGCAGAGCCACA R TCTCCAAGCTTTCCAAGA	454
2	2P26819388	26,819,389 to 26,819,923	79	Ciclev10014317m.g	S-locus glycoprotein family	F ACTTTGGGGCTTTTGTGCT R AAAATACCCTCTCGCTGTTC	535
2	2P29538734	29,538,733 to 29,539,273	104	Ciclev10015668m.g	Peptidase family M50	F GGCAGCATCAGCAAGTGATA R TTTGATGGATGGGAGGAGAC	541
2	2P30446231	30,446,230 to 30,446,704	107	Ciclev10016972m.g	Ribonuclease H2 non-catalytic subunit (Ylr154p-like)	F ACCAATGGAAGAGCGAATG R AACAGGGATTGAGGCTGATG	475
2	2P32507721	32,507,720 to 32,508,183	115	Ciclev10016236m.g	there are no functional annotations for this locus	F TACCCTGACGGAATCTGAGG R CCGGTGAATGGCTCTGTAGT	464
2	2P33506778	33,506,777 to 33,507,312	121	Ciclev10016465m.g	VQ motif	F TCGGAGCCTGAAGACTTGTT R TCGTCAAGCTTCCACCTTCT	536
2	2P33532337	33,532,339 to 33,532,930	121	Ciclev10017475m.g	Eukaryotic translation initiation factor 3 subunit 7 (eIF-3)	F CAAAACCAGAACCAGAACC R ATTCTCGCCCTTGTCTTG	593
2	2P35391362	35,391,361 to 35,391,811	130	Ciclev10017141m.g	there are no functional annotations for this locus	F TAGCGAAACCTGAGGGAATG R ATGGCACGGACTTGGTAAAG	541
2	2P36235952	36,235,951 to 36,236,487	135	Ciclev10017604m.g	PPR repeat	F TCTCATGGCCTTCCACTT R GGAAGGTGCAGCTATGAAGC	537

\*<http://www.phytozome.net>

\*\*Ollitrault et al. 2012

Table S.I.3. Multiplex genotype Identifiers (MID) and related genotypes

Mid	Seq	Genotypes	
MID-01	ACGAGTGCGT	Clausellina satsuma mandarin	IVIA19
MID-02	ACGCTCGACA	Sanguinelli sweet orange	IVIA34
MID-03	AGACGCACTC	Sevillan sour orange	IVIA117
MID-05	ATCAGACACG	Mexican lime	IVIA164
MID-06	ATATCGCGAG	Marsh grapefruit	IVIA176
MID-07	CGTGTCTCTA	Buddha's Hand citron	IVIA202
MID-08	CTCGCGTGTC	Lisbon lemon	IVIA219
MID-10	TCTCTATGCG	Clemenules clementine	IVIA22
MID-11	TGATACGTCT	Temple mandarin	IVIA81
MID-13	CATAGTAGTG	Bouquet de fleurs sour orange	IVIA139
MID-14	CGAGAGATAC	Willowleaf mandarin	IVIA154
MID-15	ATACGACGTA	Etrog citron	IVIA169
MID-16	TCACGTAATA	Star Ruby grapefruit	IVIA197
MID-17	CGTCTAGTAC	Chandler pummelo	IVIA207
MID-18	TCTACGTAGC	Washington navel sweet orange	IVIA222
MID-19	TGTACTACTC	Shekwasha mandarin	IVIA238
MID-20	ACGACTACAG	Pink pummelo	IVIA275
MID-21	CGTAGACTAG	Alemow	IVIA288
MID-22	TACGAGTATG	Kao Pan pummelo	SRA321
MID-23	TACTCTCGTG	Cleopatra mandarin	IVIA385
MID-24	TAGAGACGAG	Dancy mandarin	IVIA434
MID-25	TCGTGCTCG	King mandarin	IVIA477
MID-26	ACATACGCGT	Sun Chu Cha mandarin	IVIA483
MID-27	ACGCGAGTAT	Sunki mandarin	IVIA239
MID-28	ACTACTATGT	Deep Red pummelo	IVIA277
MID-30	AGACTATACT	Valencia late sweet orange	IVIA363
MID-31	AGCGTCTGCT	Volckamer lemon	IVIA432
MID-32	AGTACGCTAT	Sweet lemon	IVIA443
MID-33	ATAGAGTACT	Ponkan mandarin	IVIA482
MID-34	CACGCTACGT	Diamante citron	IVIA560
MID-35	CAGTAGACGT	Corsican citron	IVIA567
MID-36	CGACGTGACT	Imperial mandarin	IVIA576
MID-37	TACACACACT	Nam Roi pummelo	IVIA590
MID-38	TACACGTGAT	Castagnaro Bergamot	SRA612
MID-39	TACAGATCGT	Poncire commun citron	SRA701
MID-40	TACGCTGTCT	Sans pépins pummelo	SRA710
MID-42	TCGATCACGT	Beauty of Glen Retreat mandarin	SRA261
MID-43	TCGCACTAGT	Fuzhu mandarin	IVIA571
MID-44	TCTAGCGACT	Da Xanh pummelo	IVIA589
MID-45	TCTATACTAT	Eingedi pummelo	SRA610
MID-46	TGACGTATGT	Small flower papeda	IVIA626
MID-47	TGTGAGTAGT	Timorese pummelo	SRA707
MID-48	ACAGTATATA	Humpang citron	SRA722
MID-49	ACGCGATCGA	Nanfengmijuu mandarin	SRA839
MID-50	ACTAGCAGTA	Caffin clementine	IVIA132
MID-67	TCGATAGTGA	Meyer lemon	IVIA145
MID-68	TCGCTGCGTA	Eureka lemon	IVIA297
MID-69	TCTGACGTCA	Tahitian pummelo	SRA727
MID-98	CTCTACGCTC	Chinese box orange	IVIA147

Table S.I.4. Distribution of read numbers per gene fragment and varieties for the first Fluidifm run

Name	2P737170	2P3068140	2P4517048	2P8108334	2P11442721	2P13928427	2P21022460	2P25198627	2P26819388	2P29538734	2P30446231	2P32507721	2P33506778	2P33532337	2P35391362	2P36235952
Clauselina satsuma mandarin	155	26	37	29	21	62	52	194	5	18	1	295	26	5	100	44
Sanguinelli sweet orange	59	48	47	31	45	51	46	18	144	17	31	67	39	7	47	41
Sevilla sour orange	127	94	91	44	71	79	80	101	15	33	53	162	107	0	117	99
Mexican lime	180	48	18	26	30	74	56	231	81	10	3	427	42	23	154	45
Marsh grapefruit	147	20	26	7	18	45	22	99	107	9	2	247	30	12	74	40
Buddha's hand citron	188	146	119	91	110	125	114	187	118	67	68	216	162	1	189	127
Lisbon Frost lemon	100	21	11	14	12	49	34	153	74	4	1	264	18	17	91	28
Nules clementine	95	90	60	64	56	74	64	75	42	31	61	126	86	0	89	60
Temple mandarin	141	121	97	83	75	103	82	111	133	64	66	270	104	1	114	87
Bouquet de fleurs sour orange	151	108	98	50	68	114	50	126	5	42	65	149	71	0	125	87
Willowleaf mandarin	103	90	82	38	66	89	78	101	66	39	49	257	87	19	113	73
Etrog citron	74	73	54	33	57	70	33	101	143	41	29	104	65	5	80	51
Star Ruby grapefruit	159	116	97	63	68	111	70	104	90	52	60	197	111	7	159	105
Chandler pummelo	120	92	60	60	38	61	51	84	150	53	37	139	103	11	96	82
Washington Frost navel sweet orange	187	49	21	16	15	78	34	65	130	17	7	364	41	12	135	47
Shekwasha mandarin	196	28	20	26	19	64	41	245	28	9	6	378	32	23	129	30
Pink pummelo	199	37	28	23	22	83	42	178	105	12	7	369	36	10	118	46
Alemow	163	50	12	22	17	67	51	229	226	10	4	379	36	16	132	40
Kao Pan pummelo	155	32	10	11	15	67	47	185	164	12	2	409	30	16	111	35
Cleopatra mandarin	173	43	18	18	12	67	29	188	189	14	0	574	25	21	90	28
Dancy Mandarin	104	73	69	39	48	62	54	84	151	50	52	107	66	25	104	51
King Mandarin	102	22	13	2	20	37	21	81	42	11	1	185	27	11	74	31
Sun Chu Sha mandarin	106	81	64	50	48	78	47	82	54	53	39	190	87	12	95	55
Sunki mandarin	30	19	27	6	29	27	12	29	47	10	10	52	17	5	32	12
Seep Red pummelo	163	140	96	76	81	132	107	138	95	67	60	223	114	8	158	130
Valencia Delta sweet orange	124	94	77	58	69	94	58	38	141	27	59	194	112	13	106	91
Volkamer lemon	130	19	15	16	21	54	36	178	60	17	1	605	41	13	102	28
Sweet lemon	62	50	39	30	33	38	31	42	67	21	20	66	31	10	52	37
Ponkan Mandarin	161	43	29	24	22	63	34	156	213	13	7	347	29	23	126	29
Diamante citron	102	24	7	7	9	40	37	146	36	7	0	229	8	15	92	12
Corsican citron	173	48	8	20	19	61	60	283	99	17	4	528	34	9	151	35
Imperial mandarin	130	24	24	22	14	49	34	150	0	7	6	244	36	0	93	31
Nam Roi pummelo	128	18	7	15	8	46	35	146	107	5	0	366	18	8	76	25
Bergamot	168	27	21	22	25	55	25	170	164	10	1	370	29	9	111	42
Poncire citron	158	127	74	67	102	113	106	180	181	66	46	251	123	1	152	79
Sans pépins pummelo	133	37	24	21	22	61	37	130	182	13	2	278	39	20	86	29
Beauty of Glen Retreat mandarin	183	107	125	54	90	130	106	140	99	67	81	287	149	14	149	78
Fuzhu mandarin	69	66	57	12	51	55	51	68	35	35	44	125	54	12	88	36
Da Xanh pummelo	135	23	19	17	15	60	40	88	90	15	5	261	28	8	72	19
Eingedi pummelo	136	33	26	22	18	54	33	93	187	13	11	231	21	10	87	36
Citrus micrantha	206	47	15	30	16	76	50	106	92	15	5	330	37	39	112	38
Timor pummelo	200	28	17	22	19	68	41	168	152	21	1	424	30	15	106	42
Humpang citron	91	26	6	5	15	33	36	89	123	12	4	217	20	0	82	20
Nan Feng Mi Chu mandarin	71	13	18	5	14	32	25	100	107	5	4	182	18	20	67	19
Caffin clementine	178	123	114	69	91	130	100	123	23	51	100	250	141	2	187	124
Meyer lemon	139	100	93	58	69	100	88	112	100	47	44	157	111	0	121	87
Eureka lemon	53	53	47	33	29	54	45	62	119	20	36	67	51	7	55	45
Tahiti pummelo	78	77	57	41	44	58	43	72	104	21	38	114	76	5	82	67

6385 2874 2194 1592 1876 3393 2468 6029 4885 1270 1233 12273 2798 520 5081 2523 57394

>100  
51-100  
20-50  
15-19  
0-14



Table S.I.5. Distribution of read numbers per gene fragment and varieties for the two Fluidifm runs and solutions used for insufficient read number situations

Name	2P737170	2P3068140	2P4517048	2P8108334	2P11442721	2P13928427	2P21022460	2P25198627	2P26819388	2P29538734	2P30446231	2P32507721	2P33506778	2P33532337	2P35391362	2P36235952
Clauselina satsuma mandarin	156	26	335	734	496	62	52	194	5 (Sanger Ho)	32	709	376	743	129	100	332
Sanguinelli sw eet orange	59	48	47	31	45	51	46	18 readable	144	58	31	277	40	7 missing data	47	41
Sevillan sour orange	128	94	318	615	461	79	80	101	15 readable	73	56	351	519	33	117	327
Mexican lime	180	48	61	660	459	74	56	231	81	51	5 (Sanger)	549	559	119	154	547
Marsh grapefruit	147	20	298	788	363	45	22	99	107	68	13 (Sanger)	477	445	44	76	514
Buddha's Hand citron	188	146	221	1013	827	125	114	187	118	98	69	418	581	17 readable	189	691
Lisbon Frost lemon	100	21	11 missing data	14 readable	12 readable	49	34	153	74	66	5 (Sanger)	616	18 readable	124	91	28
Clemenules clementine	95	90	265	431	350	74	64	75	42	66	76	288	495	50	89	250
Temple mandarin	141	121	292	816	558	103	82	111	133	109	78	572	612	46	114	473
Bouquet de fleurs sour orange	151	108	303	587	443	114	50	126	5 missing data	87	68	352	452	36	125	381
Willow leaf mandarin	103	90	83	40	66	89	78	101	66	64	55	471	87	124	113	73
Etrog citron	74	73	73	626	481	70	33	101	143	89	30	297	509	62	80	430
Star Ruby grapefruit	159	116	100	67	74	111	70	104	90	86	72	377	114	42	159	108
Chandler pummelo	120	92	60	60	38	61	51	84	150	114	40	361	103	38	96	82
Washington navel sw eet orange	187	49	33	566	134	78	34	65	130	73	34	655	59	38	135	54
Shekw asha mandarin	196	28	20	26	19 readable	64	41	245	28	54	14 (readable)	662	32	139	129	30
Fnk pummelo	199	37	305	689	543	83	42	178	105	51	27	680	608	27	118	590
Alemow	163	50	37	1323	978	67	51	229	226	88	5 (Sanger)	812	657	192	132	1030
Kao Pan pummelo	155	32	247	1005	1425	67	47	185	164	60	2 (Sanger)	697	259	64	111	551
Cleopatra mandarin	173	43	362	870	603	67	29	188	189	75	4 (Sanger Ho)	941	656	83	90	538
Dancy mandarin	104	73	301	570	461	62	54	84	151	115	61	436	555	61	104	371
King mandarin	102	22	403	10 (Ho)	623	37	21	81	42	74	6 (sanger)	402	479	31	74	408
Sun Chu Sha mandarin	106	81	360	802	584	78	47	82	54	109	45	511	670	58	95	523
Sunki mandarin	30	19 (Ho)	28	9 missing data	31	27	12 (Ho)	29	47	49	16 readable	279	19 readable Ho	100	32 (readable Ho)	12
Deep Red pummelo	163	140	527	1124	471	132	107	138	95	133	63	451	493	31	158	711
Valencia late sw eet orange	124	94	78	58	72	94	58	38	141	90	78	452	114	108	106	92
Volckamer lemon	130	19 readable	300	725	627	54	36	178	60	108	2 (sanger)	1029	610	140	102	518
Sw eet lemon	62	50	39	31	34	38	31	42	67	78	22	352	31	109	52	37
Ponkan mandarin	161	43	312	800	672	63	34	156	213	82	12 (Ho)	589	442	42	126	379
Diamante citron	102	24	111	82	42	40	37	146	36	106	0 (Sanger)	598	343	34	92	257
Corsican citron	173	48	15 Missing data	730	602	61	60	283	99	88	6 (Sanger Ho)	835	578	37	151	562
Imperial mandarin	130	24	324	500	376	49	34	150	0 (Sanger Ho)	48	16 (readable)	484	537	50	93	293
Nam Roi pummelo	128	18 readable	383	437	320	46	35	146	107	53	8 (Sanger)	689	550	67	76	300
Bergamot	168	27	335	702	462	55	25	172	164	65	3 (Sanger)	661	552	95	111	423
Poncire citron	158	127	75	68	102	113	106	180	181	168	48	650	124	13 readable ho	152	82
Sans pépins pummelo	133	37	456	671	561	61	37	130	182	88	4 (Sanger)	502	652	48	86	430
Beauty of Glen Retreat mandarin	183	107	390	671	486	130	106	140	99	122	83	646	645	21	149	361
Fuzhu mandarin	69	66	57	14 missing data	52	55	51	68	35	93	55	391	55	41	88	37
Da Xanh pummelo	135	23	50	978	92	60	40	88	90	64	15 (readable)	505	34	46	72	22
Engedi pummelo	136	33	327	505	309	54	33	93	187	116	38	556	532	28	87	341
Citrus micrantha	206	47	303	451	289	76	50	106	92	102	12 readable	663	75	155	112	536
Timorese pummelo	200	28	539	1870	926	68	41	168	152	21	4 (Sanger Ho)	427	475	18 readable	106	1383
Humpang citron	91	26	182	258	393	33	36	89	123	93	5 (Sanger Ho)	507	582	1 Missing data	82	316
Nanfengmijuu mandarin	71	13 (Ho)	284	442	300	32	25	100	107	41	10 (readable)	499	405	48	67	242
Caffin clementine	178	123	436	836	577	130	100	123	23	99	108	568	436	51	187	277
Meyer lemon	139	100	407	699	466	100	88	112	100	153	48	429	592	93	121	506
Eureka lemon	53	53	49	862	874	54	45	62	119	84	36	343	1982	180	55	1176
Tahitian pummelo	78	77	697	41	44	58	43	72	104	21	38	114	76	5 (missing data)	82	67

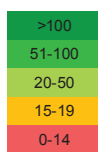






Table S.I.7. Observed heterozygosity (Ho) of secondary species

Gene fragment	Oranges	Sour Oranges	Clem.	Grapefruits	All Lemons	Regular and 'Sweet' lemons	Meyer' lemon	Volckamer lemon	Mexican' lime	Alemow	Bergamot
2P737170	0.1	0.1	0	0.2	0.5	0.5	0.5	0.5	0.5	0.5	0.1
2P3068140	0.5	0.5	0	0.5	0.5	0.46	0.5	0.6	0.4	0.4	0.62
2P4517048	0	0.2	0.2	0.3	0.2	0.25	0	0.3	0.2	0.2	0
2P8108334	0.3	0.3	0	0.2	0.2	0.26	0.2	0.3	0.3	0.1	0.21
2P11442721	0.2	0.3	0.3	0.1	0.2	0.14	0.2	0.3	0.2	0.2	0.14
2P13928427	0.1	0.1	0	0	0.4	0.35	0.3	0.5	0.5	0.4	0.35
2P21022460	0.2	0	0	0.2	0.4	0.36	0.4	0.4	0.6	0.5	0.36
2P25198627	0.3	0.3	0	0.2	0.3	0.25	0.4	0.4	0.3	0.3	0.25
2P26819388	0.1	0.1	0.1	0	0.2	0.18	0.3	N	0.3	0.2	0.23
2P29538734	0.2	0.3	0.3	0.2	0.4	0.37	0.5	0.5	0.3	0.3	0.29
2P30446231	0.1	0.2	0.2	0.2	0.2	0.23	0.2	0.2	0.4	0.2	0.15
2P32507721	0	0	0.1	0	0.3	0.09	0.4	0.7	0.2	0.4	0.09
2P33506778	0.2	0.5	0.3	0	0.5	0.5	0.5	0.5	0.5	0.5	0.33
2P33532337	0.2	0.6	0.6	0	0.4	0.44	0	0.4	0.2	0.2	0
2P35391362	0.1	0.6	0.5	0	0.6	0.56	0.6	0.6	0.3	0.3	0.56
2P36235952	0	0.6	0.5	0	0.4	0.46	0.4	0.4	0.1	0.1	0.46
Total	0.2	0.3	0.2	0.1	0.3	0.33	0.3	0.4	0.3	0.3	0.26
SD	0.36	0.44	0.40	0.33	0.43	0.47	0.47	0.49	0.46	0.44	0.44
CI	0.04	0.05	0.04	0.04	0.05	0.05	0.05	0.06	0.05	0.05	0.05

SD: standard deviation; CI: confidence interval estimated with alpha = 0.05; N: missing data; Clem.: Clementine

Table S.I.8. Average values of the Ten Structure runs at K = 4 for each cluster of each variety (confidence interval estimated with alpha = 0.05)

	<i>C. maxima</i>	<i>C. medica</i>	<i>C. reticulata</i>	<i>C. micrantha</i>
Cleopatra mandarin	0.000±0.000	0.000±0.000	0.999±0.000	0.000±0.000
Sun Chu Sha mandarin	0.000±0.000	0.000±0.000	0.999±0.000	0.000±0.000
Sunki mandarin	0.000±0.000	0.000±0.000	0.999±0.000	0.000±0.000
Satsuma mandarin	0.001±0.000	0.000±0.000	0.997±0.000	0.000±0.000
Nanfengmijuu mandarin	0.001±0.000	0.001±1.343	0.994±0.000	0.003±0.000
Shekwasha mandarin	0.001±0.000	0.002±2.687	0.955±0.001	0.041±0.001
Willowleaf mandarin	0.041±0.001	0.001±0.000	0.956±0.000	0.001±1.343
Beauty mandarin	0.047±4.300	0.000±0.000	0.952±0.000	0.001±1.343
Dancy mandarin	0.187±0.000	0.000±0.000	0.811±0.000	0.000±0.000
Fuzhu mandarin	0.227±0.000	0.000±0.000	0.771±0.000	0.001±1.343
King mandarin	0.256±0.000	0.001±1.343	0.742±0.000	0.001±1.343
Temple mandarin	0.274±0.000	0.000±0.000	0.724±0.000	0.000±0.000
Imperial mandarin	0.294±0.000	0.001±1.343	0.703±0.000	0.001±1.343
Ponkan mandarin	0.336±0.000	0.000±0.000	0.662±0.000	0.001±1.343
Chandler pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Da Xanh pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Deep Red pummelo	0.998±0.000	0.001±1.343	0.000±0.000	0.001±1.343
Eingedi pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Kao Pan pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Pink pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Tahitian pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Timorese pummelo	0.999±0.000	0.000±0.000	0.000±0.000	0.000±0.000
Sans Pépins pummelo	0.998±0.000	0.001±1.343	0.000±0.000	0.000±0.000
Nam Roi pummelo	0.994±0.000	0.003±0.000	0.001±1.343	0.001±0.000
Buddha's hand citron	0.000±0.000	0.999±0.000	0.000±0.000	0.000±0.000
Etrog citron	0.000±0.000	0.999±0.000	0.000±0.000	0.000±0.000
Diamante citron	0.000±0.000	0.999±0.000	0.000±0.000	0.000±0.000
Poncire citron	0.001±1.343	0.998±0.000	0.000±0.000	0.001±1.343
Humpang citron	0.000±0.000	0.998±0.000	0.000±0.000	0.001±1.343
Corsican citron	0.009±0.000	0.977±0.000	0.002±0.000	0.009±0.000
<i>Citrus micrantha</i>	0.001±1.343	0.001±1.343	0.000±0.000	0.998±0.000
Alemow	0.004±0.000	0.476±0.000	0.001±1.343	0.518±0.000
Mexican lime	0.001±1.343	0.517±0.000	0.000±0.000	0.481±0.000
Haploid clementine	0.000±0.000	0.000±0.000	0.999±0.000	0.000±0.000
Caffin clementine	0.314±0.000	0.000±0.000	0.684±0.000	0.001±1.343
Clemenules clementine	0.314±0.000	0.000±0.000	0.684±0.000	0.001±1.343
Bouquet de fleurs sour orange	0.493±0.000	0.002±0.000	0.501±0.000	0.003±0.000
Sevillan sour orange	0.493±0.000	0.002±0.000	0.501±0.000	0.003±0.000
Sanguinelli sweet orange	0.826±6.881	0.000±0.000	0.173±1.720	0.001±1.343
Valencia late sweet orange	0.826±6.881	0.000±0.000	0.173±1.720	0.001±1.343
Washington navel sweet orange	0.825±0.000	0.000±0.000	0.173±0.000	0.001±1.343
Eureka lemon	0.046±0.003	0.547±0.001	0.404±0.001	0.001±1.343
Lisbon lemon	0.046±0.003	0.547±0.001	0.404±0.001	0.001±1.343
Sweet lemon	0.046±0.003	0.547±0.001	0.404±0.001	0.001±1.343
Meyer lemo	0.001±0.000	0.492±0.000	0.506±0.000	0.001±1.343
Volkamer lemon	0.001±1.343	0.469±0.000	0.511±0.000	0.018±0.000
Bergamot	0.313±0.000	0.477±0.000	0.208±0.000	0.001±1.343
Marsh grapefruit	0.996±0.000	0.000±0.000	0.002±0.000	0.001±1.343
Star Ruby grapefruit	0.996±0.000	0.000±0.000	0.002±0.000	0.001±1.343

Table S.I.9. observed inherited haplotypic structure of secondary species

	Clementine		Lemons		Mexican lime		Grapefruits		Bergamot		Alemow		Meyer lemon					
	Willowleaf mandarin haplotype	Sweet orange haplotype	Sour orange haplotype	<i>C. medica</i> haplotype	<i>C. medica</i> haplotype	<i>C. micrantha</i> haplotype	<i>C. maxima</i> haplotype	Sweet orange haplotype	Sour orange haplotype	Lemons haplotype	<i>C. medica</i> haplotype	<i>C. micrantha</i> haplotype	Sweet orange haplotype	Lemons haplotype	Sour orange haplotype	<i>C. medica</i> haplotype	<i>C. reticulata</i> haplotype	<i>C. maxima</i> haplotype
2P737170	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-
2P3068140	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓	-
2P4517048	✓	✓	✓	✓	✓	✓	✓	2	-	✓	✓	-	-	-	-	-	2	-
2P8108334	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	-	✓	✓	-	-
2P11442721	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	-	✓
2P13928427	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	-	✓
2P21022460	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	3	3
2P25198627	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	-
2P26819388	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	1	-	-	✓	✓	-	✓	-
2P29538734	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	-	✓	✓	-
2P30446231	✓	✓	✓	✓	✓	✓	✓	✓	✓	1	✓	-	-	-	✓	✓	✓	-
2P32507721	✓	✓	✓	✓	✓	✓	1	✓	✓	✓	✓	-	✓	-	✓	✓	✓	-
2P33506778	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	-
2P33532337	✓	✓	✓	✓	✓	✓	✓	2	-	✓	✓	-	-	-	-	2	-	-
2P35391362	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	-
2P36235952	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	✓	-	✓	✓	✓	-
	50%	50%	50%	50%	50%	50%	47%	50%	56%	44%	44%	47%	6%	28%	16%	44%	44%	9%

1: Unidentified isolated haplotype or insufficient structuration for allowing inferring phylogenetic origin;

2: Homozygous secondary species (possible null allele); 3: missing data

#### Haplotype sharing between some secondary species and their supposed parents

Hypotheses of the origin of secondary species were proposed in previous molecular studies. We have checked for each gene fragment if the haplotypic structure of these secondary species could be inherited from the supposed parents. Several molecular studies (Barkley *et al.*, 2006; Nicolosi *et al.*, 2000) suggested that clementine resulted from hybridization between 'Willowleaf' mandarin and sweet orange, regular lemons between sour orange and citron, and 'Mexican' lime between *C. micrantha* and *C. medica*. These three secondary species shared 50% of their haplotypes with each of the supposed parents. Grapefruit is generally accepted as a hybrid between pummelo and sweet orange (Garcia-Lor *et al.*, 2013; Nicolosi *et al.*, 2000; Ollitrault and Navarro, 2012), but one of the analysed gene fragments is of unidentified origin; for all other fragments one haplotype was shared with sweet orange and the other was part of the *C. maxima* cluster. In 1811, Galesio (1811) proposed that bergamot was a hybrid between lemon and sour orange, but later molecular studies gave rise to contradictory hypotheses. For 14 gene fragments one haplotype was shared with sour orange and the other one with regular lemon. Two gene fragments have been found to be homozygous (ma1/ma1 for the 2P4517048 gene fragment

and ma2/ma2 for 2P33532337), both haplotypes shared with sour orange. It is possible that PCR competition led to the un-amplification of the lemon haplotype in this specific genotype. 'Alemow' was proposed to be a direct hybrid between a papeda closely related with *C. micrantha* and a citron (Ollitrault and Navarro, 2012). It displayed one haplotype from the *C. micrantha* cluster and one from *C. medica* for 13 gene fragments out of 16. For five gene fragments the 'Alemow' haplotype of the *C. micrantha* cluster was different from the analysed *C. micrantha* ones. For two gene fragments one of the haplotype was in a phylogenetically undetermined cluster, and the second one was associated with the *C. medica* or *C. micrantha* clusters. For the last gene fragment it displayed only one haplotype within the *C. medica* cluster. Several origin hypotheses have been proposed for 'Meyer' lemon. Chen (1991), according to GOT-2 banding patterns, proposed that it was a hybrid of lime and lemon. Gulsen and Roose (2001) proposed that it has a different pummelo as its female grandparent than lemon. It has also been suggested that it was a sweet orange hybrid (Scora *et al.*, 1982). According to our data, 'Meyer' lemon is neither a sweet orange nor a sour orange hybrid. It shared *C. medica* haplotypes for 44% of the gene fragments; the two missing *C. medica* fragments correspond to a situation where 'Meyer' lemon displayed homozygous haplotype. Therefore it is possible that for these fragments PCR competition leads to un-amplification of the *C. medica* haplotype, as we already supposed in the case of bergamot for the same gene fragments. In this case, the results are consistent with the hypothesis that 'Meyer' lemon is a direct hybrid of *C. medica*. In this hypothesis the other parent should have an interspecific *C. reticulata* / *C. maxima* genomic structure because respectively 10 and three *C. reticulata* and *C. maxima* fragments were identified.

## References

- Barkley, N.A. Roose, M.L. Krueger, R.R. and Federici, C.T. 2006. Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor.Appl.Genet.* 112:8:1519-1531.
- Chen, L.G. Omura, M. and Hidaka, T. 1991. A study on the taxonomy of citrus with GOT isozymes. *Acta Horticulturae Sinica* 18:1:27-32.
- Gallesio, G. 1811. *Traité du citrus*. Chez Louis Fantin Libraire, Paris.
- Garcia-Lor, A. Curk, F. Snoussi-Trifa, H. Morillon, R. Ancillo, G. Luro, F. Navarro, L. and Ollitrault, P. 2013. A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the 'true citrus fruit trees' group (Citrinae, Rutaceae) and the origin of cultivated species. *Ann.Bot.* 111:1:1-19.
- Gulsen, O. and Roose, M.L. 2001. Chloroplast and nuclear genome analysis of the parentage of lemons. *J.Am.Soc.Hort.Sci.* 126:2:210-215.
- Nicolosi, E. Deng, Z.N. Gentile, A. La Malfa, S. Continella, G. and Tribulato, E. 2000. Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theoretical and Applied Genetics* 100:1155-1166.
- Ollitrault, P. and Navarro, L. 2012. Citrus. p. 623-662. In: M. Badenes and D.e. Byrne (eds.), *Fruit Breeding*, Springer New York, London.
- Scora, R.W. Kumamoto, J. Soost, R.K. and Nauer, E.M. 1982. Contribution to the origin of the grapefruit *Citrus paradisi* (Rutaceae). *Syst.Bot.* 7:170-177.

Table S.I.10. Haplotypic structure of each accession.

Seq. name	Mandarins	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>	<i>C. reticulata</i>
		Willowleaf mandarin	Cleopatra mandarin	Dancy mandarin	Fuzhu mandarin	Imperial mandarin	King mandarin	Nanfengmijuu mandarin	Ponkan mandarin	Clauselina satsuma mandarin	Shekwasha mandarin	Sun Chu Sha mandarin	Sunki mandarin	Temple mandarin	Beauty mandarin
2P737170	5 (4)	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re3	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1
2P3068140	5 (4)	re1 / ma1	re1 / re1	re1 / re2	re1 / re2	re1 / re1	re4 / re4	re1 / re1	re1 / re2	re1 / re1	re1 / re3	re1 / re1	re1 / re1	re1 / re1	re1 / re2
2P4517048	5 (5)	re1 / re3	re2 / re4	re2 / re3	re2 / re3	re1 / re3	re1 / re2	re2 / re2	re1 / re3	re2 / re2	re1 / re2	re1 / re2	re2 / re5	re2 / re2	re2 / re3
2P8108334	10 (10)	re2 / re10	re2 / re6	re2 / re2	re2 / re4	re10 / re10	re5 / re5	re2 / re7	re8 / re8	re2 / re2	re3 / re9	re2 / re2	re1 / re2	re2 / re2	re2 / re2
2P11442721	10 (8)	re7 / re8	re1 / re3	re5 / re6	re5 / re6	re6 / re8	re2 / re6	re4 / re8	re5 / re6	re1 / ma1	re3 / ma4	re1 / re7	re3 / re7	re8 / ma4	re7 / re8
2P13928427	5 (4)	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re1	re1 / re3	re1 / re1	re1 / ma3	re1 / re2	re1 / re1	re1 / re1	re1 / ma1	re1 / re1
2P21022460	4 (2)	re1 / ma-re-mi	ma-re-mi / ma-re-mi	re1 / ma-re-mi	re1 / ma-re-mi	re1 / re1	ma-re-mi / ma-re-mi	ma-re-mi / ma-re-mi	re1 / ma-re-mi	re1 / ma1	ma-re-mi / re2	ma-re-mi / ma-re-mi	ma-re-mi / ma-re-mi	re1 / re1	ma-re-mi / ma-re-mi
2P25198627	7 (6)	re1 / re3	re1 / re4	re-ma / re-ma	re-ma / re-ma	re1 / ma1	re1 / re1	re1 / re1	re2 / ma1	re1 / re1	re1 / re5	re1 / re5	re1 / re1	re-ma / re-ma	re-ma / re-ma
2P26819388	12 (10)	re1 / re2	re2 / re5	?	?	re1 / re1	ma1 / ma-re	re8 / re9	ma1 / ma1	re6 / re6	re3 / re3	?	re4 / re7	ma2 / re1	re1 / re1
2P29538734	10 (7)	re5 / re7	re2 / re5	re5 / ma3	re5 / ma3	re7 / ma3	re2 / ma1	re5 / re7	re1 / ma3	re5 / re5	re3 / re5	re2 / re4	re2 / re5	re1 / ma4	re1 / re2
2P30446231	6 (5)	re1 / re1	re2 / re2	re5 / ma1	re5 / ma1	re1 / ma1	re2 / ma1	re1 / re4	ma1 / ma1	re1 / re1	re1 / re1	re2 / re2	re1 / re2	re2 / ma1	re2 / re5
2P32507721	6 (3)	ma-re1 / re1	re1 / re1	re3 / ma1	re3 / ma1	ma1 / ma-re1	ma1 / ma-re1	re2 / re3	ma1 / ma1	re2 / re2	ma-re1 / re3	re1 / re3	re1 / re3	re1 / ma-re1	re1 / ma1
2P33506778	4 (2)	re1 / re1	re1 / re1	re1 / ma1	re1 / ma1	re1 / ma1	re1 / ma1	re1 / re1	ma1 / ma2	re1 / re1	re1 / re2	re1 / re1	re1 / re1	re1 / ma1	re1 / ma2
2P33532337	4 (2)	re1 / re2	re1 / re1	re1 / re1	re1 / ma2	re1 / ma2	re1 / ma2	re1 / re2	re1 / ma2	re2 / re2	re2 / re2	re1 / re2	re2 / re2	re2 / ma2	re2 / ma1
2P35391362	4 (3)	re2 / re2	re1 / re1	re1 / ma3	re1 / ma3	re2 / ma3	re2 / ma3	re2 / re2	re1 / ma3	re2 / re2	re1 / re3	re1 / re2	re1 / re2	re1 / ma3	re1 / re1
2P36235952	6 (6)	re2 / re2	re3 / re4	re4 / re4	re4 / re4	re2 / re2	re2 / re2	re2 / re4	re1 / re1	re4 / re4	re5 / re6	re2 / re4	re4 / re4	re2 / re2	re1 / re2



Table S.I.10. continued

		<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>	<i>C. maxima</i>
	Pummelos	Chandler pummelo	Da Xanh pummelo	Deep red pummelo	Eingedi pummelo	Kao Pan pummelo	Nam Roi pummelo	Pink pummelo	Sans Pépins pummelo	Tahitian pummelo	Timorese pummelo	
Seq. name												
2P737170	6 (6)	ma1 / ma2	ma4 / ma5	ma1 / ma3	ma1 / ma3	ma1 / ma2	ma1 / ma3	ma1 / ma2	ma4 / ma5	ma3 / ma3	ma4 / ma6	
2P3068140	2 (2)	ma1 / ma1	ma1 / ma1	ma1 / ma2	ma1 / ma1	ma1 / ma1	ma1 / ma2	ma1 / ma1	ma1 / ma1	ma1 / ma2	ma1 / ma1	
2P4517048	3 (3)	ma1 / ma2	ma1 / ma1	ma1 / ma1	ma2 / ma3	ma1 / ma2	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma3	
2P8108334	7 (7)	ma2 / ma6	ma1 / ma4	ma1 / ma7	ma2 / ma6	ma2 / ma7	ma1 / ma5	ma2 / ma7	ma1 / ma7	ma2 / ma6	ma1 / ma7	
2P11442721	5 (5)	ma1 / ma2	ma1 / ma4	ma5 / ma5	ma3 / ma5	ma1 / ma1	ma1 / ma2	ma1 / ma1	ma1 / ma4	ma3 / ma4	ma1 / ma4	
2P13928427	3 (3)	ma1 / ma1	ma1 / ma1	ma1 / ma2	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma1	
2P21022460	3 (2)	ma1 / ma- re-mi	ma1 / ma2	ma- re-mi / ma2	ma- re-mi / ma2	ma1 / ma- re-mi	ma1 / ma- re-mi	ma1 / ma- re-mi	ma1 / ma- re-mi	ma1 / ma- re-mi	ma1 / ma- re-mi	
2P25198627	3 (1)	ma1 / re-ma	ma1 / re-ma	ma1 / re-ma	ma1 / re1	re1 / re-ma	ma1 / re1	re-ma / re1	ma1 / re1	ma1 / ma1	ma1 / re-ma	
2P26819388	2 (2)	ma1 / ma2	ma1 / ma1	ma1 / ma2	ma1 / ma1	ma1 / ma2	ma1 / ma1	ma1 / ma2	ma1 / ma2	ma1 / ma2	ma1 / ma1	
2P29538734	6 (6)	ma3 / ma3	ma3 / ma5	ma5 / ma6	ma1 / ma5	ma3 / ma5	ma3 / ma5	ma3 / ma5	ma5 / ma6	ma2 / ma2	ma5 / ma5	
2P30446231	8 (7)	ma2 / ma8	ma4 / ma4	ma2 / ma8	ma3 / ma4	ma7 / ma8	ma3 / ma7	ma5 / ma5	ma2 / ma7	ma3 / ma7	ma7 / ma7	
2P32507721	3 (2)	ma1 / ma- re1	ma1 / ma1	ma1 / ma- re1	ma1 / ma- re1	ma1 / ma- re1	ma1 / ma- re1	ma1 / ma- re1	ma1 / ma- re1	ma- re1 / ma- re1	ma1 / ma- re1	
2P33506778	2 (2)	ma2 / ma2	ma1 / ma1	ma2 / ma2	ma2 / ma2	ma2 / ma2	ma1 / ma2	ma2 / ma2	ma1 / ma2	ma2 / ma2	ma1 / ma2	
2P33532337	2 (2)	ma1 / ma2	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma1 / ma2	ma2 / ma2	ma1 / ma1	ma1 / ma1	ma1 / ma1	ma2 / ma2	
2P35391362	2 (2)	ma2 / ma2	ma2 / ma3	ma2 / ma2	ma2 / ma3	ma2 / ma2	ma2 / ma2	ma2 / ma2	ma2 / ma3	ma2 / ma3	ma2 / ma2	
2P36235952	8 (8)	ma1 / ma2	ma2 / ma5	ma1 / ma2	ma4 / ma7	ma1 / ma5	ma6 / ma8	ma1 / ma5	ma1 / ma2	ma4 / ma6	ma1 / ma2	

Table S.I.10. continued

Seq. name	Citrons	<i>C. medica</i>						<i>C. micrantha</i>	<i>C. clementina</i>
		Corsican citron	Diamante citron	Etrog citron	Humpang citron	Buddha's hand citron	Ponicre citron		
2P737170	2 (2)	me2 / me2	me1 / me2	me1 / me1	me2 / me2	me1 / me1	me1 / me2	2 (2)	re1
2P3068140	1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	1	re1
2P4517048	2 (2)	me1 / me2	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	1	re3
2P8108334	3 (3)	me1 / me2	me2 / me2	me2 / me2	me3 / me3	me2 / me2	me1 / me2	2 (2)	re2
2P11442721	2 (2)	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me2	1	re8
2P13928427	2 (2)	?	me1 / me2	me1 / me1	me1 / me1	me2 / me2	me1 / me2	2 (2)	re1
2P21022460	2 (2)	me1 / me1	me1 / me1	me1 / me1	me2 / me2	me2 / me2	me1 / me1	2 (1)	re1
2P25198627	3 (3)	me1 / me2	me2 / me3	me1 / me1	me3 / me3	me1 / me2	me1 / me2	2 (2)	re1
2P26819388	2 (2)	me1 / me1	me1 / me1	me1 / me2	me1 / me4	me1 / me2	me1 / me2	2 (2)	re1
2P29538734	4 (4)	me2 / me2	me1 / me1	me2 / me2	me4 / me4	me2 / me2	me1 / me1	2 (2)	re5
2P30446231	3 (3)	me2 / me2	me1 / me2	me1 / me1	me1 / me1	me1 / me1	me1 / me1	2 (2)	re1
2P32507721	1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	2 (2)	ma- re1
2P33506778	1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	1	re1
2P33532337	1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	me1 / me1	1	re1
2P35391362	2 (2)	me2 / me2	me2 / me2	me2 / me2	me1 / me2	me2 / me2	me2 / me2	1	re2
2P36235952	3 (3)	me3 / me3	me3 / me3	me3 / me3	me3 / me3	me3 / me3	me3 / me3	2 (2)	re2

Table S.I.10. continued

Seq. name	Haploid clementine <i>C. clementina</i>	Clementines <i>C. clementina</i>	Sweet oranges <i>C. sinensis</i>	Sour oranges <i>C. aurantium</i>	Grapefruits <i>C. paradisi</i>	Lemons <i>C. limon</i>	Volckamer lemon Hybrid of lemon	Meyer lemon Hybrid of lemon	Mexican lime <i>C. aurantifolia</i>	Bergamot <i>C. aurantium</i>	Alemow <i>C. macrophylla</i>
2P737170	re1	re1 / re1	re1 / ma1	re2 / ma1	re1 / ma2	ma1 / me1	re4 / me1	re1 / me1	me1 / mi1	re2 / ma1	me1 / mi1
2P3068140	re1	re1 / re1	re1 / ma1	re1 / ma1	re1 / ma1	ma1 / me1	re1 / me1	re4 / me1	me1 / mi1	re1 / me1	me1 / ?
2P4517048	re3	re2 / re3	re2 / re2	re1 / ma1	re2 / ma1	re1 / me1	re1 / me1	re3 / re3	me1 / mi1	ma1 / ma1	me1 / mi1
2P8108334	re2	re2 / re2	re2 / ma4	re5 / ma3	ma4 / ma7	re5 / me1	re3 / me2	re9 / me2	me1 / mi1	ma3 / me1	? / ?
2P11442721	re8	re1 / re8	re1 / ma4	re3 / ma1	ma1 / ma4	ma1 / me2	re7 / me1	ma1 / me1	me1 / mi1	ma1 / me2	me1 / mi1
2P13928427	re1	re1 / re1	re1 / ma1	re1 / ma1	ma1 / ma1	ma1 / me1	re1 / me1	ma3 / me2	me1 / mi1	ma1 / me1	me1 / mi2
2P21022460	re1	re1 / re1	re1 / ma1	ma-re- mi / ma-re- mi	ma1 / ma2	ma- re-mi / me1	ma- re-mi / me1	me1 / ma-re- mi	me1 / mi1	me1 / ma-re- mi	me2 / mi1
2P25198627	re1	re1 / re1	re1 / ma1	re3 / re- ma	ma1 / re-ma	re-ma / me1	re1 / me3	re1 / me1	me3 / mi1	re-ma / me1	me1 / mi1
2P26819388	re1	re1 / ma-re	ma-re / ma1	ma-re / ma2	ma1 / ma1	ma-re / me2	?	ma2 / me1	me1 / mi1	ma-re / me1	ma-re / me3
2P29538734	re5	re5 / ma3	ma1 / ma3	re7 / ma5	ma1 / ma3	re7 / me4	re1 / me3	re6 / me2	me4 / mi1	ma5 / me4	me3 / mi1
2P30446231	re1	re1 / ma1	ma1 / ma5	re3 / ma7	ma5 / ma7	re3 / me3	re2 / me2	re5 / me1	me1 / mi1	ma7 / me3	mi2 / mi?
2P32507721	ma- re1	ma1 / ma- re1	ma1 / ma1	ma-re1 / ma- re2	ma1 / ?	ma- re2 / me1	re1 / me1	re2 / me1	me1 / mi1	ma-re1 / me1	me1 / mi?
2P33506778	re1	re1 / ma1	ma1 / ma2	re1 / ma2	ma2 / ma2	re1 / me1	re1 / me1	re1 / me1	me1 / mi1	ma2 / me1	me1 / mi1
2P33532337	re1	re1 / ma2	ma1 / ma2	re1 / ma2	ma1 / ma1	re1 / me1	re1 / me1	re2 / re2	me1 / mi1	ma2 / ma2	me1 / mi1
2P35391362	re2	re2 / ma3	ma2 / ma3	re2 / ma1	ma2 / ma2	re2 / me2	re1 / me2	re2 / me2	me1 / mi1	re2 / me2	me2 / mi1
2P36235952	re2	re2 / ma4	ma4 / ma4	re3 / ma3	ma4 / ma5	re3 / me3	re4 / me1	re2 / me3	me1 / mi1	re3 / me3	me2 / mi3?

\*number of haplotype observed for this group and (number of haplotypes specific of the group)

re: *C. reticulata*ma: *C. maxima*me: *C. medica*mi: *C. micrantha*

?: non determined

Table S.I.11. SNP G<sub>st</sub> values of each taxa before and after introgression information correction.

Gene fragment	Position on chromosome 2	Variation	Initial Data				After Correction				
			Citrons G <sub>st</sub>	Mandarins G <sub>st</sub>	Pummelos G <sub>st</sub>	<i>C. micrantha</i> G <sub>st</sub>	<i>C. medica</i> G <sub>st</sub>	<i>C. reticulata</i> G <sub>st</sub>	<i>C. maxima</i> G <sub>st</sub>	<i>C. micrantha</i> G <sub>st</sub>	
2P737170	737201	[A/G]	1	0.2	0.2	0.2	1	0.21429	0.1714286	0.1132075	
	737203	[A/T/C]	0.27659574	0.37142857	0.37142857	0.538461538	0.54419	0.18408	0.148	0.6780252	
	737205	[T/G]	0.2	0.2	0.2	1	0.02083	0.0303	0.025	1	
	737240	[A/T]	1	0.2	0.2	0.2	1	0.21429	0.1714286	0.1132075	
	737253	[C/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.01031	0.01493	0.0123457	0.3333333	
	737276	[A/G]	0.17647059	0.17647059	0.06703146	0.155327343	0.10112	0.15254	0.2109889	0.1531263	
	737307	[T/G]	0.5	0.5	0.5	0.5	0.92157	0.25926	0.2058824	0.6619718	
	737315	[T/C]	0.2	0.2	0.2	1	0.02083	0.0303	0.025	1	
	737366	[T/C]	1	0.2	0.2	0.2	1	0.21429	0.1714286	0.1132075	
	737402	[T/G]	1	0.2	0.2	0.2	1	0.21429	0.1714286	0.1132075	
	737414	[A/T]	0.2	0.2	0.2	1	0.02083	0.0303	0.025	1	
	737442	[A/T]	0.1965812	0.13595532	0.88130166	0.196581197	0.24051	0.38776	0.9047619	0.1919192	
	737455	[C/G]	0.19148936	0.93103448	0.19148936	0.191489362	0.38028	1	0.4909091	0.2967033	
	737456	[T/C]	0.5	0.5	0.5	0.5	0.92157	0.25926	0.2058824	0.6619718	
	737471	[A/G]	0.5	0.5	0.5	0.5	0.92157	0.25926	0.2058824	0.6619718	
	737479	[T/G]	N	N	N	N	N	N	N	N	
	737509	[A/C]	0.00598802	0.01818182	0.00598802	0.005988024	0.01031	0.01887	0.0123457	0.008547	
737545	[T/C]	0.02564103	0.02564103	0.08108108	0.025641026	0.03297	0.04615	0.0810811	0.027027		
737555	[A/T]	0.09090909	0.09090909	0.09090909	0.333333333	0.01075	0.01493	0.012987	0.3333333		
737572	[T/C]	0.07142857	0.07142857	0.25	0.071428571	0.09302	0.13333	0.25	0.0754717		
737600	[C/G]	1	0.2	0.2	0.2	1	0.21429	0.1818182	0.1176471		
2P3068140	3068226	[A/C]	0.47368421	0.3804017	0.47368421	0.527272727	0.36364	0.80605	0.48	0.4102564	
	3068273	[T/G]	0.18309859	0.86666667	0.18309859	0.183098592	0.32353	0.91667	0.4230769	0.2643678	
	3068352	[A/G]	0.00598802	0.01818182	0.00598802	0.005988024	N	N	N	N	
	3068353	[A/G]	1	0.2	0.2	0.2	1	0.21429	0.1935484	0.122449	
	3068386	[A/G]	1	0.2	0.2	0.2	1	0.21429	0.1935484	0.122449	
	3068418	[A/C]	0.15862069	0.6969697	0.15862069	0.15862069	0.32353	0.91667	0.4230769	0.2643678	
	3068452	[A/G]	1	0.2	0.2	0.2	1	0.21429	0.1935484	0.122449	
	3068453	[T/C]	N	N	N	N	N	N	N	N	
	3068454	[T/G]	0.18309859	0.86666667	0.18309859	0.183098592	0.32353	0.91667	0.4230769	0.2643678	
	3068477	[T/C]	0.2173913	0.10623354	0.95348837	0.217391304	0.30435	0.34333	0.9473684	0.2222222	
	3068493	[A/C]	0.02439024	0.07692308	0.02439024	0.024390244	0.04651	0.09524	0.0571429	0.0377358	
	3068528	[T/C]	1	0.2	0.2	0.2	1	0.21429	0.1935484	0.122449	
	3068529	[A/G]	0.2173913	0.10623354	0.95348837	0.217391304	0.30435	0.34333	0.9473684	0.2222222	
	3068559	[T/C]	0.02564103	0.02564103	0.08108108	0.025641026	0.03448	0.04615	0.0810811	0.0280374	
	2P4517048	4517069	[A/G]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1
		4517076	[A/G]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1
		4517101	[T/C]	0.2	1	0.2	0.2	0.38889	1	0.5	0.3043478
4517111		[T/C]	0.02564103	0.02564103	0.08108108	0.025641026	0.03093	0.04615	0.0810811	0.025641	
4517124		[C/G]	0.09803922	0.36585366	0.09803922	0.098039216	0.17647	0.36585	0.2173913	0.1428571	
4517152		[A/C]	0.5	0.5	0.5	0.5	0.92308	0.25926	0.2	0.6666667	
4517161		[A/G]	0.05	0.16666667	0.05	0.05	0.08696	0.16667	0.1052632	0.0714286	
4517205		[A/G]	0.2	1	0.2	0.2	0.38889	1	0.5	0.3043478	
4517279		[A/G]	0.01694915	0.01694915	0.05263158	0.016949153	0.02041	0.0303	0.0526316	0.0169492	
4517280		[A/G]	0.01694915	0.01694915	0.05263158	0.016949153	0.02041	0.0303	0.0526316	0.0169492	
4517476		[A/T]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034	
4517483		[T/C]	0.04347826	0.01408451	0.01408451	0.014084507	0.04348	0.01493	0.0120482	0.0084034	
2P8108334		8108354	[C/G]	0.14285714	0.6	0.14285714	0.142857143	0.26582	0.6	0.3333333	0.2121212
	8108366	[C/G]	0.71428571	0.16129032	0.16129032	0.161290323	0.71429	0.17241	0.1351351	0.0909091	
	8108367	[C/G]	0.14285714	0.6	0.14285714	0.142857143	0.26582	0.6	0.3333333	0.2121212	
	8108368	[A/G]	0.09090909	0.02857143	0.02857143	0.028571429	0.09091	0.0303	0.0243902	0.0169492	
	8108372	[C/G]	0.2	0.2	1	0.2	0.25	0.41667	1	0.2	
	8108373	[T/C]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	8108387	[A/G]	0.00840336	0.00840336	0.02564103	0.008403361	0.0101	0.01493	0.025641	0.0084034	
	8108388	[T/G]	N	N	N	N	N	N	N	N	
	8108394	[T/C]	0.09090909	0.09090909	0.33333333	0.090909091	0.11111	0.17241	0.3333333	0.0909091	
	8108411	[C/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	8108431	[T/C]	0.2	1	0.2	0.2	0.38889	1	0.5	0.3043478	
	8108433	[A/G]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034	
	8108438	[T/C]	0.05263158	0.05263158	0.17647059	0.052631579	0.06383	0.09677	0.1764706	0.0526316	
	8108442	[A/T]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	8108451	[T/C]	0.09090909	0.02857143	0.02857143	0.028571429	0.09091	0.0303	0.0243902	0.0169492	
	8108466	[T/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	8108534	[C/G]	0.04347826	0.04347826	0.14285714	0.043478261	0.05263	0.07937	0.1428571	0.0434783	
	8108537	[T/G]	0.5	0.5	0.5	0.5	0.42857	0.88889	0.5555556	0.3636364	
	8108556	[A/C]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	8108562	[T/G]	0.5	0.5	0.5	0.5	0.42857	0.88889	0.5555556	0.3636364	
	8108566	[A/G]	0.48193887	0.50045746	0.43398426	0.48193887	0.2837	0.12459	0.0598153	0.1812961	
	8108576	[A/C]	0.04347826	0.04347826	0.14285714	0.043478261	0.05263	0.07937	0.1428571	0.0434783	
	8108600	[A/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	8108613	[T/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	8108614	[A/G]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034	
	8108654	[A/T/C]	0.09138155	0.06247604	0.31807273	0.09138155	0.11223	0.13765	0.3039711	0.0918114	
	8108664	[T/G]	0.09090909	0.02857143	0.02857143	0.028571429	0.09091	0.0303	0.0243902	0.0169492	
	8108665	[A/G]	0.13513514	0.55555556	0.13513514	0.135135135	0.25	0.55556	0.3125	0.2	
	8108693	[A/G]	0.38842975	0.11920624	0.63106796	0.388429752	0.63934	0.00879	0.3770492	0.4814815	
	8108700	[A/T]	0.09090909	0.02857143	0.02857143	0.028571429	0.09091	0.0303	0.0243902	0.0169492	
	8108706	[A/T]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	8108737	[C/G]	0.12751678	0.51351351	0.12751678	0.127516779	0.23457	0.51351	0.2923077	0.1881188	
	8108754	[T/C]	1	0.2	0.2	0.2	1	0.21429	0.1666667	0.1111111	
8108772	[T/C]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034		
8108778	[A/G]	0.04347826	0.04347826	0.14285714	0.043478261	0.05263	0.07937	0.1428571	0.0434783		
8108779	[A/C]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333		
8108783	[A/T]	0.41176471	0.41176471	0.17482517	0.6	0.20482	0.33333	0.5144082	0.6		
8108789	[A/G]	0.04347826	0.04347826	0.14285714	0.043478261	0.05263	0.07937	0.1428571	0.0434783		
8108825	[C/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333		
8108834	[T/C]	0.09090909	0.02857143	0.02857143	0.028571429	0.09091	0.0303	0.0243902	0.0169492		

Table S.I.11. continued

Gene fragment	Position on chromosome 2	Variation	Initial Data				After Correction				
			Citrons Gst	Mandarins Gst	Pummelos Gst	<i>C. micrantha</i> Gst	<i>C. medica</i> Gst	<i>C. reticulata</i> Gst	<i>C. maxima</i> Gst	<i>C. micrantha</i> Gst	
2P30446231	30446250	[A/C]	0.08108108	0.08108108	0.29032258	0.081081081	0.0989	0.15254	0.2903226	0.0810811	
	30446274	[C/G]	0.01818182	0.05660377	0.01818182	0.018181818	0.03093	0.0566	0.037037	0.025641	
	30446295	[A/G]	0.02564103	0.02564103	0.08108108	0.025641026	0.03093	0.04615	0.0810811	0.025641	
	30446309	[T/C]	N	N	N	N	N	N	N	N	
	30446313	[T/C]	0.07142857	0.07843137	0.01078167	0.071428571	0.11111	0.04656	0.0005211	0.0909091	
	30446316	[A/G]	0.6	0.14285714	0.14285714	0.142857143	0.6	0.15254	0.12	0.0810811	
	30446336	[T/C]	0.14285714	0.04347826	0.04347826	0.043478261	0.14286	0.04615	0.037037	0.025641	
	30446338	[A/C]	0.02564103	0.02564103	0.08108108	0.025641026	0.03093	0.04615	0.0810811	0.025641	
	30446344	[A/T]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	30446360	[T/C]	0.14285714	0.6	0.14285714	0.142857143	0.26582	0.6	0.3333333	0.2121212	
	30446363	[C/G]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
	30446393	[A/G]	0.6	0.14285714	0.14285714	0.142857143	0.6	0.15254	0.12	0.0810811	
	30446411	[A/T]	0.15384615	0.00654664	0.25160462	0.153846154	0.21951	0.00661	0.1597705	0.1764706	
	30446416	[A/G]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	30446419	[A/G]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034	
	30446429	[T/C]	0.08108108	0.08108108	0.29032258	0.081081081	0.0989	0.15254	0.2903226	0.0810811	
	30446430	[A/G]	0.9047619	0.2371134	0.0485807	0.237113402	0.88679	0.28302	0.027023	0.1428571	
	30446432	[A/T]	0.15384615	0.00654664	0.25160462	0.153846154	0.21951	0.00661	0.1597705	0.1764706	
	30446448	[T/G]	0.04347826	0.04347826	0.14285714	0.043478261	0.05263	0.07937	0.0434783	0.0434783	
	30446475	[T/G]	0.15384615	0.00654664	0.25160462	0.153846154	0.21951	0.00661	0.1597705	0.1764706	
	30446501	[T/C]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	30446509	[T/C]	0.14285714	0.6	0.14285714	0.142857143	0.26582	0.6	0.3333333	0.2121212	
	30446512	[T/C]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	30446534	[A/G]	0.05	0.16666667	0.05	0.05	0.08696	0.16667	0.1052632	0.0714286	
	30446549	[A/G]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	30446560	[A/C]	0.02564103	0.02564103	0.08108108	0.025641026	0.03093	0.04615	0.0810811	0.025641	
	30446585	[T/C]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
	30446673	[C/G]	N	N	N	N	N	N	N	N	
	2P32507721	32507773	[T/C]	0.15662651	0.09090909	0.15555556	0.156626506	0.21982	0.01971	0.0632911	0.1767109
		32507779	[T/C]	0.04347826	0.14285714	0.04347826	0.043478261	0.03627	0.06667	0.0434783	0.0300429
32507784		[A/C]	0.21212121	0.09923664	0.21212121	0.858064516	0.04956	0.0352	0.0595655	0.8020273	
32507820		[A/T]	0.04347826	0.14285714	0.04347826	0.043478261	0.03627	0.06667	0.0434783	0.0300429	
32507829		[T/G]	0.05299145	0.17782027	0.05299145	0.052991453	0.06641	0.12511	0.0800701	0.0547343	
32507830		[A/C]	0.05299145	0.17782027	0.05299145	0.052991453	0.06641	0.12511	0.0800701	0.0547343	
32507853		[T/G]	N	N	N	N	N	N	N	N	
32507866		[A/G]	0.5	0.5	0.5	0.5	0.92308	0.25926	0.2	0.6666667	
32507875		[A/G]	0.10526316	0.4	0.10526316	0.105263158	0.19048	0.4	0.2352941	0.1538462	
32507897		[A/G]	N	N	N	N	N	N	N	N	
32507899		[A/G]	N	N	N	N	N	N	N	N	
32507956		[A/G]	N	N	N	N	N	N	N	N	
32508068		[T/C]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
32508099		[T/G]	N	N	N	N	N	N	N	N	
32508122		[T/G]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1	
32508132		[T/C]	0.09090909	0.09090909	0.09090909	0.333333333	0.0101	0.01493	0.0120482	0.3333333	
2P33506778		33506802	[C/G]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1
		33507103	[A/C/G]	0.12789527	0.51556157	0.12789527	0.127895267	0.23533	0.51556	0.2933025	0.1887073
		33507112	[A/G]	1	0.2	0.2	0.2	1	0.21429	0.1666667	0.1111111
		33507145	[A/T]	0.2	0.2	0.2	1	0.02041	0.0303	0.0243902	1
	33507184	[A/G]	0.13513514	0.55555556	0.13513514	0.135135135	0.25	0.55556	0.3125	0.2	
	33507277	[T/C]	0.15862069	0.05910165	0.5557879	0.15862069	0.20482	0.17931	0.5144082	0.1650485	
	33532469	[A/G]	0.15068493	0.64705882	0.15068493	0.150684932	0.33333	1	0.44	0.255814	
2P33532337	33532481	[A/G]	0.2	0.2	0.2	1	0.02326	0.0303	0.0285714	1	
	33532486	[C/G]	0.15068493	0.64705882	0.15068493	0.150684932	0.33333	1	0.44	0.255814	
	33532594	[A/G]	0.15068493	0.64705882	0.15068493	0.150684932	0.33333	1	0.44	0.255814	
	33532669	[A/G]	0.15068493	0.64705882	0.15068493	0.150684932	0.33333	1	0.44	0.255814	
	33532717	[T/C]	0.13975577	0.08385777	0.51632985	0.139755767	0.18919	0.25926	0.5384615	0.1489362	
	33532778	[A/C]	0.2	0.2	0.2	1	0.02326	0.0303	0.0285714	1	
	33532825	[A/G]	0.07006369	0.24444444	0.07006369	0.070063694	0.14286	0.33333	0.1803279	0.1134021	
	33532907	[T/G]	0.08667529	0.01287371	0.09804977	0.086675291	0.07317	0.09677	0.1764706	0.0888235	
	2P35391362	35391391	[T/G]	0.24076809	0.01343838	0.77224236	0.240768095	0.28358	0.40426	1	0.2183908
		35391435	[C/G]	0.96721311	0.21212121	0.12953832	0.212121212	1	0.22222	0.2	0.1276596
		35391442	[A/C]	0.15068493	0.64705882	0.15068493	0.150684932	0.34375	1	0.44	0.2619048
		35391449	[T/C]	0.15068493	0.64705882	0.15068493	0.150684932	0.34375	1	0.44	0.2619048
		35391452	[A/T]	0.04347826	0.01408451	0.01408451	0.014084507	0.04348	0.01538	0.0140845	0.0953208
35391453		[C/G]	0.24076809	0.01343838	0.77224236	0.240768095	0.28358	0.40426	1	0.2183908	
35391486		[T/C]	0.96721311	0.21212121	0.12953832	0.212121212	1	0.22222	0.2	0.1276596	
35391506		[T/G]	0.14285714	0.14285714	0.6	0.142857143	0.21127	0.29412	0.6521739	0.1648352	
35391545		[A/G]	0.15068493	0.64705882	0.15068493	0.150684932	0.34375	1	0.44	0.2619048	
35391574		[A/G]	0.24076809	0.01343838	0.77224236	0.240768095	0.28358	0.40426	1	0.2183908	
35391642		[T/C]	0.07006369	0.24444444	0.07006369	0.070063694	0.14667	0.33333	0.1803279	0.1157895	
35391656		[T/C]	0.46853147	0.45535341	0.42652567	0.468531469	0.86957	0.19962	0.2631579	0.6060606	
35391668		[A/G]	0.15068493	0.64705882	0.15068493	0.150684932	0.34375	1	0.44	0.2619048	
35391693		[A/T]	0.2	0.2	0.2	1	0.02381	0.03125	0.0285714	1	
35391705		[T/G]	N	N	N	N	N	N	N	N	
35391710		[A/T]	0.96721311	0.21212121	0.12953832	0.212121212	1	0.22222	0.2	0.1276596	
35391734		[A/G]	0.24076809	0.01343838	0.77224236	0.240768095	0.28358	0.40426	1	0.2183908	
35391759	[T/G]	0.14285714	0.6	0.14285714	0.142857143	0.32308	0.91304	0.4117647	0.2470588		
35391772	[A/G]	1	0.2	0.2	0.2	1	0.22222	0.2	0.1276596		
2P36235952	36235980	[A/G]	0.2	1	0.2	0.2	0.38889	1	0.5	0.3043478	
	36235991	[T/C]	N	N	N	N	N	N	N	N	
	36236023	[A/T]	0.2	1	0.2	0.2	0.38889	1	0.5	0.3043478	
	36236036	[C/G]	0.00840336	0.00840336	0.02564103	0.008403361	0.0101	0.01493	0.025641	0.0084034	
	36236039	[T/C]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034	
	36236040	[T/G]	0.03448276	0.03448276	0.11111111	0.034482759	0.04167	0.0625	0.1111111	0.0344828	
	36236071	[T/C]	N	N	N	N	N	N	N	N	
	36236072	[T/G]	N	N	N	N	N	N	N	N	
	36236096	[A/G]	0.2	0.2	1	0.2	0.25	0.41667	1	0.2	
	36236166	[A/T/G]	0.2047141	0.90930788	0.10677545	0.204714099	0.39451	0.85108	0.3441	0.308477	
	36236177	[A/T]	0.5	0.5	0.5	0.5	0.28205	0.47826	0.9090909	0.5	
	36236182	[A/G]	0.2	1	0.2	0.2	0.38889	1	0.5	0.3043478	
	36236201	[A/G]	0.00598802	0.01818182	0.00598802	0.005988024	0.0101	0.01818	0.0120482	0.0084034	
	36236228	[A/G]	0.19148936	0.93103448	0.19148936	0.191489362	0.36986	0.93103	0.4736842	0.2903226	

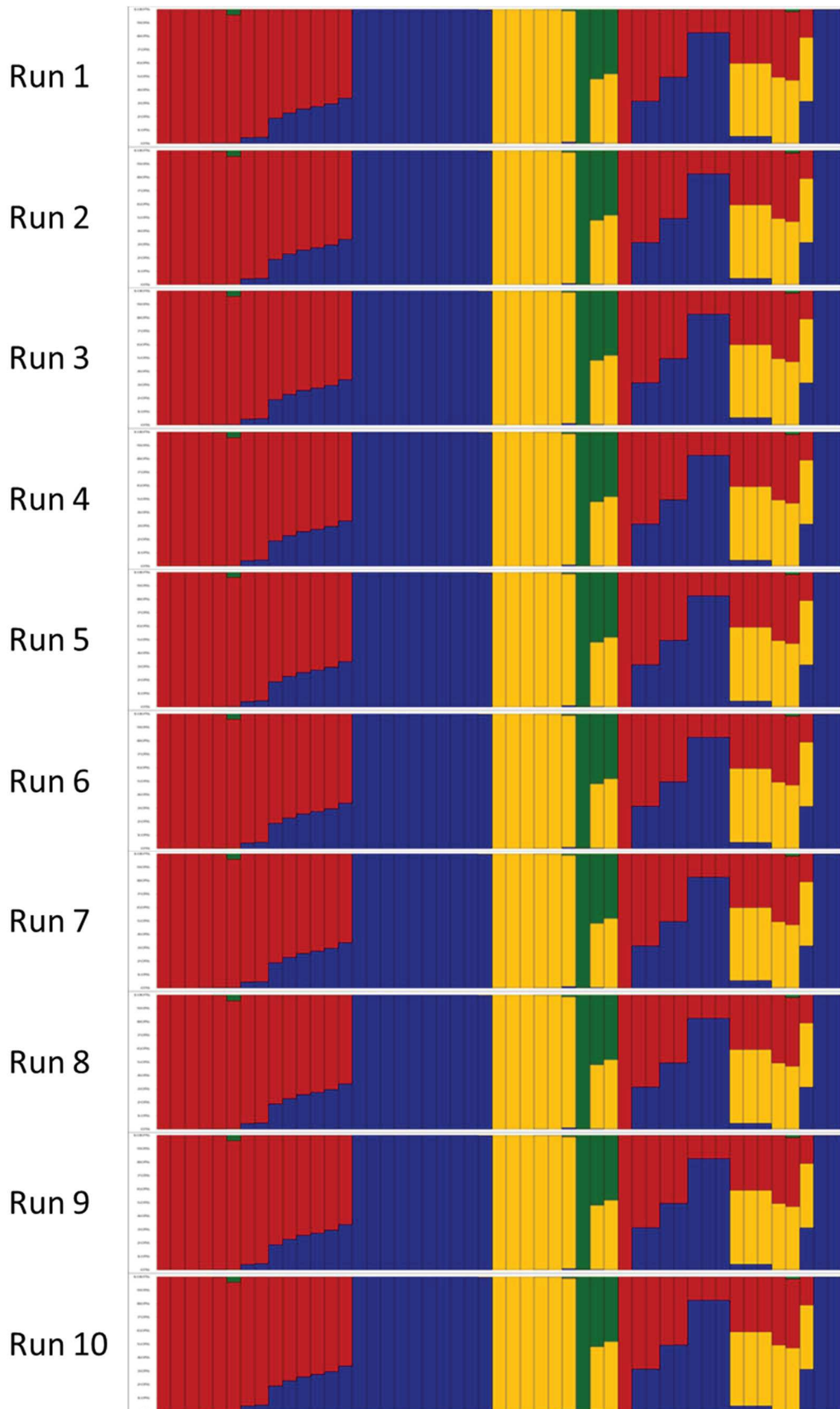


Figure S.I.1. 10 independent Structure run clusters output permuted and aligned in order to match up as closely as possible

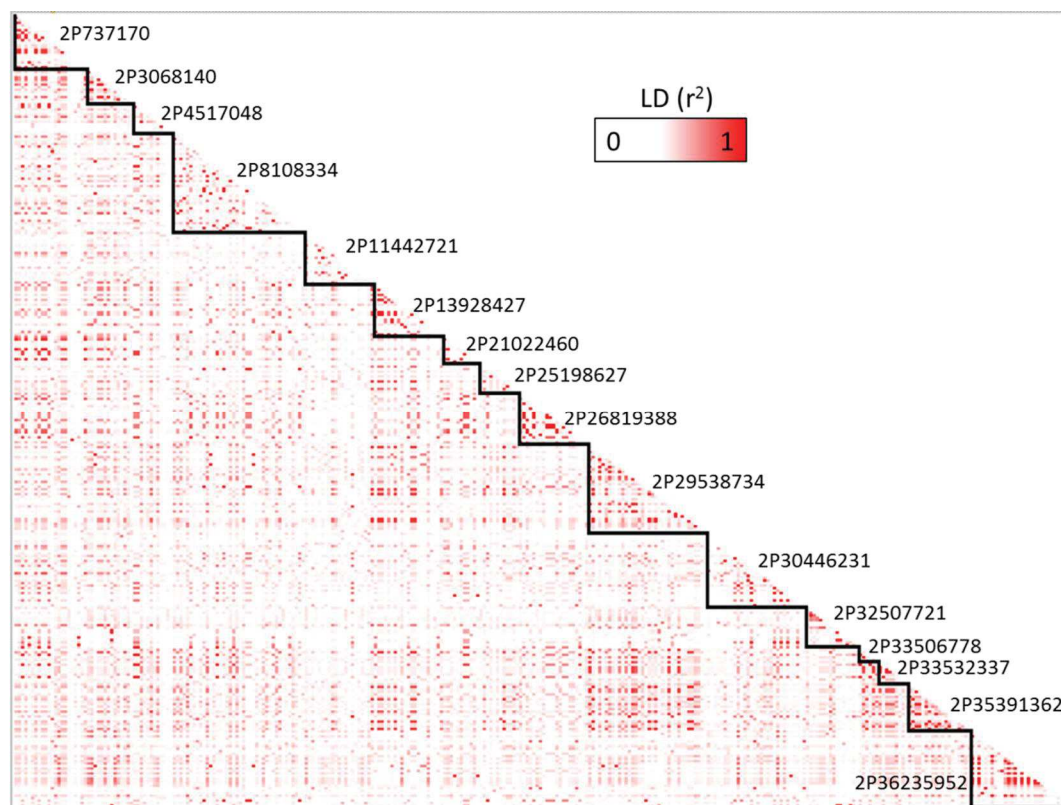


Figure S.I.2. Linkage disequilibrium (LD) between all SNPs along the chromosome 2

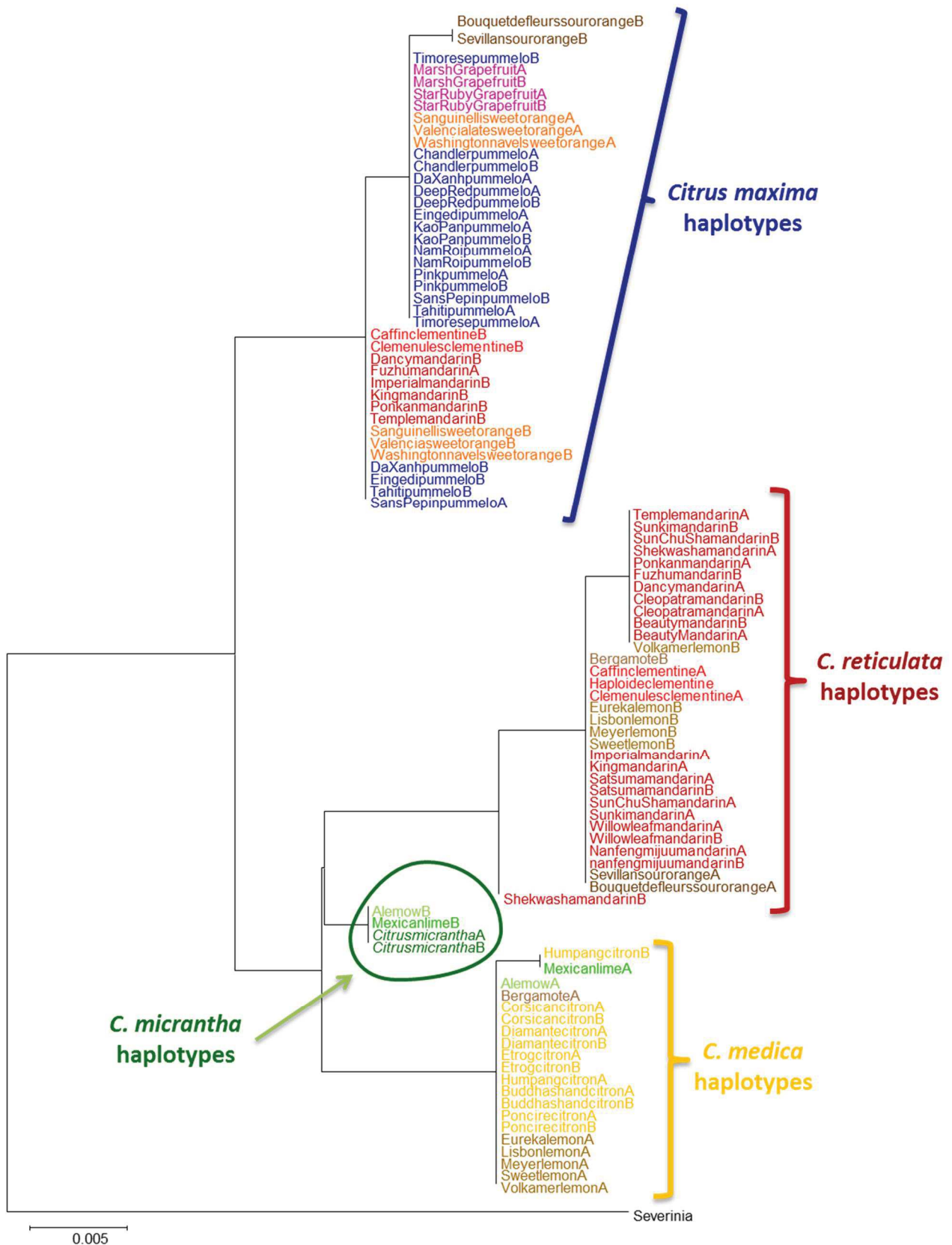


Figure S.I.3. Maximum likelihood phylogenetic tree of the haplotypic data of the 2P35391362 gene fragment.



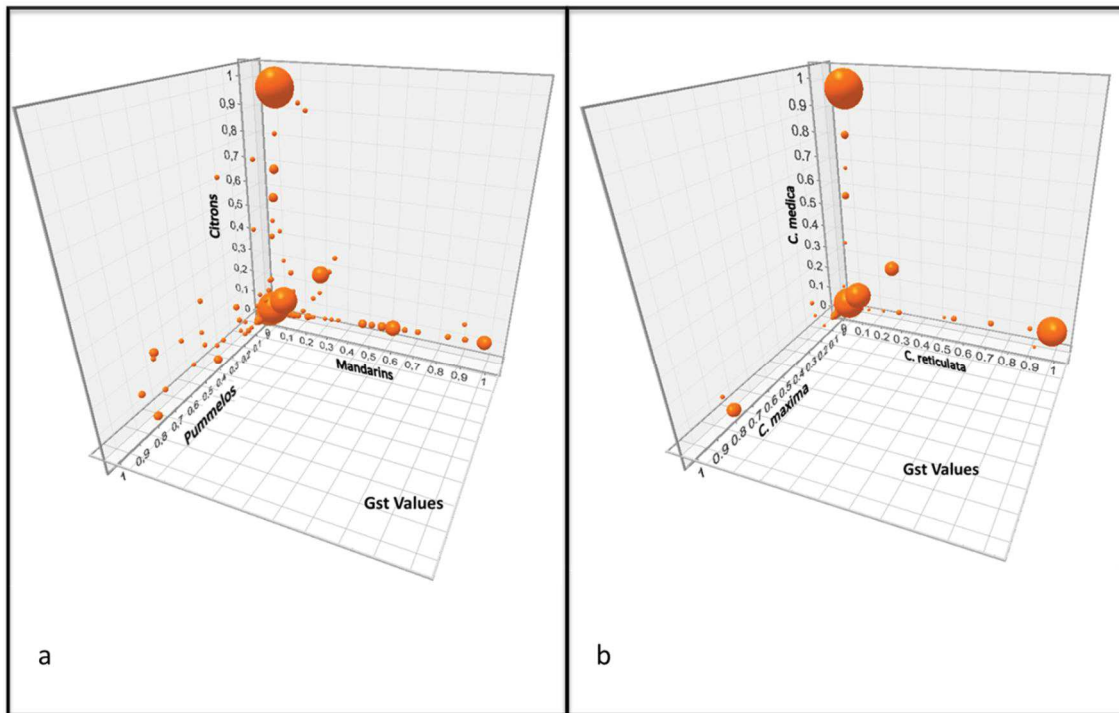


Figure S.I.4. 3D distribution of gene sequence SNPs according to their haplotypic  $G_{ST}$  value ; a:  $G_{ST}$  value for three horticultural groups (mandarins, pummelos and citrons); b:  $G_{ST}$  values for three basic taxa (*C. reticulata*, *C. maxima*, *C. medica*) after introgression information corrections..

# Annexe III

---

Supplementary information Chapitre II

Table S.II.1. Accession list

Horticultural group	Common name	Latin name (Tanaka, 1961)	Latin name (Swingle and Reece, 1967)	Germplasm ID	ID number	Accessions used for Kaspar SNP genotyping	Accessions used for 454 SNP mining
Mandarins	Beauty mandarin	<i>Citrus tangerina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	SRA	261	x	x
	Shekwasha mandarin	<i>Citrus depressa</i> Hayata	<i>Citrus reticulata</i> Blanco	IVIA	238	x	x
	Cleopatra mandarin	<i>Citrus reshni</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	IVIA	385	x	x
	Dancy mandarin	<i>Citrus tangerina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	IVIA	434	x	x
	King mandarin	<i>Citrus nobilis</i> Lour.	<i>Citrus reticulata</i> Blanco	IVIA	477	x	x
	Ponkan mandarin	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	IVIA	482	x	x
	Clausena Satsuma mandarin	<i>Citrus unshiu</i> Marcow	<i>Citrus reticulata</i> Blanco	IVIA	019	x	x
	Sunki mandarin	<i>Citrus sunki</i> (Hayata) hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	IVIA	239	x	x
	Willowleaf mandarin	<i>Citrus deliciosa</i> Ten.	<i>Citrus reticulata</i> Blanco	IVIA	154	x	x
	Ladu mandarin	<i>Citrus paratangerina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	SRA	595	x	
	Se Hui Gan mandarin	<i>Citrus suhuiensis</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	SRA	586	x	
	Szibat mandarin	<i>Citrus suhuiensis</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	SRA	596	x	
	Carvalhal mandarin	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	IVIA	568	x	
	Citrus daoxianensis	Citrus daoxianensis S.W. He & G.F. Liu	<i>Citrus reticulata</i> Blanco	IVIA	359	x	
	Fuzhu mandarin	<i>Citrus erythroa</i> hort. ex Tanaka	<i>Citrus tachibana</i> (Makino) Tanaka	IVIA	571	x	x
	Nanfengmiju mandarin	<i>Citrus kinokuni</i> hort. ex Tanaka	<i>Citrus tachibana</i> (Makino) Tanaka	SRA	839	x	x
	San Hu Hong Chu mandarin	<i>Citrus erythroa</i> hort. ex Tanaka	<i>Citrus tachibana</i> (Makino) Tanaka	SRA	769	x	
	Pummelos	Chandler pummelo	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	IVIA	207	x
Da Xanh pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	IVIA	589	x	x
Deep Red pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	IVIA	277	x	x
Kao Pan pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	SRA	321	x	x
Nam Roi pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	IVIA	590	x	x
Pink pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	IVIA	275	x	x
Tahiti pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	SRA	727	x	x
Eingedi pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	SRA	610		x
Sans pépins pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	SRA	710		x
Timorese pummelo		<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr	SRA	707	x	
Citrons	Corsican citron	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	IVIA	567	x	x
	Buddha's hand citron	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	IVIA	202	x	x
	Etrog citron	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	IVIA	169	x	x
	Humpang citron	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	SRA	722	x	x
	Poncire citron	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	SRA	701	x	x
	Mac Veu de montagne citron	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	SRA	760	x	
Papedas	Citrus micrantha	<i>Citrus micrantha</i> Wester	<i>Citrus micrantha</i> Wester	IVIA	626	x	x
	Citrus micrantha	<i>Citrus micrantha</i> Wester	<i>Citrus micrantha</i> Wester	SRA	1144	x	
Limes	Alemow	<i>Citrus macrophylla</i> Wester	<i>Citrus aurantiifolia</i> (Christm.) Swingle	IVIA	288	x	
	Excelsa lime	<i>Citrus excelsa</i> Wester	<i>Citrus aurantiifolia</i> (Christm.) Swingle	IVIA	167	x	
	Mexican lime	<i>Citrus aurantiifolia</i> (Christm.) Swingle	<i>Citrus aurantiifolia</i> (Christm.) Swingle	IVIA	164	x	
	Palestinian sweet lime	<i>Citrus limettioides</i> Tanaka	<i>Citrus aurantiifolia</i> (Christm.) Swingle	IVIA	305	x	
Sour oranges	Bouquet de fleurs sour orange	<i>Citrus aurantium</i> L.	<i>Citrus aurantium</i> L.	IVIA	139	x	
	Citrus myrtifolia	<i>Citrus aurantium</i> L.	<i>Citrus aurantium</i> L.	IVIA	136	x	
	Sevillan sour orange	<i>Citrus aurantium</i> L.	<i>Citrus aurantium</i> L.	IVIA	117	x	
Lemons	Eureka lemon	<i>Citrus limon</i> (L.) Burm. F.	<i>Citrus limon</i> (L.) Burm. F.	IVIA	297	x	
	Lisbon lemon	<i>Citrus limon</i> (L.) Burm. F.	<i>Citrus limon</i> (L.) Burm. F.	IVIA	214	x	
	Marrakech limonette	<i>Citrus limetta</i> Risso	<i>Citrus limon</i> (L.) Burm. F.	IVIA	484	x	
	Meyer lemon	<i>Citrus meyeri</i> Yu. Tanaka	<i>Citrus limon</i> (L.) Burm. F.	IVIA	145	x	
	Rangpur lime	<i>Citrus limonia</i> Osbeck	<i>Citrus limon</i> (L.) Burm. F.	IVIA	334	x	
	Volckamer lemon	<i>Citrus limonia</i> Osbeck	<i>Citrus limon</i> (L.) Burm. F.	IVIA	432	x	
Grapefruits	Duncan grapefruit	<i>Citrus paradisi</i> Macfad.	<i>Citrus paradisi</i> Macfad.	IVIA	274	x	
	Marsh grapefruit	<i>Citrus paradisi</i> Macfad.	<i>Citrus paradisi</i> Macfad.	IVIA	176	x	
	Star Ruby grapefruit	<i>Citrus paradisi</i> Macfad.	<i>Citrus paradisi</i> Macfad.	IVIA	197	x	
Sweet oranges	Tarocco sweet orange	<i>Citrus sinensis</i> (L.) Osbeck	<i>Citrus sinensis</i> (L.) Osbeck	IVIA	271	x	
	Valencia late sweet orange	<i>Citrus sinensis</i> (L.) Osbeck	<i>Citrus sinensis</i> (L.) Osbeck	IVIA	363	x	
	Washington navel sweet orange	<i>Citrus sinensis</i> (L.) Osbeck	<i>Citrus sinensis</i> (L.) Osbeck	IVIA	123	x	
Combava	Combava	<i>Citrus hystrix</i> DC.	<i>Citrus hystrix</i> DC.	IVIA	178	x	
Amblycarpa	Citrus amblycarpa	<i>Citrus amblycarpa</i> (Hassk.) Ochse	<i>Citrus reticulata</i> Blanco	IVIA	478	x	

Table S.II.1. continued

Horticultural group	Common name	Latin name (Tanaka, 1961)	Latin name (Swingle and Reece, 1967)	Germplasm ID	ID number	Accessions used for Kaspar SNP genotyping	Accessions used for 454 SNP mining
Small citrus	Clemenules clementine	<i>Citrus clementina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	IVIA	022	x	
	Fina clementine	<i>Citrus clementina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco	IVIA	039	x	
	Fallglo mandarin	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	IVIA	466	x	
	Fortune mandarin	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	IVIA	080	x	
	Fremont mandarin	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	IVIA	082	x	
	Wilking mandarin	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	IVIA	028	x	
	Minneola tangelo	<i>Citrus reticulata</i> Blanco x <i>Citrus paradisi</i> Macf.	<i>Citrus reticulata</i> Blanco x <i>Citrus paradisi</i> Macf.	IVIA	084	x	
	Nova tangelo	<i>Citrus reticulata</i> Blanco x <i>Citrus paradisi</i> Macf.	<i>Citrus reticulata</i> Blanco x <i>Citrus paradisi</i> Macf.	IVIA	074	x	
	Orlando tangelo	<i>Citrus reticulata</i> Blanco x <i>Citrus paradisi</i> Macf.	<i>Citrus reticulata</i> Blanco x <i>Citrus paradisi</i> Macf.	IVIA	101	x	
	Afourer tangor	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	SRA	741	x	
	Ellendale tangor	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	IVIA	194	x	
	Kiyomi tangor	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	SRA	704	x	
	Murcott tangor	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	IVIA	196	x	
	Ortanique tangor	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	IVIA	276	x	
	Temple tangor	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	SRA	280	x	
Triumph orangelo	<i>Citrus sinensis</i> (L.) Osbeck x <i>Citrus paradisi</i> Macf.	<i>Citrus sinensis</i> (L.) Osbeck x <i>Citrus paradisi</i> Macf.	SRA	474	x		
Other true citrus	Calamondin	<i>Citrus madurensis</i> Lour.	<i>Citrus reticulata</i> Blanco x <i>Citrus sinensis</i> (L.) Osb.	IVIA	135	x	
	Clymenia	<i>Clymenia polyandra</i> (Tanaka)	<i>Clymenia polyandra</i> (Tanaka)	IVIA	584	x	
	Desert lime	<i>Eremocitrus glauca</i> (Lindl.) Swingle	<i>Eremocitrus glauca</i> (Lindl.) Swingle	IVIA	346	x	
	Meiwa kumquat	<i>Fortunella crassifolia</i> Swingle	<i>Fortunella crassifolia</i> Swingle	IVIA	280	x	
	Marumi kumquat	<i>Fortunella japonica</i> (Thunb.) Swingle	<i>Fortunella japonica</i> (Thunb.) Swingle	IVIA	381	x	
	Australian round Lime	<i>Microcitrus australis</i> (Planchon) Swingle	<i>Microcitrus australis</i> (Planchon) Swingle	IVIA	313	x	
	Pomeroy poncirus	<i>Poncirus trifoliata</i> (L.) Raf.	<i>Poncirus trifoliata</i> (L.) Raf.	IVIA	374	x	
	Rubidoux poncirus	<i>Poncirus trifoliata</i> (L.) Raf.	<i>Poncirus trifoliata</i> (L.) Raf.	IVIA	217	x	
Near citrus	Ceylon atalantia	<i>Atalantia ceylanica</i> (Arn.) Oliv	<i>Atalantia ceylanica</i> (Arn.) Oliv	IVIA	172	x	
	Gillet's Cherry orange	<i>Citropsis gillettiana</i> Swingle & M. Kellerm	<i>Citropsis gillettiana</i> Swingle & M. Kellerm	IVIA	517	x	
Primitive citrus	Box orange	<i>Severinia buxifolia</i> (Poir.) Tenore	<i>Severinia buxifolia</i> (Poir.) Tenore	IVIA	147	x	
Triphasilineae	Limeberry	<i>Triphasia trifolia</i> (Burm. F.) P.Wils.	<i>Triphasia trifolia</i> (Burm. F.) P.Wils.	IVIA	182	x	
Clauseniae	Pink wampee	<i>Clausena excavata</i> Burm. f	<i>Clausena excavata</i> Burm. f	IVIA	311	x	
	Curry leaf	<i>Murraya koenigii</i> (L.) Spreng.	<i>Murraya koenigii</i> (L.) Spreng.	IVIA	377	x	

Table S.II.2; Multiplex genotype Identifiers (MID) and related genotypes

Mid	Seq	Genotypes		Latin name (Tanaka, 1961)	Latin name (Swingle and Reece, 1967)
MID-01	ACGAGTGCCT	Clausellina satsuma mandarin	IVIA 19	<i>Citrus unshiu</i> Marcow	<i>Citrus reticulata</i> Blanco
MID-07	CGTGTCTCTA	Buddha's hand citron	IVIA 202	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.
MID-14	CGAGAGATAC	Willowleaf mandarin	IVIA 154	<i>Citrus deliciosa</i> Ten.	<i>Citrus reticulata</i> Blanco
MID-15	ATACGACGTA	Etrog citron	IVIA 169	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.
MID-17	CGTCTAGTAC	Chandler pummelo	IVIA 207	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-19	TGTACTACTC	Shekwasha mandarin	IVIA 238	<i>Citrus depressa</i> Hayata	<i>Citrus reticulata</i> Blanco
MID-20	ACGACTACAG	Pink pummelo	IVIA 275	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-22	TACGAGTATG	Kao Pan pummelo	SRA 321	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-23	TACTCTCGTG	Cleopatra mandarin	IVIA 385	<i>Citrus reshni</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco
MID-24	TAGAGACGAG	Dancy mandarin	IVIA 434	<i>Citrus tangerina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco
MID-25	TCGTGCTCG	King mandarin	IVIA 477	<i>Citrus nobilis</i> Lour.	<i>Citrus reticulata</i> Blanco
MID-27	ACGCGAGTAT	Sunki mandarin	IVIA 239	<i>Citrus sunki</i> (Hayata) hort. ex Tanaka	<i>Citrus reticulata</i> Blanco
MID-28	ACTACTATGT	Deep Red pummelo	IVIA 277	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-33	ATAGAGTACT	Ponkan mandarin	IVIA 482	<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco
MID-35	CAGTAGACGT	Corsican citron	IVIA 567	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.
MID-37	TACACACACT	Nam Roi pummelo	IVIA 590	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-39	TACAGATCGT	Poncire citron	SRA 701	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.
MID-40	TACGCTGTCT	Sans pépins pummelo	SRA 710	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-42	TCGATCACGT	Beauty mandarin	SRA 261	<i>Citrus tangerina</i> hort. ex Tanaka	<i>Citrus reticulata</i> Blanco
MID-43	TCGCACTAGT	Fuzhu mandarin	IVIA 571	<i>Citrus erythroa</i> hort. ex Tanaka	<i>Citrus tachibana</i> (Makino) Tanaka
MID-44	TCTAGCGACT	Da Xanh pummelo	IVIA 589	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-45	TCTATACTAT	Eingedi pummelo	SRA 610	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr
MID-46	TGACGTATGT	Citrus micrantha	IVIA 626	<i>Citrus micrantha</i> Wester	<i>Citrus micrantha</i> Wester
MID-48	ACAGTATATA	Humpang citron	SRA 722	<i>Citrus medica</i> L.	<i>Citrus medica</i> L.
MID-49	ACGCGATCGA	Nanfengmiju mandarin	SRA 839	<i>Citrus kinokuni</i> hort. ex Tanaka	<i>Citrus tachibana</i> (Makino) Tanaka
MID-69	TCTGACGTCA	Tahitian pummelo	SRA 727	<i>Citrus maxima</i> (Burm.) Merr	<i>Citrus maxima</i> (Burm.) Merr





TableS.II.3. continued

C8P19	CTCACTG00CA TTCTCA TCA TCA CCATCA CCGAGAA GCTTGTCCGA GAATAA GTAA CAA ACOCTGA TTTTGTCTGGAA ATAA GTCA TCCAAAA TTGA GTT	550	Oiclev 10030162mg
C8P24	ATAACA TTTGGGGCTTACG AACTCCAAAGGCTCTCCAT	502	Oiclev 10030097mg
C9P9	AGGGGACA CAAGAGA ATGAT TCAGACCCAAA TTGSA TTCC	490	Oiclev 10005777mg
C9P14	TCTGCA GGA TCTGCTTCC ATCCCTGGCATGCA GTAA TC	515	Oiclev 10005987mg
C9P15	GTCCA GGGTAT TCCCTCAAT CTCTGAGAA CCGCTGCTTGG	517	Oiclev 10006583mg
C9P2846	TGATCTGGCA GCTGAA ATG AGGACA CA GCA TCA AA GG	540	Oiclev 10006208mg
C9P21	GCTCA ACA GCGTAGCCACA GAAA CA GGGAGAA GGTGAG	507	Oiclev 10006230mg
C9P31	CA CCATATCCGGCTTCAACT CTGGCTCTGAGTCCCTTG	464	Oiclev 10006644mg









Table S.II.5. continued

2	C2P16	2P21022495	21022495	0.00	0.00	0.48	0.00	0.23	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.11	0.11	0.43	0.11	0.43	C	0.11	0.11	0.43	0.11	0.43	C	
2	C2P16	2P21022508	21022508	0.00	0.00	0.00	0.00	0.34	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
2	C2P16	2P21022555	21022555	0.17	0.47	0.00	0.00	0.37	0.00	0.25	0.00	0.00	0.08	1.00	0.47	-	-	0.78	0.03	0.40	0.14	0.14	0.40	N	0.15	0.67	0.15	0.15	0.67	P	
2	C2P16	2P21022565	21022565	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
2	C2P16	2P21022579	21022579	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
2	C2P16	2P21022580	21022580	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
2	C2P16	2P21022588	21022588	0.00	0.12	0.00	0.00	0.04	0.00	0.13	0.00	0.00	0.04	-0.07	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	-	-	-	-	-	N	
2	C2P16	2P21022692	21022692	0.30	0.00	0.00	0.00	0.15	0.18	0.00	0.00	0.00	0.08	0.39	-	-	-	0.46	0.10	0.03	0.03	0.03	0.10	M	0.11	0.03	0.03	0.03	0.11	M	
2	C2P16	2P21022733	21022733	0.30	0.38	0.00	0.00	0.46	0.18	0.25	0.00	0.00	0.16	0.39	0.33	-	-	0.65	0.07	0.02	0.75	0.31	0.75	C	0.05	0.05	0.76	0.30	0.76	C	
2	C2P16	2P21022734	21022734	0.00	0.12	0.00	0.00	0.04	0.00	0.13	0.00	0.00	0.04	-0.07	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.02	0.05	0.02	0.02	0.05	P	
2	C2P16	2P21022735	21022735	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	-	-	-	-	-	N	
2	C2P16	2P21022737	21022737	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
2	C2P16	2P21022741	21022741	0.17	0.00	0.18	0.00	0.11	0.18	0.00	0.20	0.00	0.12	0.40	-	-	-	0.27	0.03	0.02	0.02	0.02	0.03	N	0.03	0.03	0.02	0.03	0.03	N	
2	C2P16	2P21022742	21022742	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	-	-	-0.02	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C	
2	C2P16	2P21022743	21022743	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	-	-	-0.02	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C	
2	C2P16	2P21022776	21022776	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
2	C2P16	2P21022846	21022846	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
2	C2P16	2P21022864	21022864	0.00	0.22	0.00	0.00	0.08	0.00	0.25	0.00	0.00	0.08	-0.14	-	-	-	-0.04	0.02	0.07	0.02	0.02	0.07	P	0.02	0.05	0.02	0.02	0.05	P	
2	C2P16	2P21022876	21022876	0.00	0.00	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
2	C2P16	2P21022903	21022903	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
2	C2P16	Average		0.07	0.06	0.05	0.00	0.13	0.05	0.05	0.03	0.00	0.05	0.19	0.10	0.17	-	0.35	0.07	0.08	0.20	0.16			0.08	0.10	0.22	0.17			
2	C2P16	SD		0.10	0.14	0.12	0.00	0.13	0.07	0.09	0.07	0.00	0.04	0.35	0.28	0.56	-	0.45	0.07	0.11	0.33	0.30			0.08	0.16	0.34	0.31			
2	C2P16	CI		0.04	0.06	0.05	-	0.06	0.03	0.04	0.03	-	0.02	0.22	0.24	0.54	-	0.20	0.03	0.05	0.14	0.13			0.04	0.07	0.16	0.14			
2	C2P18	2P25198709	25198709	0.10	0.00	0.00	0.00	0.04	0.10	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.03	0.01	0.01	0.01	0.03	M	0.04	0.01	0.01	0.01	0.04	M	
2	C2P18	2P25198777	25198777	0.18	0.00	0.42	0.00	0.30	0.20	0.00	0.60	0.00	0.20	-0.11	-	-	-	0.32	0.03	0.15	0.48	0.15	0.48	N	0.13	0.13	0.54	0.13	0.54	C	
2	C2P18	2P25198800	25198800	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic
2	C2P18	2P25198803	25198803	0.50	0.40	0.00	0.50	0.45	0.30	0.56	0.00	1.00	0.36	0.39	-0.39	-	-	-1.00	0.20	0.09	0.01	0.28	0.05	0.28	N	0.10	0.01	0.29	0.05	0.29	N
2	C2P18	2P25198884	25198884	0.10	0.00	0.00	0.00	0.04	0.10	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.03	0.01	0.01	0.01	0.03	M	-	-	-	-	-	N	
2	C2P18	2P25198897	25198897	0.50	0.40	0.00	0.00	0.44	0.30	0.56	0.00	0.00	0.32	0.39	-0.39	-	-	-	0.27	0.24	0.01	0.16	0.16	0.24	N	0.26	0.01	0.16	0.16	0.26	N
2	C2P18	2P25198918	25198918	0.10	0.00	0.00	0.00	0.34	0.10	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.88	0.13	0.21	0.97	0.21	0.97	C	0.20	0.20	1.00	0.20	1.00	C	
2	C2P18	2P25198921	25198921	0.10	0.49	0.00	0.00	0.30	0.10	0.67	0.00	0.00	0.28	-0.05	-0.35	-	-	-	0.05	0.03	0.26	0.09	0.09	0.26	N	0.08	0.29	0.08	0.08	0.29	P
2	C2P18	2P25198922	25198922	0.10	0.49	0.00	0.00	0.30	0.10	0.67	0.00	0.00	0.28	-0.05	-0.35	-	-	-	0.05	0.03	0.26	0.09	0.09	0.26	N	0.08	0.29	0.08	0.08	0.29	P
2	C2P18	2P25198955	25198955	0.10	0.00	0.32	0.00	0.30	0.10	0.00	0.00	0.00	0.04	-0.05	-	-	-	1.00	0.86	0.09	0.17	0.63	0.17	0.63	C	0.15	0.15	0.67	0.15	0.67	C
2	C2P18	2P25199040	25199040	0.00	0.00	0.32	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C	
2	C2P18	Average		0.16	0.16	0.10	0.09	0.24	0.13	0.22	0.05	0.18	0.15	0.04	-0.37	-	-	-1.00	0.32	0.07	0.11	0.27	0.12		0.12	0.12	0.30	0.12			
2	C2P18	SD		0.17	0.23	0.17	0.20	0.16	0.10	0.31	0.18	0.40	0.14	0.20	0.02	0.83	0.00	0.40	0.07	0.10	0.31	0.10			0.07	0.11	0.33	0.10			
2	C2P18	CI		0.10	0.13	0.10	0.12	0.09	0.06	0.18	0.11	0.24	0.08	0.13	0.02	0.93	-	0.24	0.04	0.06	0.18	0.06			0.04	0.07	0.20	0.06			
2	C2P25	2P33506802	33506802	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
2	C2P25	2P33507103	33507103	0.46	0.00	0.00	0.00	0.44	0.45	0.00	0.00	0.00	0.19	0.01	-	-	-	-	0.56	0.52	0.13	0.13	0.13	0.52	M	0.52	0.13	0.13	0.13	0.52	M
2	C2P25	2P33507112	33507112	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
2	C2P25	2P33507145	33507145	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
2	C2P25	2P33507184	33507184	0.40	0.00	0.00	0.00	0.43	0.36	0.00	0.00	0.00	0.15	0.08	-	-	-	-	0.64	0.57	0.14	0.14	0.14	0.57	M	0.57	0.14	0.14	0.14	0.57	M
2	C2P25	2P33507277	33507277	0.17	0.35	0.00	0.00	0.43	0.18	0.22	0.00	0.00	0.15	-0.10	0.36	-	-	-	0.64	0.05	0.58	0.17	0.17	0.58	P	0.05	0.58	0.17	0.17	0.58	P
2	C2P25	Average		0.17	0.06	0.00	0.00	0.29	0.17	0.04	0.00	0.00	0.08	0.00	0.36	-	-	-	0.81	0.29	0.24	0.31	0.44		0.29	0.24	0.31	0.44			
2	C2P25	SD		0.21	0.14	0.00	0.00	0.17	0.20	0.09	0.00	0.09	0.09	0.09	-	-	-	-	0.21	0.21	0.17	0.34	0.44		0.21	0.17	0.34	0.44			
2	C2P25	CI		0.17	0.11	-	-	0.14	0.16	0.07	-	-	0.07	0.10	0.14	0.27	0.35	-	0.17	0.17	0.14	0.27	0.35		0.17	0.14	0.27	0.			

Table S.II.5. continued

3	C3P3	3P2073715	2073715	0.50	0.00	0.00	0.00	0.33	0.56	0.00	0.00	0.00	0.21	-0.11	-	-	-	0.36	0.36	0.10	0.10	0.10	0.36	M	0.26	0.07	0.07	0.07	0.26	M
3	C3P3	3P2073751	2073751	0.26	0.00	0.00	0.00	0.45	0.33	0.00	0.00	0.00	0.13	-0.31	-	-	-	0.72	0.74	0.17	0.17	0.17	0.74	M	0.85	0.18	0.18	0.18	0.85	M
3	C3P3	3P2073775	2073775	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic
3	C3P3	3P2073789	2073789	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic
3	C3P3	3P2073796	2073796	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic
3	C3P3	3P2073812	2073812	0.30	0.00	0.00	0.00	0.46	0.36	0.00	0.00	0.00	0.16	-0.22	-	-	-	0.65	0.69	0.16	0.16	0.16	0.69	M	1.00	0.20	0.20	0.20	1.00	M
3	C3P3	3P2073865	2073865	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic
3	C3P3	3P2073898	2073898	0.40	0.00	0.00	0.00	0.20	0.36	0.00	0.00	0.00	0.15	0.08	-	-	-	0.25	0.16	0.05	0.05	0.05	0.16	M	0.33	0.09	0.09	0.09	0.33	M
3	C3P3	3P2073938	2073938	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	C	0.20	0.20	0.20	1.00	1.00	C
3	C3P3	3P2073997	2073997	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.04	0.01	0.01	0.01	0.04	M
3	C3P3	3P2074039	2074039	0.00	0.20	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-1.00	-	-	-	1.00	0.02	0.06	0.02	0.02	0.06	P	0.02	0.06	0.02	0.02	0.06	P
3	C3P3	3P2074061	2074061	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic
3	C3P3	3P2074098	2074098	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	C	0.20	0.20	0.20	1.00	1.00	C
3	C3P3	3P2074144	2074144	0.35	0.00	0.00	0.00	0.50	0.45	0.00	0.00	0.00	0.19	-0.29	-	-	-	0.62	0.20	0.42	0.59	0.42	0.59	C	0.50	0.50	0.50	0.50	0.50	N
3	C3P3	3P2074165	2074165	0.28	0.22	0.00	0.00	0.47	0.33	0.25	0.00	0.00	0.22	-0.20	-0.14	-	-	0.53	0.02	0.67	0.21	0.21	0.67	P	0.17	0.78	0.17	0.17	0.78	P
3	C3P3	Average		0.14	0.03	0.00	0.07	0.23	0.17	0.02	0.00	0.13	0.08	-0.16	0.43	-	-	-1.00	0.60	0.23	0.19	0.27	0.35		0.29	0.21	0.27	0.35		
3	C3P3	SD		0.18	0.07	0.00	0.18	0.18	0.21	0.06	0.00	0.35	0.09	0.14	0.81	-	-	0.00	0.41	0.22	0.17	0.33	0.36		0.28	0.19	0.32	0.36		
3	C3P3	CI		0.09	0.04	-	0.09	0.09	0.10	0.03	-	0.18	0.04	0.10	1.12	-	-	-	0.21	0.11	0.08	0.16	0.18		0.14	0.10	0.16	0.18		
3	C3P8	3P9086304	9086304	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M
3	C3P8	3P9086335	9086335	0.17	0.40	0.00	0.00	0.41	0.18	0.11	0.00	0.00	0.12	-0.10	0.72	-	-	0.72	0.04	0.51	0.16	0.16	0.51	P	0.17	0.75	0.17	0.17	0.75	P
3	C3P8	3P9086350	9086350	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	-	-	-	-	-	N
3	C3P8	3P9086366	9086366	0.00	0.20	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	1.00	-	-	1.00	0.02	0.06	0.02	0.02	0.06	P	-	-	-	-	-	N
3	C3P8	3P9086368	9086368	0.17	0.28	0.00	0.00	0.46	0.18	0.11	0.00	0.00	0.12	-0.10	0.60	-	-	0.75	0.30	0.23	0.47	0.53	0.53	Mic	0.50	0.50	0.50	0.50	0.50	N
3	C3P8	3P9086373	9086373	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	-	-	-	-	-	N
3	C3P8	3P9086381	9086381	0.17	0.28	0.00	0.00	0.46	0.18	0.11	0.00	0.00	0.12	-0.10	0.60	-	-	0.75	0.30	0.23	0.47	0.53	0.53	Mic	0.50	0.50	0.50	0.50	0.50	N
3	C3P8	3P9086388	9086388	0.24	0.28	0.00	0.00	0.50	0.09	0.11	0.00	0.00	0.08	0.61	0.60	-	-	0.85	0.17	0.38	0.33	0.64	0.64	Mic	0.33	0.64	0.33	0.64	0.64	P
3	C3P8	3P9086423	9086423	0.35	0.28	0.00	0.00	0.49	0.09	0.11	0.00	0.00	0.08	0.74	0.60	-	-	0.84	0.15	0.19	0.52	0.48	0.52	C	0.50	0.50	0.50	0.50	0.50	N
3	C3P8	3P9086425	9086425	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.04	0.01	0.01	0.04	P
3	C3P8	3P9086431	9086431	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	-	-	-	-	-	N
3	C3P8	3P9086435	9086435	0.35	0.28	0.00	0.00	0.49	0.09	0.11	0.00	0.00	0.08	0.74	0.60	-	-	0.84	0.15	0.19	0.52	0.48	0.52	C	0.50	0.50	0.50	0.50	0.50	N
3	C3P8	3P9086469	9086469	0.35	0.11	0.18	0.00	0.23	0.09	0.11	0.20	0.00	0.12	0.74	-0.06	-0.11	-	0.51	0.06	0.01	0.00	0.07	0.07	N	0.14	0.07	0.00	0.07	0.14	N
3	C3P8	3P9086470	9086470	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	-	-	-	-	-	N
3	C3P8	3P9086471	9086471	0.35	0.28	0.00	0.00	0.47	0.09	0.11	0.00	0.00	0.08	0.74	0.60	-	-	0.84	0.00	0.58	0.21	0.21	0.58	P	0.20	1.00	0.20	0.20	1.00	P
3	C3P8	3P9086473	9086473	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-0.11	-	-	-0.02	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C
3	C3P8	3P9086474	9086474	0.30	0.28	0.00	0.00	0.48	0.00	0.11	0.00	0.00	0.04	1.00	0.60	-	-	0.92	0.19	0.20	0.51	0.49	0.51	C	0.50	0.50	0.50	0.50	0.50	N
3	C3P8	3P9086503	9086503	0.17	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.07	0.02	0.02	0.02	0.07	M
3	C3P8	3P9086529	9086529	0.40	0.28	0.00	0.00	0.49	0.18	0.11	0.00	0.00	0.12	0.54	0.60	-	-	0.77	0.12	0.18	0.54	0.46	0.54	C	0.39	0.48	0.52	0.48	0.52	C
3	C3P8	3P9086540	9086540	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M
3	C3P8	3P9086569	9086569	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	C	0.20	0.20	0.20	1.00	1.00	C
3	C3P8	3P9086574	9086574	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M
3	C3P8	3P9086599	9086599	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.04	0.01	0.01	0.04	P
3	C3P8	3P9086615	9086615	0.46	0.00	0.00	0.00	0.26	0.18	0.00	0.00	0.00	0.08	0.61	-	-	-	0.71	0.22	0.06	0.06	0.06	0.22	M	0.33	0.09	0.09	0.09	0.33	M
3	C3P8	3P9086683	9086683	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.00	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M
3	C3P8	Average		0.16	0.14	0.01	0.00	0.23	0.07	0.07	0.02	0.00	0.06	0.36	0.39	-0.11	-	0.45	0.09	0.13	0.20	0.18			0.23	0.29	0.25	0.22		
3	C3P8	SD		0.15	0.13	0.05	0.00	0.20	0.07	0.06	0.06	0.00	0.04	0.43	0.37	0.00	-	0.44	0.10	0.16	0.27	0.22			0.20	0.31	0.28	0.23		
3	C3P8	CI		0.06	0.05	0.02	-	0.08	0.03	0.02	0.02	-	0.01	0.21	0.18	-	-	0.17	0.04	0.06	0.11	0.09			0.09	0.13	0.12	0.10		
3	C3P9	3P11001129	11001129	0.00	0.00	0.00	0.00	0.49	0.0																					

Table S.II.5. continued

4	C4P5	4P2154320	2154320	0.17	0.00	0.00	0.00	0.47	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.83	0.18	0.18	0.18	0.83	M	0.83	0.18	0.18	0.18	0.83	M		
4	C4P5	4P2154399	2154399	0.07	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic			
4	C4P5	4P2154420	2154420	0.17	0.00	0.00	0.00	0.07	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.04	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M		
4	C4P5	4P2154508	2154508	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.92	0.91	0.19	0.19	0.19	0.91	M	0.91	0.19	0.19	0.19	0.91	M		
4	C4P5	4P2154540	2154540	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	-	0.11	-	-	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C	
4	C4P5	4P2154541	2154541	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	-	0.11	-	-	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C	
4	C4P5	4P2154567	2154567	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M		
4	C4P5	4P2154589	2154589	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	-	0.11	-	-	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C	
4	C4P5	4P2154642	2154642	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M		
4	C4P5	4P2154652	2154652	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
4	C4P5	4P2154736	2154736	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M			
4	C4P5	Average		0.05	0.00	0.05	0.00	0.19	0.06	0.00	0.05	0.00	0.03	-0.07	-	-	-	0.42	0.30	0.10	0.18	0.17			0.30	0.10	0.18	0.17				
4	C4P5	SD		0.07	0.00	0.08	0.00	0.20	0.07	0.00	0.09	0.00	0.03	0.03	-	-	-	0.51	0.40	0.09	0.28	0.29			0.40	0.09	0.28	0.29				
4	C4P5	CI		0.04	-	0.05	-	0.12	0.04	-	0.06	-	0.02	0.02	-	-	-	0.30	0.24	0.06	0.17	0.17			0.24	0.06	0.17	0.17				
4	C4P14	4P10567127	10567127	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.05	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
4	C4P14	4P10567129	10567129	0.00	0.00	0.00	0.00	0.47	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M			
4	C4P14	4P10567210	10567210	0.00	0.13	0.00	0.00	0.04	0.00	0.14	0.00	0.00	0.05	-	-	-	0.08	-	-	0.02	0.01	0.04	0.01	0.04	P	0.01	0.04	0.01	0.01	0.04	P	
4	C4P14	4P10567255	10567255	0.11	0.00	0.00	0.00	0.04	0.11	0.00	0.00	0.00	0.05	-0.06	-	-	-	-	0.03	0.01	0.01	0.01	0.03	M	0.03	0.01	0.01	0.01	0.03	M		
4	C4P14	4P10567300	10567300	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M			
4	C4P14	4P10567350	10567350	0.00	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
4	C4P14	4P10567352	10567352	0.00	0.25	0.00	0.00	0.09	0.00	0.29	0.00	0.00	0.09	-	-	-	0.17	-	-	0.02	0.08	0.02	0.02	0.08	P	0.02	0.08	0.02	0.02	0.08	P	
4	C4P14	4P10567353	10567353	0.00	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
4	C4P14	4P10567366	10567366	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P			
4	C4P14	4P10567424	10567424	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.05	-	-	-	1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
4	C4P14	4P10567547	10567547	0.00	0.34	0.00	0.00	0.13	0.03	0.43	0.00	0.00	0.14	-	-	-	0.27	-	-	0.04	0.12	0.04	0.04	0.12	P	0.04	0.12	0.04	0.04	0.12	P	
4	C4P14	4P10567556	10567556	0.00	0.00	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
4	C4P14	4P10567566	10567566	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.05	-	-	-	1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
4	C4P14	4P10567597	10567597	0.00	0.00	0.00	0.00	0.43	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P			
4	C4P14	Average		0.01	0.05	0.00	0.11	0.24	0.01	0.06	0.00	0.21	0.03	-0.06	-	-	-	0.48	0.24	0.25	0.30	0.18			0.24	0.25	0.30	0.18				
4	C4P14	SD		0.03	0.11	0.00	0.21	0.19	0.03	0.13	0.00	0.43	0.04	-	-	-	0.10	0.54	0.33	0.32	0.39	0.12			0.33	0.32	0.39	0.12				
4	C4P14	CI		0.01	0.06	-	0.11	0.10	0.02	0.07	-	0.22	0.02	-	-	-	0.28	0.17	0.17	0.20	0.06				0.17	0.17	0.20	0.06				
4	C4P16	4P14276848	14276848	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.50	0.50	0.50	0.50	0.50	N	0.50	0.50	0.50	0.50	0.50	N			
4	C4P16	4P14276953	14276953	0.48	0.00	0.00	0.00	0.29	0.27	0.00	0.00	0.00	0.12	0.44	-	-	-	0.60	0.26	0.07	0.07	0.07	0.26	M	0.26	0.07	0.07	0.07	0.26	M		
4	C4P16	4P14276954	14276954	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic			
4	C4P16	4P14276984	14276984	0.00	0.48	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.12	-	-	-	0.30	-	-	0.65	0.11	0.44	0.11	0.44	P	0.11	0.44	0.11	0.11	0.44	P	
4	C4P16	4P14277017	14277017	0.00	0.40	0.00	0.00	0.17	0.00	0.33	0.00	0.00	0.12	-	-	-	0.17	-	-	0.34	0.05	0.16	0.05	0.16	P	0.05	0.16	0.05	0.16	0.16	P	
4	C4P16	4P14277059	14277059	0.17	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M		
4	C4P16	4P14277100	14277100	0.24	0.00	0.00	0.00	0.11	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-	0.07	0.02	0.02	0.02	0.07	M	0.07	0.02	0.02	0.02	0.07	M		
4	C4P16	4P14277120	14277120	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M			
4	C4P16	4P14277155	14277155	0.09	0.00	0.00	0.00	0.11	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.65	0.13	0.21	0.21	0.97	0.97	Mic	0.13	0.21	0.21	0.97	0.97	Mic		
4	C4P16	4P14277197	14277197	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M			
4	C4P16	Average		0.10	0.09	0.00	0.00	0.25	0.06	0.07	0.00	0.00	0.05	0.31	0.23	-	-	0.72	0.34	0.20	0.16	0.31			0.34	0.20	0.16	0.31				
4	C4P16	SD		0.16	0.19	0.00	0.00	0.16	0.11	0.14	0.00	0.00	0.06	0.53	0.09	-	-	0.36	0.37	0.16	0.14	0.38			0.37	0.16	0.14	0.38				
4	C4P16	CI		0.10	0.11	-	-	0.10	0.07	0.09	-	0.04	0.04	0.52	0.13	-	-	0.22	0.23	0.10	0.09	0.24			0.23	0.10	0.09	0.24				
4	C4P20	4P18900188	18900188	0.00	0.00	0.18	0.00	0.05	0.00	0.00	0.20	0.00	0.05	-	-	-	0.11	-	-	0.03	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C
4	C4P20	4P18900190	18900190	0.42	0.00	0.00	0.00	0.48	0.20	0.00	0.00	0.00	0.05	0.52	-	-	-	0.89	0.13	0.62	0.40	0.40	0.62	P	0.13	0.62	0.40	0.40	0.62	P		
4	C4P20	4P18900194	18900194	0.42	0.00	0.00	0.00	0.15	0.20	0.00	0.00	0.00	0.05	0.52	-	-	-	0.64	0.18	0.05	0.05											



Table S.II.5. continued

5	C5P13	5P22687265	22687265	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic				
5	C5P13	5P22687304	22687304	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
5	C5P13	5P22687409	22687409	0.24	0.00	0.00	0.00	0.46	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-	0.75	0.76	0.17	0.17	0.17	0.76	M	1.00	0.20	0.20	0.20	1.00	M			
5	C5P13	5P22687428	22687428	0.24	0.00	0.00	0.00	0.46	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-	0.75	0.76	0.17	0.17	0.17	0.76	M	1.00	0.20	0.20	0.20	1.00	M			
5	C5P13	5P22687463	22687463	0.09	0.11	0.00	0.00	0.07	0.09	0.11	0.00	0.00	0.08	-0.05	-0.06	-	-	-	-	0.01	0.01	0.02	0.02	0.02	N	0.01	0.03	0.01	0.01	0.03	P			
5	C5P13	5P22687587	22687587	0.35	0.00	0.00	0.00	0.17	0.27	0.00	0.00	0.00	0.12	0.22	-	-	-	-	0.34	0.13	0.04	0.04	0.04	0.13	M	0.19	0.05	0.05	0.05	0.19	M			
5	C5P13	5P22687598	22687598	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-	-1.00	-0.02	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic			
5	C5P13	5P22687640	22687640	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
5	C5P13	Average		0.11	0.01	0.00	0.13	0.23	0.11	0.01	0.00	0.25	0.06	-0.04	-0.06	-	-	-	-1.00	0.47	0.28	0.12	0.32	0.18		0.35	0.13	0.33	0.19					
5	C5P13	SD		0.14	0.04	0.00	0.23	0.18	0.14	0.04	0.00	0.46	0.05	0.18	-	-	-	-	0.00	0.46	0.30	0.07	0.42	0.12		0.41	0.07	0.42	0.11					
5	C5P13	CI		0.10	0.03	-	0.16	0.12	0.09	0.03	-	0.32	0.03	0.18	-	-	-	-	0.32	0.21	0.05	0.29	0.08			0.28	0.05	0.29	0.08					
5	C5P20	5P33038503	33038503	0.17	0.41	0.00	0.00	0.38	0.18	0.00	0.00	0.00	0.08	-0.10	1.00	-	-	-	-	0.78	0.04	0.50	0.16	0.16	0.50	P	0.09	0.94	0.22	0.22	0.94	P		
5	C5P20	5P33038516	33038516	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
5	C5P20	5P33038524	33038524	0.40	0.00	0.00	0.00	0.22	0.18	0.00	0.00	0.00	0.08	0.54	-	-	-	-	-	0.62	0.16	0.05	0.05	0.05	0.16	M	0.16	0.05	0.05	0.05	0.16	M		
5	C5P20	5P33038556	33038556	0.00	0.41	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	-	1.00	-	-	-	-	1.00	0.05	0.17	0.05	0.05	0.17	P	-	-	-	-	-	N		
5	C5P20	5P33038610	33038610	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
5	C5P20	5P33038613	33038613	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M		
5	C5P20	5P33038626	33038626	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic		
5	C5P20	5P33038638	33038638	0.17	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M		
5	C5P20	5P33038671	33038671	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
5	C5P20	5P33038740	33038740	0.50	0.41	0.00	0.00	0.44	0.55	0.00	0.00	0.00	0.25	-0.10	1.00	-	-	-	-	0.44	0.23	0.02	0.16	0.16	0.23	N	0.38	0.10	0.10	0.10	0.38	M		
5	C5P20	5P33038742	33038742	0.00	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
5	C5P20	5P33038791	33038791	0.00	0.00	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.50	0.50	0.50	0.50	1.00	N	0.50	0.50	0.50	0.50	1.00	N		
5	C5P20	5P33038827	33038827	0.43	0.00	0.00	0.00	0.25	0.64	0.00	0.00	0.00	0.29	-0.47	-	-	-	-	-	-0.17	0.19	0.06	0.06	0.06	0.19	M	0.19	0.06	0.06	0.06	0.19	M		
5	C5P20	5P33038885	33038885	0.00	0.41	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	-	1.00	-	-	-	-	1.00	0.14	0.56	0.14	0.14	0.56	P	0.20	1.00	0.20	0.20	1.00	P		
5	C5P20	5P33038886	33038886	0.00	0.41	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.00	-	1.00	-	-	-	-	1.00	0.14	0.56	0.14	0.14	0.56	P	0.20	1.00	0.20	0.20	1.00	P		
5	C5P20	5P33038926	33038926	0.00	0.13	0.00	0.00	0.12	0.00	0.14	0.00	0.00	0.04	-	-0.08	-	-	-	-	-	0.65	0.20	0.14	0.20	0.95	0.95	Mic	0.20	0.12	0.20	0.94	0.94	Mic	
5	C5P20	5P33038928	33038928	0.00	0.13	0.00	0.00	0.04	0.00	0.14	0.00	0.00	0.04	-	-0.08	-	-	-	-	-	0.02	0.01	0.04	0.01	0.01	0.04	P	0.02	0.05	0.02	0.02	0.05	P	
5	C5P20	5P33038929	33038929	0.00	0.13	0.00	0.00	0.04	0.00	0.14	0.00	0.00	0.04	-	-0.08	-	-	-	-	-	-0.02	0.01	0.04	0.01	0.01	0.04	P	0.02	0.05	0.02	0.02	0.05	P	
5	C5P20	5P33038933	33038933	0.00	0.13	0.00	0.00	0.04	0.00	0.14	0.00	0.00	0.04	-	-0.08	-	-	-	-	-	-0.02	0.01	0.04	0.01	0.01	0.04	P	0.02	0.05	0.02	0.02	0.05	P	
5	C5P20	Average		0.09	0.14	0.00	0.03	0.21	0.09	0.03	0.00	0.05	0.05	0.14	0.52	-	-	-	-	-1.00	0.59	0.14	0.19	0.25	0.22		0.16	0.27	0.27	0.24				
5	C5P20	SD		0.17	0.18	0.00	0.11	0.14	0.19	0.06	0.00	0.23	0.08	0.53	0.57	-	-	-	-	-	0.47	0.12	0.19	0.35	0.29		0.13	0.35	0.36	0.30				
5	C5P20	CI		0.07	0.08	-	0.05	0.06	0.08	0.03	-	0.10	0.04	0.43	0.37	-	-	-	-	-	0.21	0.05	0.09	0.16	0.13		0.06	0.16	0.16	0.14				
5	C5P23	5P38011692	38011692	0.00	0.28	0.00	0.00	0.11	0.00	0.33	0.00	0.00	0.12	-	-0.20	-	-	-	-	-	0.03	0.09	0.03	0.03	0.09	P	0.07	0.23	0.07	0.07	0.23	P		
5	C5P23	5P38011705	38011705	0.00	0.48	0.00	0.00	0.23	0.00	0.33	0.00	0.00	0.12	-	-0.30	-	-	-	-	-	0.51	0.07	0.24	0.07	0.07	0.24	P	-	-	-	-	-	N	
5	C5P23	5P38011733	38011733	0.00	0.48	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.12	-	-0.30	-	-	-	-	-	0.65	0.11	0.44	0.11	0.11	0.44	P	0.20	1.00	0.20	0.20	1.00	P	
5	C5P23	5P38011736	38011736	0.00	0.54	0.00	0.50	0.49	0.00	0.33	0.00	1.00	0.15	-	-0.38	-	-	-	-	-	0.69	0.35	0.01	0.57	0.18	0.57	C	0.24	0.24	0.76	0.17	0.76	C	
5	C5P23	5P38011738	38011738	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
5	C5P23	5P38011742	38011742	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-	-	-	0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M	
5	C5P23	5P38011743	38011743	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-	-	0.01	0.03	0.01	0.01	0.01	0.03	P	-	-	-	-	-	N	
5	C5P23	5P38011753	38011753	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-	-	-	0.02	0.01	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M	
5	C5P23	5P38011762	38011762	0.46	0.00	0.00	0.00	0.39	0.36	0.00	0.00	0.00	0.15	0.21	-	-	-	-	-	-	0.61	0.47	0.12	0.12	0.47	M	0.47	0.12	0.12	0.12	0.47	M		
5	C5P23	5P38011770	38011770	0.00	0.48	0.00	0.50	0.26	0.00	0.56	0.00	1.00	0.23	-	-0.17	-	-	-	-	-	-1.00	0.11	0.17	0.06	0.17	0.16	0.17	N	0.23	0.22	0.23	0.09	0.23	N
5	C5P23	5P38011771	38011771	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
5	C5P23	5P38011795	38011795	0.00	0.49	0.00	0.00	0.31	0.00	0.22	0.00																							



Table S.II.5. continued

6	C6P6	6P7496136	7496136	0.43	0.00	0.00	0.00	0.29	0.27	0.00	0.00	0.00	0.12	0.37	-	-	-	0.60	0.00	0.28	0.28	0.81	0.81	Mic	0.00	0.28	0.28	0.81	0.81	Mic			
6	C6P6	6P7496184	7496184	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	1.00	1.00	M				
6	C6P6	6P7496223	7496223	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic			
6	C6P6	6P7496245	7496245	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	1.00	1.00	1.00	1.00	C	0.20	0.20	1.00	0.20	1.00	C				
6	C6P6	6P7496280	7496280	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic			
6	C6P6	6P7496318	7496318	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M			
6	C6P6	6P7496346	7496346	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic				
6	C6P6	6P7496397	7496397	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic			
6	C6P6		Average	0.07	0.00	0.00	0.19	0.16	0.05	0.00	0.00	0.38	0.03	0.16	-	-	-	-1.00	0.44	0.21	0.15	0.25	0.40		0.21	0.15	0.25	0.40					
6	C6P6		SD	0.15	0.00	0.00	0.26	0.17	0.10	0.00	0.00	0.52	0.04	0.30	-	-	-	0.00	0.51	0.33	0.09	0.32	0.33		0.33	0.09	0.32	0.33					
6	C6P6		CI	0.11	-	-	0.18	0.12	0.07	-	-	0.36	0.03	0.41	-	-	-	0.35	0.23	0.06	0.22	0.23			0.23	0.06	0.22	0.23					
6	C6P11	6P13421415	13421415	0.24	0.00	0.00	0.00	0.11	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-0.06	0.07	0.02	0.02	0.02	0.07	M	0.08	0.03	0.03	0.03	0.08	M			
6	C6P11	6P13421487	13421487	0.00	0.20	0.00	0.00	0.07	0.00	0.22	0.00	0.00	0.08	-	-0.13	-	-	-0.04	0.02	0.06	0.02	0.02	0.06	P	0.02	0.06	0.02	0.02	0.06	P			
6	C6P11	6P13421509	13421509	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M			
6	C6P11	6P13421524	13421524	0.09	0.48	0.00	0.00	0.26	0.09	0.11	0.00	0.00	0.08	-0.05	0.77	-	-	-	0.71	0.02	0.22	0.08	0.08	0.22	N	0.07	0.24	0.07	0.07	0.24	P		
6	C6P11	6P13421525	13421525	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.92	0.91	0.19	0.19	0.19	0.91	M	1.00	0.20	0.20	0.20	1.00	M			
6	C6P11	6P13421556	13421556	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic			
6	C6P11	6P13421631	13421631	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.20	0.00	0.04	-	-	-	0.52	-	0.65	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
6	C6P11	6P13421651	13421651	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic			
6	C6P11	6P13421659	13421659	0.17	0.00	0.00	0.00	0.47	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	-	1.00	0.09	0.94	0.22	0.22	0.94	P	0.20	1.00	0.20	0.20	1.00	P		
6	C6P11	6P13421666	13421666	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M			
6	C6P11	6P13421697	13421697	0.43	0.00	0.00	0.00	0.23	0.45	0.00	0.00	0.00	0.19	-0.05	-	-	-	-	0.18	0.19	0.06	0.06	0.06	0.19	M	0.21	0.06	0.06	0.06	0.21	M		
6	C6P11		Average	0.11	0.06	0.04	0.00	0.18	0.10	0.03	0.02	0.00	0.06	0.09	0.32	0.52	-	-	0.48	0.16	0.18	0.11	0.24		0.19	0.19	0.11	0.24					
6	C6P11		SD	0.13	0.15	0.13	0.00	0.16	0.14	0.07	0.06	0.00	0.06	0.41	0.63	-	-	-	0.47	0.26	0.27	0.09	0.38		0.28	0.28	0.09	0.38					
6	C6P11		CI	0.08	0.09	0.07	-	0.10	0.08	0.04	0.04	-	0.03	0.30	0.87	-	-	-	0.28	0.15	0.16	0.05	0.23		0.17	0.17	0.05	0.23					
6	C6P14	6P15755635	15755635	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	-	1.00	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C		
6	C6P14	6P15755637	15755637	0.00	0.28	0.00	0.00	0.11	0.00	0.11	0.00	0.00	0.04	-	-	-	0.60	-	-	0.65	0.03	0.09	0.03	0.03	0.09	P	0.03	0.09	0.03	0.03	0.09	P	
6	C6P14	6P15755669	15755669	0.17	0.00	0.00	0.00	0.47	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	-	0.84	0.83	0.18	0.18	0.18	0.83	M	0.83	0.18	0.18	0.18	0.83	M		
6	C6P14	6P15755672	15755672	0.10	0.00	0.00	0.00	0.04	0.10	0.00	0.00	0.00	0.04	-0.05	-	-	-	-	-0.02	0.03	0.01	0.01	0.01	0.03	M	0.03	0.01	0.01	0.01	0.03	M		
6	C6P14	6P15755686	15755686	0.00	0.00	0.32	0.00	0.26	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	1.00	0.15	0.15	0.67	0.15	0.67	C	0.15	0.15	0.67	0.15	0.67	C		
6	C6P14	6P15755714	15755714	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P			
6	C6P14	6P15755741	15755741	0.00	0.20	0.00	0.50	0.11	0.00	0.22	0.00	1.00	0.12	-	-0.13	-	-	-	-0.05	0.09	0.04	0.09	0.29	0.29	N	0.09	0.04	0.09	0.29	0.29	N		
6	C6P14	6P15755745	15755745	0.24	0.00	0.00	0.00	0.11	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-	-0.06	0.07	0.02	0.02	0.02	0.07	M	0.07	0.02	0.02	0.02	0.07	M		
6	C6P14	6P15755746	15755746	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P			
6	C6P14	6P15755762	15755762	0.17	0.00	0.00	0.00	0.07	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	-	-0.04	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M		
6	C6P14	6P15755763	15755763	0.00	0.35	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	-	-	1.00	0.04	0.13	0.04	0.13	P	0.04	0.13	0.04	0.04	0.13	P	
6	C6P14	6P15755768	15755768	0.46	0.00	0.00	0.00	0.26	0.36	0.00	0.00	0.00	0.15	0.21	-	-	-	-	0.41	0.22	0.06	0.06	0.06	0.22	M	0.22	0.06	0.06	0.06	0.22	M		
6	C6P14	6P15755784	15755784	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P	
6	C6P14	6P15755824	15755824	0.40	0.00	0.00	0.00	0.20	0.18	0.00	0.00	0.00	0.08	0.54	-	-	-	-	0.62	0.16	0.05	0.05	0.05	0.16	M	0.16	0.05	0.05	0.05	0.16	M		
6	C6P14	6P15755827	15755827	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	-	1.00	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C	
6	C6P14	6P15755839	15755839	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.50	0.50	0.50	0.50	0.50	N	0.50	0.50	0.50	0.50	0.50	N			
6	C6P14	6P15755857	15755857	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P			
6	C6P14	6P15755860	15755860	0.46	0.00	0.00	0.00	0.26	0.36	0.00	0.00	0.00	0.15	0.21	-	-	-	-	0.41	0.22	0.06	0.06	0.06	0.22	M	0.22	0.06	0.06	0.06	0.22	M		
6	C6P14	6P15755895	15755895	0.00	0.35	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	-	-	1.00	0.15	0.64	0.15	0.15	0.64	P	0.15	0.64	0.15	0.15	0.64	P
6	C6P14	6P15755914	15755914	0.17	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M		
6	C6P14	6P15755927	15755927	0.00	0.35	0.00	0.00	0.39	0.00	0.00	0.00	0.00	0.00	-	-	-	-																

Table S.II.5. continued

7	C7P2173	7P1214715	1214715	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.92	0.91	0.19	0.19	0.91	M	1.00	0.20	0.20	0.20	1.00	M			
7	C7P2173	7P1214723	1214723	0.48	0.00	0.00	0.00	0.29	0.45	0.00	0.00	0.00	0.19	0.06	-	-	-	0.33	0.26	0.07	0.07	0.07	0.26	M	0.29	0.08	0.08	0.08	0.29	M		
7	C7P2173	7P1214754	1214754	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	0.65	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P2173	7P1214763	1214763	0.24	0.00	0.00	0.00	0.11	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-0.06	0.07	0.02	0.02	0.02	0.07	M	0.08	0.03	0.03	0.03	0.08	M		
7	C7P2173	7P1214772	1214772	0.00	0.00	0.48	0.00	0.50	0.00	0.00	0.40	0.00	0.08	-	-	0.17	-	0.85	0.67	0.36	0.07	0.36	0.67	M	0.67	0.36	0.07	0.36	0.67	M		
7	C7P2173	7P1214781	1214781	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic		
7	C7P2173	7P1214793	1214793	0.40	0.00	0.00	0.00	0.20	0.36	0.00	0.00	0.00	0.15	0.08	-	-	-	0.25	0.16	0.05	0.05	0.05	0.16	M	0.18	0.05	0.05	0.05	0.18	M		
7	C7P2173	7P1214802	1214802	0.00	0.00	0.48	0.00	0.20	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.11	0.11	0.43	0.11	0.43	C	0.11	0.11	0.43	0.11	0.43	C		
7	C7P2173	7P1214871	1214871	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	0.65	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P2173	7P1214883	1214883	0.50	0.00	0.48	0.00	0.39	0.55	0.00	0.40	0.00	0.31	-0.10	-	0.17	-	0.22	0.12	0.17	0.08	0.17	0.17	N	0.16	0.18	0.07	0.18	0.18	N		
7	C7P2173	7P1214913	1214913	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
7	C7P2173	7P1214931	1214931	0.00	0.28	0.00	0.00	0.11	0.00	0.11	0.00	0.00	0.04	-	0.60	-	-	0.65	0.03	0.09	0.03	0.03	0.09	P	0.03	0.09	0.03	0.03	0.09	P		
7	C7P2173	7P1214934	1214934	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C		
7	C7P2173	7P1214996	1214996	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C		
7	C7P2173	7P1215018	1215018	0.00	0.00	0.48	0.00	0.14	0.00	0.00	0.40	0.00	0.08	-	-	0.17	-	0.46	0.07	0.07	0.25	0.07	0.25	C	0.07	0.07	0.25	0.07	0.25	C		
7	C7P2173	7P1215025	1215025	0.43	0.00	0.00	0.00	0.23	0.45	0.00	0.00	0.00	0.19	-0.05	-	-	-	0.18	0.19	0.06	0.06	0.06	0.19	M	0.21	0.06	0.06	0.06	0.21	M		
7	C7P2173	7P1215063	1215063	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic		
7	C7P2173	Average		0.13	0.02	0.20	0.00	0.20	0.13	0.01	0.09	0.00	0.07	-0.04	0.60	0.57	-	0.65	0.20	0.12	0.19	0.21				0.21	0.12	0.19	0.21			
7	C7P2173	SD		0.20	0.07	0.22	0.00	0.14	0.20	0.03	0.16	0.00	0.09	0.09	0.09	0.39	-	0.36	0.24	0.09	0.23	0.31				0.25	0.09	0.23	0.31			
7	C7P2173	CI		0.09	0.03	0.11	-	0.07	0.10	0.01	0.08	-	0.04	0.07	0.27	-	0.17	0.11	0.04	0.11	0.15						0.12	0.04	0.11	0.15		
7	C7P8	7P7137409	7137409	0.00	0.28	0.00	0.00	0.11	0.00	0.33	0.00	0.00	0.12	-	-	0.20	-	-0.06	0.03	0.09	0.03	0.03	0.09	P	0.03	0.09	0.03	0.03	0.09	P		
7	C7P8	7P7137412	7137412	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137413	7137413	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137438	7137438	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic		
7	C7P8	7P7137444	7137444	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137447	7137447	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-1.00	-0.02	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
7	C7P8	7P7137453	7137453	0.00	0.40	0.00	0.00	0.17	0.00	0.33	0.00	0.00	0.12	-	0.17	-	-	0.34	0.05	0.16	0.05	0.05	0.16	P	0.05	0.16	0.05	0.05	0.16	P		
7	C7P8	7P7137458	7137458	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	0.11	-	-0.02	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C		
7	C7P8	7P7137465	7137465	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137474	7137474	0.00	0.20	0.00	0.00	0.07	0.00	0.22	0.00	0.00	0.08	-	-	0.13	-	-0.04	0.02	0.06	0.02	0.02	0.06	P	0.02	0.06	0.02	0.02	0.06	P		
7	C7P8	7P7137480	7137480	0.00	0.28	0.00	0.00	0.11	0.00	0.33	0.00	0.00	0.12	-	-	0.20	-	-0.06	0.03	0.09	0.03	0.03	0.09	P	0.03	0.09	0.03	0.03	0.09	P		
7	C7P8	7P7137481	7137481	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137489	7137489	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137500	7137500	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.40	0.00	0.08	-	-	0.25	-	-0.04	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C		
7	C7P8	7P7137513	7137513	0.50	0.00	0.00	0.00	0.36	0.18	0.00	0.00	0.00	0.08	0.63	-	-	-	0.78	0.38	0.10	0.10	0.10	0.38	M	0.44	0.11	0.11	0.11	0.44	M		
7	C7P8	7P7137516	7137516	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.40	0.00	0.08	-	-	0.25	-	-0.04	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C		
7	C7P8	7P7137532	7137532	0.00	0.00	0.50	0.00	0.17	0.00	0.00	0.60	0.00	0.12	-	-	0.20	-	0.34	0.09	0.09	0.33	0.09	0.33	C	0.09	0.09	0.33	0.09	0.33	C		
7	C7P8	7P7137552	7137552	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137554	7137554	0.43	0.00	0.00	0.00	0.23	0.09	0.00	0.00	0.00	0.04	0.79	-	-	-	0.84	0.19	0.06	0.06	0.06	0.19	M	0.24	0.07	0.07	0.07	0.24	M		
7	C7P8	7P7137559	7137559	0.00	0.00	0.18	0.00	0.04	0.00	0.00	0.20	0.00	0.04	-	-	0.11	-	-0.02	0.02	0.02	0.05	0.02	0.05	C	0.02	0.02	0.05	0.02	0.05	C		
7	C7P8	7P7137564	7137564	0.00	0.00	0.42	0.00	0.11	0.00	0.00	0.60	0.00	0.12	-	-	0.43	-	-0.06	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C		
7	C7P8	7P7137591	7137591	0.00	0.44	0.00	0.00	0.20	0.00	0.44	0.00	0.00	0.15	-	0.00	-	-	0.25	0.06	0.20	0.06	0.06	0.20	P	0.06	0.20	0.06	0.06	0.20	P		
7	C7P8	7P7137595	7137595	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-	0.06	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P		
7	C7P8	7P7137621	7137621	0.00	0.00	0.48	0.00	0.20	0.00	0.00	0.40	0.00	0.08	-	-	0.17	-	0.62	0.11	0.11	0.43	0.11	0.43	C	0.11	0.11	0.43	0.11	0.43	C		
7	C7P8	7P7137647	7137647	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-	0.06	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P		
7	C7P8	7P7137650	7137650	0.00	0.00	0.48	0.00	0.14	0.00	0.00	0.80	0.00	0.15	-	-	0.67	-	-0.08	0.07													

Table S.II.5. continued

7	C7P19	7P19024974	19024974	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
7	C7P19	7P19024991	19024991	0.00	0.00	0.32	0.00	0.26	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.15	0.15	0.67	0.15	0.67	C	0.15	0.15	0.67	0.15	0.67	C	
7	C7P19	7P19024999	19024999	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	-	-	-	-	-	N	
7	C7P19	7P19025027	19025027	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
7	C7P19	7P19025075	19025075	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
7	C7P19	7P19025080	19025080	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.92	0.91	0.19	0.19	0.19	0.91	M	1.00	0.20	0.20	0.20	1.00	M	
7	C7P19	7P19025099	19025099	0.17	0.00	0.00	0.00	0.47	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	-	0.84	0.83	0.18	0.18	0.83	M	0.90	0.19	0.19	0.19	0.90	M	
7	C7P19	7P19025113	19025113	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P
7	C7P19	7P19025117	19025117	0.17	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M
7	C7P19	7P19025170	19025170	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
7	C7P19	7P19025180	19025180	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.92	0.91	0.19	0.19	0.19	0.91	M	1.00	0.20	0.20	0.20	1.00	M	
7	C7P19	7P19025186	19025186	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P
7	C7P19	7P19025196	19025196	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
7	C7P19	7P19025197	19025197	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M	
7	C7P19	7P19025227	19025227	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
7	C7P19	7P19025335	19025335	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
7	C7P19	7P19025354	19025354	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
7	C7P19	7P19025375	19025375	0.19	0.00	0.00	0.00	0.08	0.30	0.00	0.00	0.00	0.13	-0.59	-	-	-	-0.53	0.06	0.02	0.02	0.02	0.06	M	0.06	0.02	0.02	0.02	0.06	M	
7	C7P19	7P19025380	19025380	0.10	0.00	0.00	0.00	0.04	0.10	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.03	0.01	0.01	0.01	0.03	M	0.03	0.01	0.01	0.01	0.03	M	
7	C7P19	7P19025383	19025383	0.46	0.00	0.00	0.00	0.39	0.36	0.00	0.00	0.00	0.15	0.21	-	-	-	-	0.61	0.47	0.12	0.12	0.12	0.47	M	0.43	0.11	0.11	0.11	0.43	M
7	C7P19	7P19025384	19025384	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	-0.01	0.01	0.02	M	0.03	0.01	-0.01	0.01	0.03	M	
7	C7P19	7P19025385	19025385	0.00	0.00	0.42	0.00	0.23	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	-	0.84	0.13	0.13	0.54	0.13	0.54	C	0.13	0.13	0.54	0.13	0.54	C
7	C7P19	7P19025390	19025390	0.09	0.20	0.00	0.00	0.46	0.09	0.22	0.00	0.00	0.12	-0.05	-0.13	-	-	-	0.75	0.38	0.31	0.48	0.53	0.53	Mic	0.46	0.32	0.46	0.54	0.54	Mic
7	C7P19	7P19025408	19025408	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P
7	C7P19	Average		0.07	0.02	0.03	0.02	0.20	0.07	0.02	0.01	0.04	0.04	0.01	-0.08	0.76	-1.00	0.51	0.22	0.11	0.32	0.16				0.24	0.12	0.34	0.17		
7	C7P19	SD		0.10	0.05	0.11	0.10	0.17	0.10	0.06	0.04	0.20	0.04	0.36	0.03	0.34	-	0.52	0.28	0.09	0.40	0.22				0.31	0.09	0.40	0.22		
7	C7P19	CI		0.04	0.02	0.04	0.04	0.07	0.04	0.02	0.02	0.08	0.02	0.20	0.03	0.47	-	0.21	0.11	0.04	0.16	0.09				0.13	0.04	0.16	0.09		
7	C7P23	7P20541603	20541603	0.00	0.00	0.00	0.00	0.48	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M	
7	C7P23	7P20541609	20541609	0.40	0.00	0.00	0.00	0.20	0.18	0.00	0.00	0.00	0.08	0.54	-	-	-	-	0.62	0.16	0.05	0.05	0.05	0.16	M	0.16	0.05	0.05	0.05	0.16	M
7	C7P23	7P20541625	20541625	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M	
7	C7P23	7P20541633	20541633	0.40	0.00	0.00	0.00	0.20	0.18	0.00	0.00	0.00	0.08	0.54	-	-	-	-	0.62	0.16	0.05	0.05	0.05	0.16	M	0.16	0.05	0.05	0.05	0.16	M
7	C7P23	7P20541635	20541635	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M	
7	C7P23	7P20541676	20541676	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M	
7	C7P23	7P20541910	20541910	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M	
7	C7P23	7P20541931	20541931	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M	
7	C7P23	7P20541968	20541968	0.48	0.00	0.00	0.00	0.29	0.27	0.00	0.00	0.00	0.12	0.44	-	-	-	-	0.60	0.26	0.07	0.07	0.07	0.26	M	0.26	0.07	0.07	0.07	0.26	M
7	C7P23	7P20541969	20541969	0.48	0.00	0.00	0.00	0.29	0.27	0.00	0.00	0.00	0.12	0.44	-	-	-	-	0.60	0.26	0.07	0.07	0.07	0.26	M	0.26	0.07	0.07	0.07	0.26	M
7	C7P23	7P20541977	20541977	0.35	0.00	0.00	0.00	0.17	0.27	0.00	0.00	0.00	0.12	0.22	-	-	-	-	0.34	0.13	0.04	0.04	0.04	0.13	M	0.13	0.04	0.04	0.04	0.13	M
7	C7P23	7P20541982	20541982	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
7	C7P23	7P20542019	20542019	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-0.02	0.02	0.01	0.01	0.01	0.02	M	0.02	0.01	0.01	0.01	0.02	M	
7	C7P23	Average		0.18	0.00	0.00	0.04	0.25	0.11	0.00	0.00	0.08	0.05	0.25	-	-	-1.00	0.52	0.39	0.09	0.09	0.11				0.39	0.09	0.09	0.11		
7	C7P23	SD		0.20	0.00	0.00	0.14	0.19	0.11	0.00	0.00	0.28	0.05	0.27	-	-	-	0.42	0.43	0.08	0.08	0.10				0.43	0.08	0.08	0.10		
7	C7P23	CI		0.11	-	-	0.08	0.10	0.06	-	-	0.15	0.02	0.19	-	-	-	0.23	0.23	0.04	0.04	0.06				0.23	0.04	0.04	0.06		
8	C8P1	8P44289	44289	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
8	C8P1	8P44309	44309	0.35	0.28	0.00	0.00	0.45	0.45	0.33	0.00	0.00	0.31	-0.29	-0.20	-	-	-	0.32	0.03	0.07	0.77	0.30	0.77	C	0.24	0.04	0.89	0.24	0.89	C
8	C8P1	8P44312	44312	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20										

Table S.II.5. continued

8	C8P7	8P3997998	3997998	0.17	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.05	0.02	0.02	0.02	0.05	M			
8	C8P7	8P3998008	3998008	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic			
8	C8P7	8P3998012	3998012	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
8	C8P7	8P3998031	3998031	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic			
8	C8P7	8P3998069	3998069	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
8	C8P7	8P3998105	3998105	0.40	0.11	0.00	0.00	0.59	0.36	0.11	0.00	0.00	0.21	0.08	-0.06	-	-	-	0.65	0.30	0.22	0.68	0.29	0.68	C	0.39	0.29	0.67	0.29	0.67	C		
8	C8P7	8P3998116	3998116	0.40	0.11	0.00	0.00	0.43	0.36	0.11	0.00	0.00	0.21	0.08	-0.06	-	-	-	0.52	0.10	0.16	0.82	0.22	0.82	C	0.10	0.15	0.85	0.21	0.85	C		
8	C8P7	8P3998117	3998117	0.00	0.48	0.00	0.00	0.52	0.00	0.33	0.00	0.00	0.13	-	0.30	-	-	-	0.76	0.26	0.23	0.72	0.26	0.72	C	0.27	0.29	0.70	0.27	0.70	C		
8	C8P7	8P3998124	3998124	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
8	C8P7	8P3998130	3998130	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C			
8	C8P7	8P3998161	3998161	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-	0.02	0.01	0.01	0.01	0.01	0.02	M	0.03	0.01	0.01	0.01	0.03	M		
8	C8P7	8P3998172	3998172	0.43	0.11	0.00	0.00	0.28	0.45	0.11	0.00	0.00	0.25	-0.05	-0.06	-	-	-	0.10	0.16	0.01	0.07	0.07	0.16	N	0.21	0.06	0.06	0.06	0.21	M		
8	C8P7	8P3998181	3998181	0.00	0.00	0.00	0.50	0.04	0.09	0.00	0.00	1.00	0.04	-	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
8	C8P7	8P3998191	3998191	0.00	0.44	0.00	0.00	0.38	0.00	0.22	0.00	0.00	0.08	-	0.50	-	-	-	0.78	0.13	0.50	0.12	0.12	0.50	P	0.14	0.60	0.14	0.14	0.60	P		
8	C8P7	8P3998196	3998196	0.00	0.48	0.00	0.00	0.35	0.00	0.33	0.00	0.00	0.13	-	0.30	-	-	-	0.65	0.11	0.44	0.11	0.11	0.44	P	0.13	0.52	0.13	0.13	0.52	P		
8	C8P7	8P3998212	3998212	0.00	0.11	0.00	0.00	0.04	0.09	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P		
8	C8P7	8P3998214	3998214	0.40	0.20	0.00	0.00	0.28	0.36	0.22	0.00	0.00	0.25	0.08	-0.13	-	-	-	-	0.11	0.00	0.07	0.07	0.11	N	0.08	0.00	0.05	0.05	0.08	N		
8	C8P7	8P3998215	3998215	0.00	0.11	0.00	0.00	0.04	0.09	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P		
8	C8P7	8P3998217	3998217	0.09	0.00	0.00	0.00	0.04	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	-	-	0.02	0.01	0.01	0.01	0.01	0.02	M	-	-	-	-	-	N	
8	C8P7	8P3998235	3998235	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
8	C8P7	8P3998236	3998236	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
8	C8P7	8P3998237	3998237	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
8	C8P7	8P3998239	3998239	0.17	0.28	0.00	0.00	0.19	0.18	0.33	0.00	0.00	0.21	-0.10	-0.20	-	-	-	-	-0.12	0.00	0.05	0.04	0.04	0.05	N	0.00	0.06	0.05	0.05	0.06	N	
8	C8P7	8P3998286	3998286	0.35	0.11	0.00	0.00	0.22	0.27	0.11	0.00	0.00	0.17	0.22	-0.06	-	-	-	-	0.25	0.11	0.02	0.04	0.04	0.11	N	0.12	0.02	0.05	0.05	0.12	N	
8	C8P7	8P3998304	3998304	0.24	0.00	0.00	0.00	0.12	0.09	0.00	0.00	0.00	0.04	0.61	-	-	-	-	-	0.64	0.07	0.02	0.02	0.02	0.07	M	0.08	0.03	0.03	0.03	0.08	M	
8	C8P7	8P3998309	3998309	0.00	0.20	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	1.00	-	-	-	-	1.00	0.02	0.06	0.02	0.02	0.06	P	0.02	0.07	0.02	0.02	0.07	P	
8	C8P7	8P3998312	3998312	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
8	C8P7	8P3998313	3998313	0.40	0.44	0.00	0.00	0.37	0.36	0.11	0.00	0.00	0.21	0.08	0.75	-	-	-	-	0.44	0.07	0.04	0.08	0.08	0.08	N	0.06	0.03	0.05	0.05	0.06	N	
8	C8P7	8P3998320	3998320	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
8	C8P7	8P3998348	3998348	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-	-	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P	
8	C8P7	8P3998355	3998355	0.00	0.20	0.00	0.00	0.08	0.00	0.22	0.00	0.00	0.08	-	-0.13	-	-	-	-	-	0.02	0.06	0.02	0.02	0.06	P	0.02	0.07	0.02	0.02	0.07	P	
8	C8P7	8P3998357	3998357	0.40	0.20	0.00	0.00	0.28	0.36	0.22	0.00	0.00	0.25	0.08	-0.13	-	-	-	-	0.10	0.11	0.00	0.07	0.07	0.11	N	0.08	0.00	0.05	0.05	0.08	N	
8	C8P7	8P3998402	3998402	0.00	0.20	0.00	0.00	0.08	0.00	0.22	0.00	0.00	0.08	-	-0.13	-	-	-	-	-	0.02	0.06	0.02	0.02	0.06	P	0.02	0.07	0.02	0.02	0.07	P	
8	C8P7	8P3998411	3998411	0.00	0.49	0.00	0.00	0.32	0.00	0.38	0.00	0.00	0.13	-	0.24	-	-	-	-	0.59	0.10	0.39	0.10	0.10	0.39	P	0.12	0.47	0.12	0.12	0.47	P	
8	C8P7	8P3998412	3998412	0.30	0.11	0.00	0.00	0.19	0.18	0.11	0.00	0.00	0.13	0.39	-0.06	-	-	-	-	0.33	0.07	0.00	0.04	0.04	0.07	N	0.05	0.02	0.02	0.02	0.05	M	
8	C8P7	8P3998436	3998436	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
8	C8P7	8P3998439	3998439	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
8	C8P7	8P3998460	3998460	0.53	0.29	0.00	0.00	0.38	0.18	0.11	0.00	0.00	0.13	0.58	0.62	-	-	-	-	0.64	0.10	0.00	0.01	0.07	0.10	N	0.09	0.01	-0.03	0.06	0.09	N	
8	C8P7	Average		0.11	0.12	0.00	0.03	0.20	0.09	0.09	0.00	0.05	0.08	0.21	0.13	-	-	-	-	1.00	0.53	0.12	0.13	0.34	0.19		0.13	0.15	0.35	0.19			
8	C8P7	SD		0.17	0.16	0.00	0.11	0.14	0.14	0.12	0.00	0.23	0.09	0.32	0.34	-	-	-	-	0.00	0.44	0.08	0.13	0.42	0.26		0.09	0.15	0.42	0.26			
8	C8P7	CI		0.05	0.05	-	0.04	0.04	0.05	0.04	-	0.07	0.03	0.17	0.15	-	-	-	-	-	0.04	0.13	0.08					0.03	0.05	0.14	0.08		
8	C8P13	8P14277656	14277656	0.00	0.00	0.32	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.03	0.03	0.11	0.03	0.11	C	0.03	0.03	0.11	0.03	0.11	C		
8	C8P13	8P14277701	14277701	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic
8	C8P13	8P14277706	14277706	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-	-	0.01	0.03	0.01	0.01	0.03	P	-	-	-	-	-	N	
8	C8P13	8P14277747	14277747	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
8	C8P1																																

Table S.II.5. continued

8	C8P14	8P16570144	16570144	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P	
8	C8P14	8P16570181	16570181	0.50	0.00	0.00	0.00	0.31	0.11	0.00	0.00	0.00	0.04	0.78	-	-	-	0.86	0.33	0.09	0.09	0.09	0.33	M	1.00	0.20	0.20	0.20	1.00	M	
8	C8P14	8P16570189	16570189	0.00	0.00	0.42	0.00	0.12	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	0.64	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C	
8	C8P14	8P16570190	16570190	0.00	0.00	0.42	0.00	0.12	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	0.64	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C	
8	C8P14	8P16570195	16570195	0.35	0.00	0.32	0.00	0.38	0.00	0.00	0.00	0.00	0.00	1.00	-	1.00	-	1.00	0.00	0.21	0.54	0.21	0.54	C	0.15	0.15	0.67	0.15	0.67	C	
8	C8P14	8P16570199	16570199	0.00	0.50	0.00	-	0.26	0.00	1.00	0.00	-	0.30	-	-1.00	-	-	-0.18	0.14	0.33	0.14	-	0.33	P	-	0.33	0.33	-	0.33	P	
8	C8P14	8P16570200	16570200	0.50	0.00	0.00	-	0.28	0.00	0.00	0.00	-	0.00	1.00	-	-	-	1.00	0.33	0.14	0.14	-	0.33	M	-	-	-	-	-	N	
8	C8P14	8P16570201	16570201	0.38	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.14	0.04	0.04	0.04	0.14	M	-	-	-	-	-	N	
8	C8P14	8P16570202	16570202	0.38	0.00	0.00	0.00	0.16	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.14	0.04	0.04	0.04	0.14	M	-	-	-	-	-	N	
8	C8P14	8P16570203	16570203	0.20	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.06	0.02	0.02	0.02	0.06	M	0.14	0.04	0.04	0.04	0.14	M	
8	C8P14	8P16570208	16570208	0.32	0.00	0.00	-	0.39	0.00	0.00	0.00	-	0.00	1.00	-	-	-	1.00	0.67	0.25	0.25	-	0.67	M	0.33	0.14	0.14	-	0.33	M	
8	C8P14	8P16570209	16570209	0.28	0.00	0.00	0.00	0.46	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.26	0.44	0.44	0.57	0.57	Mic	0.17	0.41	0.41	0.60	0.60	Mic	
8	C8P14	8P16570212	16570212	0.28	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.09	0.03	0.03	0.03	0.09	M	0.14	0.04	0.04	0.04	0.14	M	
8	C8P14	8P16570213	16570213	0.50	0.00	0.00	0.00	0.33	0.25	0.00	0.00	0.11	0.50	-	-	-	-	0.68	0.33	0.09	0.09	0.09	0.33	M	0.60	0.14	0.14	0.14	0.60	M	
8	C8P14	8P16570235	16570235	0.48	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.13	0.23	0.23	0.79	0.79	Mic	0.27	0.27	0.27	0.71	0.71	Mic	
8	C8P14	8P16570237	16570237	0.42	0.00	0.00	0.00	0.15	0.20	0.00	0.00	0.06	0.52	-	-	-	-	0.64	0.18	0.05	0.05	0.05	0.18	M	0.33	0.09	0.09	0.09	0.33	M	
8	C8P14	8P16570239	16570239	0.42	0.00	0.00	0.00	0.39	0.20	0.00	0.00	0.08	0.52	-	-	-	-	0.81	0.54	0.13	0.13	0.13	0.54	M	0.33	0.09	0.09	0.09	0.33	M	
8	C8P14	8P16570241	16570241	0.38	0.25	0.00	0.00	0.28	0.17	0.00	0.00	0.07	0.56	1.00	-	-	-	0.76	0.08	0.01	0.07	0.07	0.08	N	0.16	0.00	0.09	0.09	0.16	N	
8	C8P14	8P16570243	16570243	0.25	0.22	0.00	0.00	0.21	0.00	0.00	0.00	0.00	1.00	1.00	-	-	-	1.00	0.03	0.02	0.05	0.05	0.05	N	0.02	0.07	0.02	0.02	0.07	P	
8	C8P14	8P16570244	16570244	0.00	0.00	0.32	0.00	0.29	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.15	0.15	0.67	0.15	0.67	C	0.15	0.15	0.67	0.15	0.67	C	
8	C8P14	8P16570245	16570245	0.00	0.34	0.00	0.00	0.14	0.00	0.14	0.00	0.00	0.05	-	0.58	-	-	-	0.64	0.04	0.12	0.04	0.04	0.12	P	0.04	0.12	0.04	0.04	0.12	P
8	C8P14	8P16570246	16570246	0.00	0.22	0.00	0.00	0.06	0.00	0.25	0.00	0.00	0.06	-	-	-	-	-	0.03	0.02	0.07	0.02	0.02	0.07	P	0.02	0.07	0.02	0.02	0.07	P
8	C8P14	8P16570249	16570249	0.00	0.44	0.00	-	0.15	0.00	0.00	0.00	-	0.00	-	1.00	-	-	-	1.00	0.09	0.20	0.09	-	0.20	P	0.09	0.20	0.09	-	0.20	P
8	C8P14	8P16570252	16570252	0.00	0.00	0.32	-	0.17	0.00	0.00	0.00	-	0.00	-	-	1.00	-	-	1.00	0.05	0.05	0.11	-	0.11	C	0.05	0.05	0.11	-	0.11	C
8	C8P14	8P16570254	16570254	0.49	0.00	0.00	-	0.44	0.00	0.00	0.00	-	0.00	1.00	-	-	-	1.00	0.40	0.17	0.17	-	0.40	M	1.00	0.33	0.33	-	1.00	M	
8	C8P14	8P16570291	16570291	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P
8	C8P14	8P16570334	16570334	0.00	0.11	0.00	0.00	0.12	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	0.65	0.20	0.15	-0.18	0.96	0.96	Mic	0.20	0.15	-0.18	0.96	0.96	Mic
8	C8P14	8P16570340	16570340	0.44	0.11	0.00	0.00	0.45	0.44	0.11	0.00	0.00	0.21	0.00	-0.06	-	-	-	0.54	0.26	0.21	0.03	0.69	0.69	Mic	0.61	0.27	0.13	0.63	0.63	Mic
8	C8P14	8P16570353	16570353	0.00	0.00	0.42	0.00	0.12	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	-	0.64	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C
8	C8P14	8P16570366	16570366	0.40	0.00	0.00	0.00	0.19	0.33	0.00	0.00	0.00	0.13	0.17	-	-	-	-	0.33	0.16	0.05	0.05	0.05	0.16	M	0.23	0.07	0.07	0.23	0.23	M
8	C8P14	8P16570377	16570377	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	-	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic
8	C8P14	8P16570382	16570382	0.20	0.00	0.00	0.00	0.38	0.22	0.00	0.00	0.08	-0.13	-	-	-	-	-	0.78	0.07	0.23	0.93	0.23	0.93	C	0.20	0.20	1.00	0.20	1.00	C
8	C8P14	8P16570385	16570385	0.20	0.00	0.00	0.00	0.08	0.22	0.00	0.00	0.08	-0.09	-	-	-	-	-	-0.03	0.04	0.01	0.01	0.01	0.04	M	0.10	0.03	0.03	0.10	0.10	M
8	C8P14	8P16570415	16570415	0.20	0.00	0.00	0.00	0.08	0.22	0.00	0.00	0.08	-0.13	-	-	-	-	-	-0.04	0.06	0.02	0.02	0.02	0.06	M	0.14	0.04	0.04	0.04	0.14	M
8	C8P14	8P16570424	16570424	0.40	0.00	0.00	0.00	0.40	0.56	0.00	0.00	0.00	0.21	-0.39	-	-	-	-	0.47	0.57	0.14	0.14	0.14	0.57	M	1.00	0.20	0.20	0.20	1.00	M
8	C8P14	8P16570437	16570437	0.48	0.00	0.00	0.00	0.41	0.67	0.00	0.00	0.00	0.25	-0.40	-	-	-	-	0.38	0.50	0.13	0.13	0.13	0.50	M	0.80	0.17	0.17	0.17	0.80	M
8	C8P14	8P16570439	16570439	0.51	0.00	0.00	0.00	0.25	0.67	0.00	0.00	0.00	0.25	-0.30	-	-	-	-	0.02	0.21	0.06	0.06	0.06	0.21	M	0.36	0.10	0.10	0.10	0.36	M
8	C8P14	8P16570453	16570453	0.18	0.00	0.00	0.00	0.05	0.20	0.00	0.00	0.00	0.05	-0.11	-	-	-	-	-0.03	0.05	0.02	0.02	0.02	0.05	M	0.33	0.09	0.09	0.09	0.33	M
8	C8P14	8P16570504	16570504	0.20	0.00	0.42	0.00	0.31	0.22	0.00	0.20	0.00	0.13	-0.13	-	0.52	-	-	0.59	0.03	0.16	0.47	0.16	0.47	N	0.13	0.13	0.54	0.13	0.54	C
8	C8P14	8P16570509	16570509	0.20	0.00	0.00	0.00	0.08	0.22	0.00	0.00	0.08	-0.13	-	-	-	-	-	-0.04	0.06	0.02	0.02	0.02	0.06	M	0.14	0.04	0.04	0.04	0.14	M
8	C8P14	8P16570526	16570526	0.00	0.40	0.00	0.00	0.19	0.00	0.33	0.00	0.00	0.13	-	0.17	-	-	-	0.33	0.05	0.16	0.05	0.05	0.16	P	0.05	0.16	0.05	0.05	0.16	P
8	C8P14	8P16570528	16570528	0.20	0.00	0.00	0.00	0.38	0.22	0.00	0.00	0.08	-0.13	-	-	-	-	-	0.78	0.07	0.23	0.93	0.23	0.93	C	0.20	0.20	1.00	0.20	1.00	C
8	C8P14	8P16570548	16570548	0.00	0.00	0.42	0.00	0.12	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	-	0.64	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C
8	C8P14	8P16570553	16570553	0.00	0.00	0.42	0.00	0.12	0.00	0.00	0.20	0.00	0.04	-	-	0.52	-	-	0.64	0.05	0.05	0.18	0.05	0.18	C	0.05	0.05	0.18	0.05	0.18	C
8	C8P14	8P16570575	16570575	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	-0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P
8	C8P14	Average		0.22	0.06	0.08	0.00	0.21	0.1																						

Table S.II.5. continued

8	C8P24	8P24303353	24303353	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P		
8	C8P24	8P24303362	24303362	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic		
8	C8P24	8P24303371	24303371	0.00	0.00	0.32	0.00	0.26	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.15	0.15	0.67	0.15	0.67	C	0.15	0.15	0.67	0.15	0.67	C	
8	C8P24	8P24303412	24303412	0.00	0.00	0.32	0.00	0.26	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.15	0.15	0.67	0.15	0.67	C	0.15	0.15	0.67	0.15	0.67	C	
8	C8P24	8P24303437	24303437	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M		
8	C8P24	8P24303492	24303492	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.92	0.91	0.19	0.19	0.19	0.91	M	0.91	0.19	0.19	0.19	0.91	M	
8	C8P24	8P24303512	24303512	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.50	0.50	0.50	0.50	0.50	N	0.50	0.50	0.50	0.50	0.50	N		
8	C8P24	8P24303560	24303560	0.00	0.00	0.32	0.00	0.26	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.15	0.15	0.67	0.15	0.67	C	0.15	0.15	0.67	0.15	0.67	C	
8	C8P24	8P24303581	24303581	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M		
8	C8P24	8P24303599	24303599	0.00	0.00	0.00	0.00	0.45	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	1.00	0.20	0.20	1.00	P	0.20	1.00	0.20	0.20	1.00	P		
8	C8P24	8P24303621	24303621	0.17	0.00	0.00	0.00	0.47	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.83	0.18	0.18	0.18	0.83	M	0.83	0.18	0.18	0.18	0.83	M	
8	C8P24	8P24303628	24303628	0.00	0.00	0.00	0.00	0.36	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.50	0.50	0.50	0.50	0.50	N	0.50	0.50	0.50	0.50	0.50	N		
8	C8P24	8P24303636	24303636	0.00	0.40	0.00	0.00	0.17	0.00	0.33	0.00	0.00	0.12	-	0.17	-	-	-	0.34	0.05	0.16	0.05	0.05	0.16	P	0.05	0.16	0.05	0.05	0.16	P
8	C8P24	8P24303677	24303677	0.00	0.00	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	1.00	0.20	0.20	0.20	1.00	M	1.00	0.20	0.20	0.20	1.00	M		
8	C8P24	8P24303689	24303689	0.00	0.00	0.32	0.00	0.62	0.00	0.00	0.00	0.00	0.00	-	-	1.00	-	1.00	0.67	0.29	0.39	0.29	0.67	M	0.67	0.29	0.39	0.29	0.67	M	
8	C8P24	Average	0.02	0.03	0.09	0.00	0.38	0.02	0.02	0.02	0.00	0.02	-0.07	0.17	1.00	-	-	0.94	0.50	0.34	0.33	0.28		0.50	0.34	0.33	0.28				
8	C8P24	SD	0.05	0.10	0.15	0.00	0.15	0.05	0.09	0.00	0.00	0.04	0.04	-	0.00	-	-	0.17	0.37	0.29	0.21	0.23		0.37	0.29	0.21	0.23				
8	C8P24	CI	0.02	0.05	0.07	-	0.07	0.03	0.04	-	-	0.02	0.05	-	-	-	-	0.09	0.19	0.15	0.11	0.12		0.19	0.15	0.11	0.12				
9	C9P9	9P4699283	4699283	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
9	C9P9	9P4699292	4699292	0.00	0.28	0.00	0.50	0.43	0.00	0.11	0.00	1.00	0.08	-	0.60	-	-	1.00	0.29	0.44	0.29	0.05	0.44	N	0.32	0.60	0.32	0.04	0.60	P	
9	C9P9	9P4699327	4699327	0.00	0.11	0.00	0.00	0.49	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	0.92	0.96	0.12	0.21	0.21	0.96	M	1.00	0.20	0.20	0.20	1.00	M	
9	C9P9	9P4699334	4699334	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	1.00	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
9	C9P9	9P4699339	4699339	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	1.00	1.00	1.00	Mic		
9	C9P9	9P4699403	4699403	0.00	0.11	0.00	0.00	0.04	0.00	0.11	0.00	0.00	0.04	-	-0.06	-	-	-	0.02	0.01	0.03	0.01	0.01	0.03	P	0.01	0.03	0.01	0.01	0.03	P
9	C9P9	9P4699415	4699415	0.09	0.11	0.00	0.00	0.49	0.09	0.11	0.00	0.00	0.08	-0.05	-0.06	-	-	-	0.84	0.88	0.11	0.20	0.20	0.88	M	0.89	0.19	0.19	0.19	0.89	M
9	C9P9	9P4699423	4699423	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
9	C9P9	9P4699440	4699440	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	1.00	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
9	C9P9	9P4699443	4699443	0.00	0.20	0.00	0.00	0.07	0.00	0.22	0.00	0.00	0.08	-	-0.13	-	-	-	0.04	0.02	0.06	0.02	0.02	0.06	P	0.02	0.07	0.02	0.02	0.07	P
9	C9P9	9P4699463	4699463	0.00	0.44	0.00	0.00	0.39	0.00	0.44	0.00	0.00	0.15	-	0.00	-	-	-	0.61	0.38	0.11	0.38	0.64	0.64	Mic	0.41	0.17	0.41	0.60	0.60	Mic
9	C9P9	9P4699520	4699520	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-	1.00	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
9	C9P9	9P4699529	4699529	0.17	0.48	0.00	0.00	0.41	0.18	0.78	0.00	0.00	0.35	-0.10	-0.64	-	-	-	0.16	0.23	0.06	0.40	0.62	0.62	Mic	0.37	0.09	0.37	0.66	0.66	Mic
9	C9P9	9P4699531	4699531	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
9	C9P9	9P4699544	4699544	0.17	0.35	0.00	0.00	0.20	0.18	0.22	0.00	0.00	0.15	-0.10	0.36	-	-	-	0.25	0.00	0.08	0.06	0.06	0.08	N	0.03	0.10	0.03	0.03	0.10	P
9	C9P9	9P4699570	4699570	0.30	0.11	0.00	0.00	0.46	0.00	0.11	0.00	0.00	0.04	1.00	-0.06	-	-	-	0.92	0.66	0.09	0.17	0.17	0.66	M	1.00	0.20	0.20	0.20	1.00	M
9	C9P9	9P4699582	4699582	0.00	0.28	0.00	0.00	0.11	0.00	0.33	0.00	0.00	0.12	-	-0.20	-	-	-	0.06	0.03	0.09	0.03	0.03	0.09	P	0.03	0.10	0.03	0.03	0.10	P
9	C9P9	9P4699593	4699593	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C		
9	C9P9	9P4699608	4699608	0.17	0.59	0.00	0.00	0.38	0.18	0.22	0.00	0.00	0.15	-0.10	0.63	-	-	-	0.59	0.03	0.27	-0.02	0.09	0.27	N	0.05	0.33	-0.03	0.10	0.33	N
9	C9P9	9P4699609	4699609	0.09	0.35	0.00	0.00	0.42	0.09	0.00	0.00	0.00	0.04	-0.05	1.00	-	-	-	0.91	0.11	0.61	-0.13	0.15	0.61	P	0.12	0.75	-0.15	0.17	0.75	P
9	C9P9	9P4699610	4699610	0.00	0.65	0.00	0.00	0.58	0.00	0.22	0.00	0.00	0.08	-	0.66	-	-	-	0.87	0.30	0.14	0.60	0.30	0.60	C	0.32	0.18	0.56	0.32	0.56	C
9	C9P9	9P4699611	4699611	0.00	0.44	0.00	0.00	0.18	0.00	0.11	0.00	0.00	0.04	-	0.75	-	-	-	0.78	0.04	0.12	0.02	0.04	0.12	P	0.04	0.13	0.02	0.04	0.13	P
9	C9P9	9P4699612	4699612	0.00	0.20	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	1.00	-	-	-	1.00	0.02	0.06	0.02	0.02	0.06	P	0.02	0.07	0.02	0.02	0.07	P
9	C9P9	9P4699613	4699613	0.17	0.20	0.00	0.00	0.14	0.18	0.00	0.00	0.00	0.08	-0.10	1.00	-	-	-	0.46	0.01	0.02	0.03	0.03	0.03	N	-	-	-	-	-	N
9	C9P9	9P4699615	4699615	0.17	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.05	0.02	0.02	0.02	0.05	M	0.06	0.02	0.02	0.02	0.06	M	
9	C9P9	9P4699616	4699616	0.48	0.20	0.00	0.00	0.43	0.09	0.00	0.00	0.00	0.04	0.81	1.00	-	-	-	0.91	0.37	0.05	0.11	0.11	0.37	N	0.57	0.14	0.14	0.14	0.57	M
9	C9P9	9P4699617	4699617	0.09	0.20	0.00	0.00	0.11	0.09	0.00	0.00	0.00	0.04	-0.05	1.00	-	-	-	0.65	0.01	0.05	-0.01	0.02	0.05	N	0.03	0.01	0.01	0.01	0.03	M
9	C9P9	9P4699618	4699618	0.17	0.20	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	1.00	1.00	-	-	-	1.00	0.03	0.04	0.01	0.03	0.04	N	0.06	0.02				

Table S.II.5. continued

9	C9P2846	9P15593937	15593937	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic	
9	C9P2846	9P15594032	15594032	0.24	0.00	0.00	0.00	0.49	0.27	0.00	0.00	0.00	0.12	-0.16	-	-	-	-0.77	0.29	0.45	0.55	0.45	0.55	C	0.50	0.50	0.50	0.50	0.50	N	
9	C9P2846	9P15594052	15594052	0.17	0.40	0.00	0.00	0.41	0.18	0.33	0.00	0.00	0.19	-0.10	0.17	-	-	-	0.53	0.04	0.51	0.16	0.16	0.51	P	0.14	0.57	0.14	0.14	0.57	P
9	C9P2846	9P15594066	15594066	0.18	0.40	0.00	0.00	0.42	0.20	0.33	0.00	0.00	0.20	-0.11	0.17	-	-	-	0.52	0.04	0.50	0.16	0.16	0.50	P	0.14	0.57	0.14	0.14	0.57	P
9	C9P2846	9P15594080	15594080	0.17	0.00	0.00	0.00	0.49	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.35	0.47	0.53	0.47	0.53	C	0.50	0.50	0.50	0.50	0.50	N	
9	C9P2846	9P15594112	15594112	0.00	0.40	0.00	0.00	0.17	0.00	0.33	0.00	0.00	0.12	-	0.17	-	-	-	0.34	0.05	0.16	0.05	0.05	0.16	P	0.05	0.16	0.05	0.05	0.16	P
9	C9P2846	9P15594145	15594145	0.24	0.00	0.18	0.50	0.48	0.27	0.00	0.20	1.00	0.19	-0.16	-	-0.11	-1.00	0.60	0.44	0.32	0.16	0.03	0.44	N	0.46	0.33	0.16	0.03	0.46	N	
9	C9P2846	9P15594153	15594153	0.09	0.00	0.00	0.00	0.48	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.92	0.42	0.48	0.52	0.48	0.52	C	0.50	0.50	0.50	0.50	0.50	N	
9	C9P2846	9P15594179	15594179	0.17	0.00	0.00	0.50	0.50	0.18	0.00	0.00	1.00	0.12	-0.10	-	-	-1.00	0.77	0.25	0.50	0.40	0.26	0.50	P	0.37	0.54	0.37	0.28	0.54	P	
9	C9P2846	9P15594187	15594187	0.17	0.00	0.00	0.00	0.49	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.35	0.47	0.53	0.47	0.53	C	0.50	0.50	0.50	0.50	0.50	N	
9	C9P2846	9P15594188	15594188	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
9	C9P2846	9P15594196	15594196	0.35	0.00	0.00	0.00	0.17	0.45	0.00	0.00	0.00	0.19	-0.29	-	-	-0.11	0.13	0.04	0.04	0.04	0.13	M	0.19	0.05	0.05	0.05	0.19	M		
9	C9P2846	9P15594258	15594258	0.46	0.00	0.18	0.00	0.41	0.55	0.00	0.20	0.00	0.27	-0.18	-	-0.11	-	0.34	0.41	0.14	0.02	0.14	0.41	N	0.40	0.14	0.02	0.14	0.40	N	
9	C9P2846	9P15594259	15594259	0.46	0.00	0.18	0.00	0.41	0.55	0.00	0.20	0.00	0.27	-0.18	-	-0.11	-	0.34	0.41	0.14	0.02	0.14	0.41	N	0.40	0.14	0.02	0.14	0.40	N	
9	C9P2846	9P15594264	15594264	0.17	0.00	0.18	0.00	0.33	0.18	0.00	0.20	0.00	0.12	-0.10	-	-0.11	-	0.65	0.07	0.20	0.76	0.20	0.76	C	0.05	0.21	0.74	0.21	0.74	C	
9	C9P2846	9P15594275	15594275	0.46	0.00	0.18	0.00	0.41	0.55	0.00	0.20	0.00	0.27	-0.18	-	-0.11	-	0.34	0.41	0.14	0.02	0.14	0.41	N	0.40	0.14	0.02	0.14	0.40	N	
9	C9P2846	9P15594277	15594277	0.17	0.00	0.00	0.00	0.49	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.35	0.47	0.53	0.47	0.53	C	0.50	0.50	0.50	0.50	0.50	N	
9	C9P2846	9P15594326	15594326	0.09	0.40	0.00	0.00	0.43	0.09	0.33	0.00	0.00	0.15	-0.05	0.17	-	-	0.64	0.33	0.14	0.42	0.59	0.59	Mic	0.40	0.15	0.40	0.61	0.61	Mic	
9	C9P2846	9P15594350	15594350	0.17	0.00	0.18	0.00	0.33	0.18	0.00	0.20	0.00	0.12	-0.10	-	-0.11	-	0.65	0.07	0.20	0.76	0.20	0.76	C	0.05	0.21	0.74	0.21	0.74	C	
9	C9P2846	9P15594364	15594364	0.17	0.00	0.00	0.00	0.49	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.35	0.47	0.53	0.47	0.53	C	0.50	0.50	0.50	0.50	0.50	N	
9	C9P2846	9P15594370	15594370	0.30	0.00	0.18	0.00	0.46	0.36	0.00	0.20	0.00	0.19	-0.22	-	-0.11	-	0.59	0.63	0.18	0.05	0.18	0.63	M	0.71	0.19	0.06	0.19	0.71	M	
9	C9P2846	Average		0.20	0.08	0.06	0.07	0.38	0.23	0.06	0.07	0.14	0.14	-0.13	0.17	-0.11	-1.00	0.58	0.27	0.30	0.31	0.31				0.33	0.32	0.30	0.32		
9	C9P2846	SD		0.14	0.16	0.09	0.18	0.14	0.17	0.13	0.10	0.36	0.08	0.06	0.00	0.00	0.00	0.29	0.17	0.17	0.26	0.23				0.20	0.19	0.24	0.24		
9	C9P2846	CI		0.06	0.07	0.04	0.08	0.06	0.07	0.06	0.04	0.15	0.03	0.03	-	-	-	0.13	0.07	0.07	0.11	0.10				0.08	0.08	0.10	0.10		
9	C9P21	9P25060299	25060299	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
9	C9P21	9P25060348	25060348	0.09	0.00	0.00	0.00	0.33	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.89	0.13	0.21	0.97	0.21	0.97	C	0.20	0.20	0.10	0.20	1.00	C	
9	C9P21	9P25060356	25060356	0.00	0.20	0.00	0.00	0.07	0.00	0.22	0.00	0.00	0.08	-	-0.13	-	-	-0.04	0.02	0.06	0.02	0.02	0.06	P	0.02	0.06	0.02	0.02	0.06	P	
9	C9P21	9P25060377	25060377	0.43	0.00	0.00	0.00	0.41	0.45	0.00	0.00	0.00	0.19	-0.05	-	-	-	0.53	0.52	0.13	0.13	0.13	0.52	M	0.75	0.17	0.17	0.17	0.75	M	
9	C9P21	9P25060404	25060404	0.00	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	1.00	0.20	1.00	C	0.20	0.20	1.00	0.20	1.00	C	
9	C9P21	9P25060473	25060473	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
9	C9P21	9P25060493	25060493	0.31	0.28	0.00	0.00	0.30	0.36	0.33	0.00	0.00	0.27	-0.16	-0.20	-	-	0.11	0.06	0.12	-0.60	0.80	0.80	Mic	0.04	0.13	-0.54	0.79	0.79	Mic	
9	C9P21	9P25060505	25060505	0.35	0.28	0.00	0.00	0.54	0.09	0.33	0.00	0.00	0.15	0.74	-0.20	-	-	0.71	0.01	0.64	-0.15	0.22	0.64	P	0.17	0.76	-0.10	0.17	0.76	P	
9	C9P21	9P25060507	25060507	0.35	0.28	0.00	0.00	0.52	0.09	0.33	0.00	0.00	0.15	0.74	-0.20	-	-	0.70	0.11	0.35	0.10	0.63	0.63	Mic	0.30	0.42	0.13	0.67	0.67	Mic	
9	C9P21	9P25060517	25060517	0.30	0.20	0.00	0.00	0.27	0.00	0.22	0.00	0.00	0.08	1.00	-0.13	-	-	0.71	0.04	0.15	0.24	0.82	0.82	Mic	0.20	0.12	0.20	0.93	0.93	Mic	
9	C9P21	9P25060538	25060538	0.09	0.00	0.00	0.00	0.33	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.89	0.13	0.21	0.97	0.21	0.97	C	0.20	0.20	1.00	1.00	1.00	C	
9	C9P21	9P25060544	25060544	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
9	C9P21	9P25060607	25060607	0.09	0.00	0.00	0.00	0.33	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.89	0.13	0.21	0.97	0.21	0.97	C	0.20	0.20	1.00	0.20	1.00	C	
9	C9P21	9P25060609	25060609	0.09	0.00	0.00	0.00	0.49	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.92	0.42	0.48	0.48	0.52	0.52	Mic	0.50	0.50	0.50	0.50	0.50	N	
9	C9P21	9P25060632	25060632	0.00	0.00	0.00	0.50	0.04	0.00	0.00	0.00	1.00	0.04	-	-	-1.00	-0.02	0.09	0.09	0.09	0.33	0.33	Mic	0.09	0.09	0.09	0.33	0.33	Mic		
9	C9P21	9P25060644	25060644	0.30	0.00	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	1.00	-	-	-	1.00	0.10	0.03	0.03	0.03	0.10	M	-	-	-	-	-	N	
9	C9P21	9P25060651	25060651	0.17	0.00	0.00	0.00	0.47	0.18	0.00	0.00	0.00	0.08	-0.10	-	-	-	0.84	0.83	0.18	0.18	0.18	0.83	M	0.87	0.18	0.18	0.18	0.87	M	
9	C9P21	9P25060670	25060670	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	-	-	-	-	1.00	0.20	0.20	0.20	1.00	1.00	Mic	0.20	0.20	0.20	1.00	1.00	Mic	
9	C9P21	9P25060674	25060674	0.09	0.00	0.00	0.00	0.33	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.89	0.13	0.21	0.97	0.21	0.97	C	0.20	0.20	1.00	1.00	1.00	C	
9	C9P21	9P25060689	25060689	0.09	0.00	0.00	0.00	0.49	0.09	0.00	0.00	0.00	0.04	-0.05	-	-	-	0.92	0.42	0.48	0.48	0.52	0.52	Mic	0.50	0.50	0.50	0.50	0.50	N	
9	C9P21	9P25060768	25060768	0.00	0.20	0.00	0.00	0.50	0.00	0.22	0.00	0.00	0.08	-	-0.13	-	-	0.85	0.46	0.32	0.54	0.46	0.54	C	0.46	0.32	0.54	0.46	0.54		

Table S.II.6. SNP/kb for each gene fragment

	Sequence size (bp)	Number of SNPs	SNP/kb
C1P1	519	22	42.39
C1P6	542	19	35.06
C1P7	500	10	20.00
C1P13	525	9	17.14
C1P16	553	11	19.89
C1P18	445	23	51.69
Chromosome 1	3084	94	30.48
C2P265	490	15	30.61
C2P10	547	21	38.39
C2P12	502	20	39.84
C2P16	538	20	37.17
C2P18	454	11	24.23
C2P25	536	6	11.19
C2P27	449	18	40.09
Chromosome 2	3516	111	31.57
C3P1	516	15	29.07
C3P3	537	15	27.93
C3P8	466	25	53.65
C3P9	520	9	17.31
C3P10	504	15	29.76
Chromosome 3	2543	79	31.07
C4P1	410	21	51.22
C4P5	525	11	20.95
C4P14	535	14	26.17
C4P16	456	10	21.93
C4P20	554	43	77.62
C4P21	366	11	30.05
C4P1431	516	25	48.45
C4P28	477	17	35.64
Chromosome 4	3839	152	39.59
C5P6	502	13	25.90
C5P12	494	13	26.32
C5P13	525	8	15.24
C5P20	480	19	39.58
C5P23	474	18	37.97
Chromosome 5	2475	71	28.69
C6P1869	462	14	30.30
C6P3	570	15	26.32
C6P6	467	8	17.13
C6P11	457	11	24.07
C6P14	469	29	61.83
C6P20	473	12	25.37
C6P26	510	16	31.37
Chromosome 6	3408	105	30.81
C7P2173	485	17	35.05
C7P8	493	46	93.31
C7P13	502	24	47.81
C7P19	542	24	44.28
C7P23	453	13	28.70
Chromosome 7	2475	124	50.10
C8P1	532	21	39.47
C8P4	486	15	30.86
C8P7	512	38	74.22
C8P13	520	17	32.69
C8P14	490	45	91.84
C8P15	550	17	30.91
C8P19	550	19	34.55
C8P24	502	15	29.88
Chromosome 8	4142	187	45.15
C9P9	490	39	79.59
C9P14	514	9	17.51
C9P15	517	15	29.01
C9P2846	540	21	38.89
C9P21	500	22	44.00
C9P31	464	24	51.72
Chromosome 9	3025	130	42.98
Ttotal	28507	1053	36.94



Table S.II.7. diversity parameters of the 105 Kaspar SNP markers within and between the four basic horticultural groups

	He					Ho					Fw					Gst				Marker
	M	P	C	Mic	Total	M	P	C	Mic	Total	M	P	C	Mic	Total	M	P	C	Mic	
1p12808366	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
1P16721343	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	1.000	0.200	0.200	0.200	M
1p16721616	0.000	0.000	0.000	0.000	0.367	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	1.000	0.200	0.200	P
1p18304219	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
1P199494	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
1p3705568	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	1.000	0.200	0.200	0.200	M
1p4381281	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.156	0.207	0.981	0.207	C
1p4381520	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
2p13928722	0.208	0.000	0.000	0.000	0.496	0.118	0.000	0.000	0.000	0.061	0.433	N	N	N	0.878	0.789	0.172	0.172	0.172	M
2p21022555	0.057	0.430	0.000	0.000	0.298	0.059	0.375	0.000	0.000	0.121	-0.030	0.127	N	N	0.593	0.089	0.506	0.136	0.136	P
2p25198777	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
2P33507145	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
2p35391693	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
3p11001478	0.111	0.000	0.000	0.000	0.397	0.118	0.000	0.000	0.000	0.061	-0.063	N	N	N	0.847	0.120	0.962	0.214	0.214	P
3p11355960	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
3p165577	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
3p165889	0.389	0.000	0.000	0.000	0.471	0.412	0.000	0.000	0.000	0.212	-0.058	N	N	N	0.549	0.006	0.838	0.267	0.267	P
3p2073865	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
3p9086569	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
4p10567300	0.000	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	1.000	0.200	0.200	0.200	M
4p14276984	0.000	0.430	0.000	0.000	0.278	0.000	0.375	0.000	0.000	0.091	N	0.127	N	N	0.673	0.129	0.524	0.129	0.129	P
4p18900442	0.057	0.000	0.000	0.000	0.500	0.059	0.000	0.000	0.000	0.030	-0.030	N	N	N	0.939	0.943	0.193	0.193	0.193	M
4p20183357	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
4p2154736	0.208	0.000	0.000	0.000	0.496	0.118	0.000	0.000	0.000	0.061	0.433	N	N	N	0.878	0.789	0.172	0.172	0.172	M
4P23090348	0.000	0.000	0.000	0.000	0.312	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.147	0.071	0.885	0.246	C
4p25620409	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
4p25620847	0.111	0.117	0.000	0.000	0.382	0.118	0.125	0.000	0.000	0.091	-0.063	-0.067	N	N	0.762	0.106	0.844	0.199	0.199	P
4p261309	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
4p261349	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
5p21042485	0.057	0.000	0.000	0.000	0.500	0.059	0.000	0.000	0.000	0.030	-0.030	N	N	N	0.939	0.943	0.193	0.193	0.193	M
5p22687304	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
5p22687428	0.057	0.000	0.000	0.000	0.500	0.059	0.000	0.000	0.000	0.030	-0.030	N	N	N	0.939	0.943	0.193	0.193	0.193	M
5p33038626	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.215	0.116	0.215	0.959	mic
5P38011795	0.000	0.500	0.000	0.000	0.213	0.000	0.250	0.000	0.000	0.061	N	0.500	N	N	0.716	0.091	0.333	0.091	0.091	P
5p38011919	0.000	0.000	0.444	0.000	0.114	0.000	0.000	0.333	0.000	0.061	N	N	0.250	N	0.468	0.059	0.059	0.200	0.059	C
5p38011955	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
5p4980011	0.000	0.000	0.000	0.000	0.264	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
5p4980236	0.057	0.117	0.000	0.000	0.367	0.059	0.125	0.000	0.000	0.061	-0.030	-0.067	N	N	0.835	0.141	0.863	0.192	0.192	P
6p13421556	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
6P15755686	0.000	0.000	0.486	0.000	0.144	0.000	0.000	0.500	0.000	0.094	N	N	-0.029	N	0.349	0.075	0.075	0.263	0.075	C
6p15755895	0.057	0.459	0.000	0.000	0.264	0.059	0.429	0.000	0.000	0.125	-0.030	0.067	N	N	0.526	0.080	0.456	0.126	0.126	P
6p20377715	0.251	0.000	0.000	0.000	0.434	0.294	0.000	0.000	0.000	0.152	-0.172	N	N	N	0.651	0.048	0.907	0.236	0.236	P
6p24709277	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
6p4406341	0.117	0.000	0.000	0.000	0.404	0.125	0.000	0.000	0.000	0.063	-0.067	N	N	N	0.845	0.116	0.959	0.215	0.215	P
6p498189	0.111	0.000	0.000	0.000	0.397	0.118	0.000	0.000	0.000	0.061	-0.063	N	N	N	0.847	0.120	0.962	0.214	0.214	P
6p7496245	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
6p7496346	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
7p11128938	0.000	0.305	0.000	0.000	0.410	0.000	0.375	0.000	0.000	0.091	N	-0.231	N	N	0.778	0.574	0.235	0.433	0.433	N
7p1214781	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.215	0.116	0.215	0.959	mic
7p1902502	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.215	0.116	0.215	0.959	mic

Table S.II.7. continued

	He					Ho					Fw					Gst				Marker
	M	P	C	Mic	Total	M	P	C	Mic	Total	M	P	C	Mic	Total	M	P	C	Mic	
7p19025227	0.000	0.000	0.000	0.000	0.264	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
7p20541676	0.111	0.000	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000	1.000	N	N	N	1.000	0.889	0.186	0.186	0.186	M
7p7137853	0.000	0.000	0.153	0.000	0.278	0.000	0.000	0.167	0.000	0.030	N	N	-0.091	N	0.891	0.180	0.180	0.846	0.180	C
8p14278018	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
8P16570424	0.111	0.000	0.000	0.000	0.500	0.118	0.000	0.000	0.000	0.061	-0.063	N	N	N	0.879	0.889	0.186	0.186	0.186	M
8p18684219	0.000	0.000	0.000	0.000	0.121	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
8p18684429	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
8p20936864	0.000	0.469	0.000	0.000	0.257	0.000	0.750	0.000	0.000	0.182	N	-0.600	N	N	0.293	0.116	0.455	0.116	0.116	P
8p2427684	0.415	0.000	0.000	0.000	0.463	0.471	0.000	0.000	0.000	0.242	-0.133	N	N	N	0.476	0.545	0.133	0.133	0.133	M
8p24303677	0.061	0.000	0.000	0.000	0.500	0.063	0.000	0.000	0.000	0.031	-0.032	N	N	N	0.937	0.939	0.193	0.193	0.193	M
8p3998031	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
8p3998069	0.000	0.000	0.444	0.000	0.213	0.000	0.000	0.333	0.000	0.061	N	N	0.250	N	0.716	0.125	0.125	0.500	0.125	C
8p44391	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
8p44499	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
9p12381640	0.208	0.000	0.000	0.000	0.498	0.235	0.000	0.000	0.000	0.125	-0.133	N	N	N	0.749	0.789	0.172	0.172	0.172	M
9P13931153	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
9p13931363	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
9p15594052	0.251	0.219	0.000	0.000	0.410	0.294	0.250	0.000	0.000	0.212	-0.172	-0.143	N	N	0.483	0.031	0.686	0.205	0.205	P
9p25060404	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
9p25060473	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
9p31143176	0.000	0.000	0.000	0.000	0.114	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
9p4699283	0.000	0.000	0.000	0.000	0.298	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
2p21022776	0.000	0.000	0.000	0.000	0.117	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	0.200	1.000	mic
ACO-C601	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
AOC-C593	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
AOC-M290	0.219	0.000	0.000	0.000	0.495	0.250	0.000	0.000	0.000	0.129	-0.143	N	N	N	0.739	0.778	0.171	0.171	0.171	M
ATMR-M728	0.111	0.000	0.000	0.000	0.500	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.875	0.889	0.186	0.186	0.186	M
CCC1-P727	0.438	0.000	0.000	0.000	0.476	0.529	0.000	0.000	0.000	0.281	-0.209	N	N	N	0.409	0.000	0.805	0.283	0.283	P
CHS-P57	0.000	0.000	0.000	0.000	0.342	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	1.000	0.200	0.200	P
CiC1444-03	0.111	0.000	0.000	0.000	0.375	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.833	0.120	0.962	0.214	0.214	P
CiC1459-02	0.161	0.000	0.000	0.000	0.500	0.176	0.000	0.000	0.000	0.097	-0.097	N	N	N	0.806	0.838	0.179	0.179	0.179	M
CiC2128-01	0.291	0.000	0.000	0.000	0.492	0.353	0.000	0.000	0.000	0.188	-0.214	N	N	N	0.619	0.700	0.159	0.159	0.159	M
CiC2417-04	0.111	0.000	0.000	0.000	0.375	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.833	0.120	0.962	0.214	0.214	P
CiC2518-02	0.208	0.000	0.000	0.000	0.412	0.235	0.000	0.000	0.000	0.129	-0.133	N	N	N	0.687	0.316	0.545	0.457	0.457	N
CiC2840-01	0.111	0.000	0.000	0.000	0.500	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.875	0.889	0.186	0.186	0.186	M
CiC3056-02	0.251	0.245	0.000	0.000	0.390	0.294	0.286	0.000	0.000	0.219	-0.172	-0.167	N	N	0.439	0.028	0.659	0.201	0.201	P
CiC3674-02	0.111	0.000	0.000	0.000	0.500	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.875	0.889	0.186	0.186	0.186	M
CiC3712-01	0.327	0.000	0.000	0.000	0.488	0.412	0.000	0.000	0.000	0.219	-0.259	N	N	N	0.552	0.659	0.153	0.153	0.153	M
CiC4876-07	0.208	0.000	0.000	0.000	0.498	0.235	0.000	0.000	0.000	0.125	-0.133	N	N	N	0.749	0.789	0.172	0.172	0.172	M
CiC5164-02	0.327	0.133	0.000	0.000	0.430	0.412	0.143	0.000	0.000	0.250	-0.259	-0.077	N	N	0.418	0.014	0.740	0.233	0.233	P
CiC5796-12	0.111	0.000	0.000	0.000	0.375	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.833	0.120	0.962	0.214	0.214	P
CiC5950-02	0.000	0.000	0.000	0.000	0.498	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	1.000	0.200	0.200	0.200	M
DFR-M240	0.111	0.000	0.000	0.000	0.500	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.875	0.889	0.186	0.186	0.186	M
F3H-P30	0.000	0.153	0.000	0.000	0.299	0.000	0.167	0.000	0.000	0.033	N	-0.091	N	N	0.889	0.180	0.846	0.180	0.180	P
HYB-C433	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
LCYB-M480	0.208	0.000	0.000	0.000	0.499	0.235	0.000	0.000	0.000	0.129	-0.133	N	N	N	0.742	0.789	0.172	0.172	0.172	M
NADK2-M285	0.161	0.000	0.153	0.000	0.500	0.176	0.000	0.167	0.000	0.125	-0.097	N	-0.091	N	0.750	0.784	0.199	0.078	0.199	M
NCED3-M535	0.111	0.000	0.000	0.000	0.500	0.118	0.000	0.000	0.000	0.063	-0.063	N	N	N	0.875	0.889	0.186	0.186	0.186	M
PEPC-C328	0.000	0.000	0.000	0.000	0.271	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
PEPC-M316	0.000	0.000	0.000	0.000	0.498	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	1.000	0.200	0.200	0.200	M
PKF-C64	0.000	0.000	0.000	0.000	0.271	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
PSY-C461	0.000	0.000	0.000	0.000	0.305	0.000	0.000	0.000	0.000	0.000	N	N	N	N	1.000	0.200	0.200	1.000	0.200	C
PSY-M30	0.291	0.000	0.000	0.000	0.492	0.353	0.000	0.000	0.000	0.188	-0.214	N	N	N	0.619	0.700	0.159	0.159	0.159	M
SOS1-M50	0.057	0.000	0.000	0.000	0.500	0.059	0.000	0.000	0.000	0.031	-0.030	N	N	N	0.937	0.943	0.193	0.193	0.193	M
TSC-C80	0.057	0.000	0.000	0.000	0.292	0.059	0.000	0.000	0.000	0.032	-0.030	N	N	N	0.889	0.156	0.207	0.981	0.207	C

Ho: observed heterozygosity; He: expected heterozygosity; Nei diversity index; F<sub>w</sub>: Wright fixation index; G<sub>ST</sub>: differentiation index; M: Mandarin; P: Pummelo; C: Citron; mic: *Citrus micrantha*.

Table S.II.8. Estimation of the best number of subpopulation in structure analysis

Population structure was inferred using the Structure program version 2.3.4 (Pritchard Lab, 2014). No *a priori* population structure was defined. The linkage model option was used, with allele frequencies correlated and computed probability of the data for K estimating. Analyses were made with K value (number of subpopulations) varying from 1–10. The statistics used to select the correct K value were as in Evanno *et al.* (2005): the mean likelihood,  $L(K)$ ; the mean difference between successive likelihood values of K,  $L'(K)$ ; the absolute value of this difference,  $|L''(K)|$ ; and  $\Delta K$ , which is the mean of the absolute values of  $L''(K)$  divided by the standard deviation of  $L(K)$ . The likelihood distribution  $L(K)$  and  $\Delta K$  were the main values used to choose the optimal K value of the population. Analysis of delta K indicated that optimal results were obtained with  $K = 4$  [Additional file 5]. Minimal variability in estimated frequencies was observed between the ten runs for  $K = 4$ .

	MEDIA	DESVEST	$L'(K)$	$ L''(K) $	$\Delta K$
K=1	-6889.08	0.14757			
K=2	-4895.27	6.56236	1993.81	819.56	124.888017
K=3	-3721.02	38.9405	1174.25	509.27	13.0781418
K=4	-3056.04	1.19555	664.98	626.73	524.220165
K=5	-3017.79	2.26296	38.25	10.42	4.60458327
K=6	-2989.96	18.6591	27.83	14.29	0.76584788
K=7	-2976.42	17.5488	13.54	14.03	0.79948594
K=8	-2948.85	18.2035	27.57	13.84	0.76029357
K=9	-2935.12	20.7344	13.73	2.25	0.10851511
K=10	-2923.64	13.4779	11.48	11.48	0.85176659

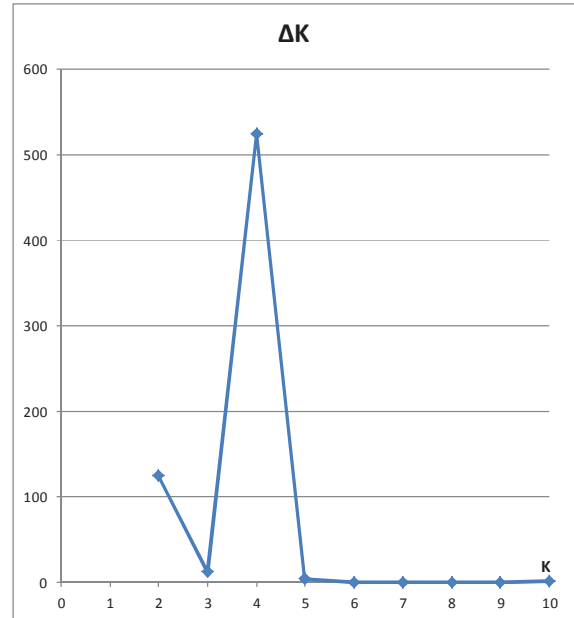


Table S.II.9. Means and confidence interval of the contribution of the four basic taxa (from 10 permuted and aligned independent Structure run cluster outputs)

Genotype	Missing Data (%)	<i>C. micrantha</i>	Citron	Mandarin	Pummelo
Fuzhu mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.963 ± 0.000	0.030 ± 0.000
Beauty mandarin	0	0.002 ± 0.000	0.003 ± 0.000	0.991 ± 0.000	0.003 ± 0.000
Shekwasha mandarin	0	0.006 ± 0.000	0.051 ± 0.000	0.931 ± 0.000	0.010 ± 0.000
Cleopatra mandarin	0	0.003 ± 0.000	0.003 ± 0.000	0.992 ± 0.000	0.002 ± 0.000
Dancy mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.963 ± 0.000	0.030 ± 0.000
King mandarin	0	0.012 ± 0.001	0.004 ± 0.000	0.750 ± 0.002	0.234 ± 0.002
Nanfengmijuu mandarin	0	0.004 ± 0.000	0.004 ± 0.000	0.984 ± 0.000	0.006 ± 5.666
Ponkan mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.958 ± 0.000	0.035 ± 0.000
Satsuma mandarin	0	0.003 ± 0.000	0.003 ± 0.000	0.722 ± 0.002	0.269 ± 0.002
Sunki mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.991 ± 0.000	0.003 ± 0.000
Willowleaf mandarin	1	0.003 ± 0.000	0.002 ± 0.000	0.953 ± 0.000	0.040 ± 0.000
Citrus daoianensis	3	0.002 ± 0.000	0.002 ± 0.000	0.991 ± 0.000	0.002 ± 0.000
Carvalho mandarin	0	0.003 ± 0.000	0.003 ± 0.000	0.815 ± 0.001	0.178 ± 0.001
Ladu mandarin	0	0.002 ± 0.000	0.002 ± 0.000	0.991 ± 0.000	0.003 ± 0.000
Se Hui Gan mandarin	0	0.002 ± 0.000	0.003 ± 0.000	0.991 ± 0.000	0.003 ± 0.000
Szibat mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.991 ± 0.000	0.003 ± 0.000
San Hu Hong Chu mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.991 ± 0.000	0.003 ± 0.000
Chandler pummelo	0	0.003 ± 0.000	0.002 ± 0.000	0.003 ± 0.000	0.991 ± 0.000
Da Xanh pummelo	0	0.003 ± 0.000	0.004 ± 0.000	0.008 ± 0.000	0.983 ± 0.000
Deep Red pummelo	0	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.991 ± 0.000
Kao Pan pummelo	30	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.989 ± 0.000
Nam Roi pummelo	4	0.003 ± 0.000	0.003 ± 0.000	0.004 ± 0.000	0.989 ± 0.000
Pink pummelo	3	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.991 ± 0.000
Tahiti pummelo	1	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.991 ± 0.000
Timorese pummelo	0	0.003 ± 0.000	0.003 ± 0.000	0.003 ± 0.000	0.991 ± 0.000
Corsican citron	0	0.002 ± 0.000	0.991 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Buddha's hand citron	0	0.003 ± 0.000	0.991 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Etrog citron	0	0.003 ± 0.000	0.991 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Humpang citron	0	0.003 ± 0.000	0.990 ± 0.000	0.003 ± 0.000	0.003 ± 0.000
Poncire citron	2	0.003 ± 0.000	0.991 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Mac Veu de montagne citron	2	0.003 ± 0.000	0.991 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Citrus micrantha	0	0.991 ± 0.000	0.003 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Citrus micrantha	0	0.991 ± 0.000	0.003 ± 0.000	0.002 ± 0.000	0.003 ± 0.000
Combava	0	0.722 ± 0.003	0.003 ± 0.000	0.032 ± 0.003	0.241 ± 0.002
Alemow	0	0.486 ± 0.002	0.505 ± 0.002	0.003 ± 0.000	0.004 ± 0.000
Excelsa lime	0	0.440 ± 0.003	0.549 ± 0.003	0.004 ± 0.000	0.006 ± 0.000
Mexican lime	0	0.503 ± 0.002	0.490 ± 0.002	0.003 ± 0.000	0.003 ± 0.000
Sweet Palestinian lime	0	0.003 ± 0.000	0.468 ± 0.003	0.319 ± 0.004	0.208 ± 0.005
Eureka lemon	0	0.013 ± 0.001	0.502 ± 0.006	0.334 ± 0.005	0.149 ± 0.005
Lisbon lemon	4	0.014 ± 0.001	0.520 ± 0.007	0.325 ± 0.007	0.140 ± 0.003
Marrakech limonette	0	0.012 ± 0.001	0.496 ± 0.005	0.337 ± 0.004	0.152 ± 0.004
Meyer lemon	0	0.003 ± 0.000	0.506 ± 0.006	0.326 ± 0.006	0.163 ± 0.003
Rangpur lime	0	0.003 ± 0.000	0.495 ± 0.002	0.496 ± 0.002	0.004 ± 0.000
Volckamer lemon	0	0.003 ± 0.000	0.475 ± 0.002	0.516 ± 0.002	0.005 ± 0.000

Table S.II.9. continued

Genotype	Missing Data (%)	<i>C. micrantha</i>	Citron	Mandarin	Pummelo
Tarocco sweet orange	0	0.003 ± 0.000	0.003 ± 0.000	0.600 ± 0.003	0.393 ± 0.003
Valencia late sweet orange	0	0.003 ± 0.000	0.003 ± 0.000	0.602 ± 0.002	0.390 ± 0.002
Washington navel sweet orange	0	0.003 ± 0.000	0.003 ± 0.000	0.602 ± 0.002	0.391 ± 0.001
Bouquet de fleurs sour orange	1	0.003 ± 0.000	0.003 ± 0.000	0.485 ± 0.002	0.507 ± 0.002
Citrus myrtifolia	2	0.003 ± 0.000	0.003 ± 0.000	0.483 ± 0.001	0.510 ± 0.001
Sevillan sour orange	0	0.003 ± 0.000	0.003 ± 0.000	0.483 ± 0.003	0.509 ± 0.003
Duncan grapefruit	0	0.003 ± 0.000	0.003 ± 0.000	0.378 ± 0.002	0.615 ± 0.002
Marsh grapefruit	1	0.003 ± 0.000	0.003 ± 0.000	0.368 ± 0.002	0.624 ± 0.002
Star Ruby grapefruit	1	0.003 ± 0.000	0.003 ± 0.000	0.364 ± 0.001	0.629 ± 0.001
Fallglo mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.713 ± 0.001	0.280 ± 0.001
Fortune mandarin	0	0.003 ± 0.000	0.004 ± 0.000	0.740 ± 0.001	0.251 ± 0.001
Fremont mandarin	0	0.003 ± 0.000	0.002 ± 0.000	0.746 ± 0.001	0.246 ± 0.001
Wilking mandarin	0	0.004 ± 0.000	0.003 ± 0.000	0.893 ± 0.000	0.097 ± 0.000
Clemenules clementine	0	0.003 ± 0.000	0.002 ± 0.000	0.758 ± 0.003	0.235 ± 0.003
Fina clementine	0	0.003 ± 0.000	0.003 ± 0.000	0.759 ± 0.001	0.233 ± 0.001
Afourer tangor	0	0.003 ± 0.000	0.002 ± 0.000	0.869 ± 0.001	0.124 ± 0.001
Ellendale tangor	0	0.007 ± 0.000	0.003 ± 0.000	0.890 ± 0.001	0.098 ± 0.001
Kiyomi tangor	1	0.003 ± 0.000	0.003 ± 0.000	0.631 ± 0.002	0.362 ± 0.002
Murcott tangor	0	0.003 ± 0.000	0.002 ± 0.000	0.844 ± 0.001	0.149 ± 0.001
Ortanique tangor	0	0.003 ± 0.000	0.003 ± 0.000	0.713 ± 0.001	0.28 ± 0.001
Temple Tangor	0	0.003 ± 0.000	0.003 ± 0.000	0.766 ± 0.002	0.227 ± 0.002
Minneola tangelo	0	0.003 ± 0.000	0.003 ± 0.000	0.710 ± 0.002	0.282 ± 0.002
Nova tangelo	0	0.003 ± 0.000	0.002 ± 0.000	0.688 ± 0.002	0.305 ± 0.002
Orlando tangelo	0	0.003 ± 0.000	0.003 ± 0.000	0.714 ± 0.001	0.279 ± 0.002
Triumph orangelo	30	0.003 ± 0.000	0.003 ± 0.000	0.322 ± 0.003	0.670 ± 0.003
Citrus amblycarpa	1	0.478 ± 0.002	0.005 ± 0.000	0.511 ± 0.003	0.005 ± 0.000

# Annexe IV

---

Supplementary information Chapitre III

Table S.III.1. List of accessions with Swingle and Tanaka classification names, ploidy level, Mytotype, chlorotype, cytoplasmic type and nuclear cluster information.

Swingle classification	Tanaka classification	Variety	Accession number	Collection	Ploidy	Mytotype	Chlorotype	Cytoplasmic Type	Redundant genotype	Nuclear cluster
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Ambilobe lime	SRA 480	INRA-CIRAD	3	2	2	2		T1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Antillaise SG lime	SRA 503	INRA-CIRAD	2	2	2	2		C1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Coppenrhad lime	SRA 838	INRA-CIRAD	3	2	2	2		T1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Gallet lemon	SRA 1057	INRA-CIRAD	2	4	4	4		F4
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Giant Key lime	SRA 785	INRA-CIRAD	4	2	2	2		T1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Kaghzi lime	SRA 818	INRA-CIRAD	2	2	2	2		C3
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Kirk lime	SRA 696	INRA-CIRAD	2	2	2	2		Unic
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Madagascar lemon	SRA 1048	INRA-CIRAD	3	2	2	2		T1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Mexican lime	IVIA 164	IVIA	2	2	2	2		C1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Mohtasseb SG lime	SRA 504	INRA-CIRAD	3	2	2	2		T1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	New Caledonia lime	SRA 809	INRA-CIRAD	2	2	2	2		C3
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Pursha lime	SRA 822	INRA-CIRAD	2	4	4	4		Unic
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	sans épine lime	SRA 619	INRA-CIRAD	2	2	2	2		C1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus aurantifolia</i> (Christm.) Swing.	Tanepao lime	SRA 836	INRA-CIRAD	3	2	2	2		T1
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus bergamia</i> Risso & Poit.	Calabria Bergamot	IVIA 254	IVIA	2	5	5	5		Unic
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus excelsa</i> Wester	Excelsa lime	IVIA 167	IVIA	2	2	2	2	5	C2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus excelsa</i> Wester	Nestour lime	SRA 698	INRA-CIRAD	2	2	2	2		C2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus latifolia</i> (Yu. Tanaka) Tanaka	Bears lime	IVIA 124	IVIA	3	5	5	5		T2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus latifolia</i> (Yu. Tanaka) Tanaka	El Kseur lime	SRA 643	INRA-CIRAD	3	5	5	5		T2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus latifolia</i> (Yu. Tanaka) Tanaka	IAC-5 lime	SRA 618	INRA-CIRAD	3	5	5	5		T2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus latifolia</i> (Yu. Tanaka) Tanaka	Persian lime	SRA 1046	INRA-CIRAD	3	5	5	5		T2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus latifolia</i> (Yu. Tanaka) Tanaka	Tahiti lime	IVIA 569	IVIA	3	5	5	5		T2
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus limettioides</i> Tanaka	Bisri Sweet lime	SRA 772	INRA-CIRAD	2	5	6	6	4	D
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus limettioides</i> Tanaka	Brasil sweet lime	SRA 697	INRA-CIRAD	2	5	6	6	4	D

<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus limettioides</i> Tanaka	Palestinian sweet lime	IVIA 305	IVIA	2	5	6	6	D	
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus macrophylla</i> Wester	Alemow	IVIA 288	IVIA	2	2	2	2	C	
<i>Citrus aurantifolia</i> (Christm.) Swing.	<i>Citrus webberii</i> Wester	Kalpi lime	IVIA 234	IVIA	2	2	2	2	5	C2
<i>Citrus limon</i> (L.) Burm.	<i>Citrus aurata</i> Risso	Pomme d'Adam lime	SRA 855	INRA-CIRAD	2	2	2	2	C2	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus jambhiri</i> Lush.	Milam lemon	0110047	INRA-CIRAD	2	4	4	4	I	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus jambhiri</i> Lush.	Rough lemon	IVIA 333	IVIA	2	4	4	4	F4	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus jambhiri</i> Lush.	Soh-jahlia lemon	SRA 858	INRA-CIRAD	2	4	4	4	F3	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus karna</i> Raf.	Indian Khatta lime	SRA 972	INRA-CIRAD	2	4	4	4	G	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus karna</i> Raf.	Khatta Karna lime	SRA 695	INRA-CIRAD	2	4	4	4	G	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limetta</i> Risso	à mamelon limette	SRA 715	INRA-CIRAD	2	5	5	5	6	B
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limetta</i> Risso	Marrakech acid limonette	SRA 1047	INRA-CIRAD	2	5	5	5	6	B
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limetta</i> Risso	Marrakech limonette	IVIA 484	IVIA	2	5	5	5	B	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Adamo lemon	SRA 787	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Adamopoulos lemon	SRA 622	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	AK lemon	SRA 635	INRA-CIRAD	2	5	5	5	Unic	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Allen lemon	SRA 1051	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Apireno Continella lemon	SRA 915	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Asaasli lemon	SRA 632	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Butnal Sweet lemon	SRA 810	INRA-CIRAD	2	5	6	6	D	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Cerza lemon	SRA 806	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Corpaci lemon	SRA 191	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Corsigliese lemon	SRA 1054	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Demre dickensiz lemon	SRA 476	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Eureka lemon	IVIA 297	IVIA	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Femminello lemon	SRA 180	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Fino lemon	SRA 344	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Imperial lemon	SRA 831	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Interdonato lemon	SRA 633	INRA-CIRAD	2	5	5	5	A	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Karystini lemon	SRA 623	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Kulu lemon	SRA 1061	INRA-CIRAD	2	5	6	6	Unic	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Kutdiken lemon	SRA 530	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Lamas lemon	SRA 477	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Lapithou lemon	SRA 541	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Limoneira 8A lemon	IVIA 214	IVIA	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Lisbon lemon	SRA 197	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Luminciana lemon	SRA 817	INRA-CIRAD	2	5	5	5	A	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Maghzalani lemon	SRA 634	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Maglini lemon	SRA 624	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Malti lemon	SRA 636	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Menton lemon	SRA 625	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Molla Mehemet lemon	SRA 1049	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Monachello lemon	SRA 621	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Moretti lemon	SRA 1056	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Sanguin lemon	SRA 1055	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Santa Teresa lemon	SRA 626	INRA-CIRAD	2	5	5	5	A1	
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Siracusano sans épine lemon	SRA 1053	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Sweet lemon	IVIA 443	IVIA	2	5	5	5	A1	



<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Ussana sans pépins lemon	SRA 762	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Vakhalou lemon	SRA 543	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Variegated lemon	SRA 548	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Villafranca lemon	SRA 253	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Yedi-veren lemon	SRA 479	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limon</i> (L.) Burm.	Zagara Bianca lemon	SRA 739	INRA-CIRAD	2	5	5	5	1	A1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limonia</i> Osbeck	Arabie Saoudite lime	SRA 725	INRA-CIRAD	2	4	4	4		F
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limonia</i> Osbeck	India sweet lime	0110048	INRA-CIRAD	2	4	4	4		F2
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limonia</i> Osbeck	Iran sweet lime	0110019	INRA-CIRAD	2	4	4	4		F1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limonia</i> Osbeck	Rangpur jaune lime	SRA 944	INRA-CIRAD	2	4	4	4		G
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limonia</i> Osbeck	Rangpur lime	IVIA 334	IVIA	2	4	4	4		G
<i>Citrus limon</i> (L.) Burm.	<i>Citrus limonia</i> Osbeck	Volkamer lemon	IVIA 432	IVIA	2	4	4	4		F1
<i>Citrus limon</i> (L.) Burm.	<i>Citrus lumia</i> Risso & Poit.	Barum lemon	0110010	INRA-CIRAD	2	5	6	6	3	Unic
<i>Citrus limon</i> (L.) Burm.	<i>Citrus lumia</i> Risso & Poit.	Bitrouni lime	SRA 816	INRA-CIRAD	2	5	5	5		Unic
<i>Citrus limon</i> (L.) Burm.	<i>Citrus lumia</i> Risso & Poit.	Borneo lemon	SRA 984	INRA-CIRAD	2	5	6	6	3	Unic
<i>Citrus limon</i> (L.) Burm.	<i>Citrus lumia</i> Risso & Poit.	Poire du Commandeur lime	SRA 919	INRA-CIRAD	2	5	6	6		Unic
<i>Citrus limon</i> (L.) Burm.	<i>Citrus lumia</i> Risso & Poit.	S.P.Jaffa lemon	SRA 823	INRA-CIRAD	2	5	6	6		Unic
<i>Citrus limon</i> (L.) Burm.	<i>Citrus meyeri</i> Y. Tan.	Indian lemon	SRA 770	INRA-CIRAD	2	4	4	4		F2
<i>Citrus limon</i> (L.) Burm.	<i>Citrus meyeri</i> Y. Tan.	Meyer lemon	IVIA 145	IVIA	2	5	6	6		Unic
<i>Citrus limon</i> (L.) Burm.	<i>Citrus pyriformis</i> Hassk.	Ponderosa lemon	IVIA 268	IVIA	2	5	6	6		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Alikioti lime	SRA 1082	INRA-CIRAD	2	4	4	4		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Big fruit lime	SRA 768	INRA-CIRAD	2	5	5	5		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Citrus voangiala lime	SRA 947	INRA-CIRAD	2	4	4	4		F
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Damas citron	SRA 837	INRA-CIRAD	2	5	5	5		E
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Emperor lemon	SRA 1060	INRA-CIRAD	2	5	6	6		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Hybride Fourny lemon	SRA 794	INRA-CIRAD	2	5	6	6		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Hybride Mac Nao N	SRA 793	INRA-CIRAD	2	5	6	6		H
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Ichang lemon	SRA 699	INRA-CIRAD	2	5	6	6		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	India lime	SRA 830	INRA-CIRAD	2	4	4	4		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Kaghi lime	SRA 808	INRA-CIRAD	2	4	4	4		F1
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Mak nao Si citron	SRA 833	INRA-CIRAD	2	5	5	5		E
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Milan lemon	SRA 969	INRA-CIRAD	2	4	4	4		J
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Otaheite lime	SRA 828	INRA-CIRAD	2	5	6	6		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	Rhobs el Arsa citron	SRA 244	INRA-CIRAD	2	5	5	5		Unic
<i>Citrus</i> sp.	<i>Citrus</i> sp.	India lemon	SRA 1058	INRA-CIRAD	2	4	4	4		F3
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Azimboa pummelo	IVIA 420	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Chandler pummelo	IVIA 207	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Da Xhang pummelo	IVIA 589	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Deep Red pummelo	IVIA 277	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Flores pummelo	SRA 673	INRA-CIRAD	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Gil pummelo	IVIA 321	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Nam Roi pummelo	IVIA 590	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Pink pummelo	IVIA 275	IVIA	2	5	6	6		C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Sans pépin pummelo	SRA 710	INRA-CIRAD	2	5	6	6		C. maxima

<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Tahitian pummelo	SRA 727	INRA-CIRAD	2	5	6	6	C. maxima
<i>Citrus maxima</i> (Burm.) Merr.	<i>Citrus maxima</i> (Burm.) Merr.	Timorese pummelo	SRA 707	INRA-CIRAD	2	5	6	6	C. maxima
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Arizona citron	IVIA 169	IVIA	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Buddha's Hand citron	IVIA 202	IVIA	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Chinese lemon	SRA 834	INRA-CIRAD	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Corsican citron	IVIA 567	IVIA	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Diamante citron	IVIA 560	IVIA	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Mac Veu de montagne citron	SRA 760	INRA-CIRAD	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Mak nao San citron	SRA 821	INRA-CIRAD	2	1	1	1	C. medica
<i>Citrus medica</i> L.	<i>Citrus medica</i> L.	Poncire commun citron	SRA 701	INRA-CIRAD	2	1	1	1	C. medica
<i>Citrus micrantha</i> Wester	<i>Citrus micrantha</i> Wester	Micrantha	SRA 1115	INRA-CIRAD	2	2	2	2	C. micrantha
<i>Citrus micrantha</i> Wester	<i>Citrus micrantha</i> Wester	Small flowered papeda	IVIA 626	IVIA	2	2	2	2	C. micrantha
<i>Citrus reticulata</i> Blanco	<i>Citrus deliciosa</i> Ten.	Willowleaf mandarin	IVIA 154	IVIA	2	3	3	3	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus depressa</i> Hayata	Shekwasha mandarin	SRA 847	INRA-CIRAD	2	4	4	4	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus paratangerina</i> Hort. ex Tanaka	Ladu ordinaire mandarin	SRA 590	INRA-CIRAD	2	3	3	3	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus reshni</i> Hort. ex Tanaka	Cleopatra mandarin	IVIA 385	IVIA	2	4	4	4	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus reticulata</i> Blanco	Sun Chu Cha mandarin	IVIA 483	IVIA	2	4	4	4	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus suhuiensis</i> Hort. ex Tanaka	Se hui gan mandarin	SRA 586	INRA-CIRAD	2	3	3	3	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus suhuiensis</i> Hort. ex Tanaka	Szinkom mandarin	SRA 597	INRA-CIRAD	2	3	3	3	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus sunki</i> (Hayata) Hort. ex Tanaka	Sunki mandarin	IVIA 239	IVIA	2	4	4	4	C. reticulata
<i>Citrus reticulata</i> Blanco	<i>Citrus tangerina</i> Hort. ex Tanaka	Dancy mandarin	IVIA 434	IVIA	2	3	3	3	C. reticulata
<i>Citrus tachibana</i> (Mak.) Tan.	<i>Citrus erythroa</i> Hort. ex Tanaka	Fuzhu mandarin	SRA 775	INRA-CIRAD	2	3	3	3	C. reticulata
<i>Citrus tachibana</i> (Mak.) Tan.	<i>Citrus erythroa</i> Hort. Ex Tanaka	San Hu Hong Chu mandarin	SRA 769	INRA-CIRAD	2	3	3	3	C. reticulata
<i>Citrus tachibana</i> (Mak.) Tan.	<i>Citrus kinokuni</i> Hort. ex Tanaka	Nan feng mi Chu mandarin	SRA 839	INRA-CIRAD	2	3	3	3	C. reticulata
<i>Citrus aurantium</i> L.	<i>Citrus aurantium</i> L.	Bouquet de fleurs sour orange	IVIA 139	IVIA	2	5	5	5	2 C. aurantium
<i>Citrus aurantium</i> L.	<i>Citrus aurantium</i> L.	Sevillan sour orange	IVIA 117	IVIA	2	5	5	5	2 C. aurantium
<i>Citrus paradisi</i> Macf.	<i>Citrus hassaku</i> Hort. ex Y. Tan.	Hassaku lemon	SRA 819	INRA-CIRAD	2	5	6	6	H
<i>Citrus paradisi</i> Macf.	<i>Citrus paradisi</i> Macf.	Marsh grapefruit	IVIA 176	IVIA	2	5	6	6	8 C. paradisi
<i>Citrus paradisi</i> Macf.	<i>Citrus paradisi</i> Macf.	Star Ruby grapefruit	IVIA 197	IVIA	2	5	6	6	8 C. paradisi
<i>Citrus reticulata</i> Blanco	<i>Citrus clementina</i> Hort. ex Tanaka	Nules clementine	IVIA 22	IVIA	2	3	3	3	Unic
<i>Citrus sinensis</i> (L.) Osb.	<i>Citrus sinensis</i> (L.) Osb.	Shamouti sweet orange	IVIA 270	IVIA	2	5	6	6	7 C. sinensis
<i>Citrus sinensis</i> (L.) Osb.	<i>Citrus sinensis</i> (L.) Osb.	Valencia late delta sweet orange	IVIA 363	IVIA	2	5	6	6	7 C. sinensis

Table S.III.2. Marker detailed information

Marker	Scaffold	Position	Marker type	GeneBank; Phytozome accession/gene numbers	Publication reference	Specific allele	Gst tax spe	Selected 73
IDEMA	scaffold_1	7 406 773	Indel	CB417399	Garcia-Lor et al. (2012) MGG. 287:77-94	N	0.8581	N
IDHYB1	scaffold_9	29 490 887	Indel	AF315289	Garcia-Lor et al. (2012) MGG. 287:77-94	N	0.5584	N
IDHYB2	scaffold_9	29 491 230	Indel	AF315289	Garcia-Lor et al. (2012) MGG. 287:77-94	N	0.5211	N
IDLCY2	scaffold_8	19 261 284	Indel	FJ516403	Garcia-Lor et al. (2012) MGG. 287:77-94	N	0.8131	N
IDPEPC1	scaffold_2	11 091 910	Indel	EF058158	Garcia-Lor et al. (2012) MGG. 287:77-94	C. max	0.913	1
IDPEPC2	scaffold_2	11 092 145	Indel	EF058158	Garcia-Lor et al. (2012) MGG. 287:77-94	C. med	1	1
IDPSY	scaffold_6	21 395 739	Indel	AB037975	Garcia-Lor et al. (2012) MGG. 287:77-94	C. med	1	1
IDPSY2	scaffold_6	21 395 929	Indel	AB037975	Garcia-Lor et al (2013) Annals of botany. 111: 1-19	C. ret	0.4667	N
1p12808366	scaffold_1	12 808 366	SNP	Ciclev10010796m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
1P16721616	scaffold_1	16 721 616	SNP	Ciclev10007960m.g	Curk et al. (submitted), PlosOne	C. max	0.9636	1
1p3705568	scaffold_1	3 705 568	SNP	Ciclev10010157m.g	Curk et al. (submitted), PlosOne	C. ret	1	1
1p4381520	scaffold_1	4 381 520	SNP	Ciclev10008549m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
2P21022776	scaffold_2	21 022 776	SNP	Ciclev10018135m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
2p25198777	scaffold_2	25 198 777	SNP	Ciclev10015267m.g	Curk et al. (submitted), PlosOne	C. med	1	1
3p11355960	scaffold_3	11 355 960	SNP	Ciclev10023509m.g	Curk et al. (submitted), PlosOne	C. med	1	1
3p165889	scaffold_3	165 889	SNP	Ciclev10023360m.g	Curk et al. (submitted), PlosOne	C. max	0.8909	1
3p2073865	scaffold_3	2 073 865	SNP	Ciclev10020017m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
3p35931624	scaffold_3	35 931 624	SNP	Ciclev10023979m.g	Curk et al. (submitted), PlosOne	N	0.6571	N
3p9086569	scaffold_3	9 086 569	SNP	Ciclev10021171m.g	Curk et al. (submitted), PlosOne	C. med	0.75	1
4p10567300	scaffold_4	10 567 300	SNP	Ciclev10033209m.g	Curk et al. (submitted), PlosOne	C. ret	0.9592	1
4p14276984	scaffold_4	14 276 984	SNP	Ciclev10031207m.g	Curk et al. (submitted), PlosOne	C. max	0.5714	1
4p20183357	scaffold_4	20 183 357	SNP	Ciclev10033475m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
4p2154736	scaffold_4	2 154 736	SNP	Ciclev10032524m.g	Curk et al. (submitted), PlosOne	C. ret	0.8333	1
4p25620409	scaffold_4	25 620 409	SNP	Ciclev10032624m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
4P261349	scaffold_4	261 349	SNP	Ciclev10031876m.g	Curk et al. (submitted), PlosOne	C. med	0.8824	1
5p21042485	scaffold_5	21 042 485	SNP	Ciclev10003688m.g	Curk et al. (submitted), PlosOne	C. ret	0.9636	1
5p22687304	scaffold_5	22 687 304	SNP	Ciclev10001185m.g	Curk et al. (submitted), PlosOne	C. med	0.6842	1
5p22687428	scaffold_5	22 687 428	SNP	Ciclev10001185m.g	Curk et al. (submitted), PlosOne	C. ret	1	1

5p33038626	scaffold_5	33 038 626	SNP	Ciclev10001741m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
5p38011919	scaffold_5	38 011 919	SNP	Ciclev10003376m.g	Curk et al. (submitted), PlosOne	N	0.2308	N
5P38011955	scaffold_5	38 011 955	SNP	Ciclev10003376m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
6p13421556	scaffold_6	13 421 556	SNP	Ciclev10013094m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
6p24709277	scaffold_6	24 709 277	SNP	Ciclev10011921m.g	Curk et al. (submitted), PlosOne	C. med	0.6842	1
6p7496245	scaffold_6	7 496 245	SNP	Ciclev10013603m.g	Curk et al. (submitted), PlosOne	C. med	0.8824	1
6p7496346	scaffold_6	7 496 346	SNP	Ciclev10013603m.g	Curk et al. (submitted), PlosOne	C. mic	0.9672	1
7p19025227	scaffold_7	19 025 227	SNP	Ciclev10025343m.g	Curk et al. (submitted), PlosOne	C. med	0.6	1
7p6084876	scaffold_7	6 084 876	SNP	Ciclev10027284m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
7p7137438	scaffold_7	7 137 438	SNP	Ciclev10027413m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
8p2427684	scaffold_8	2 427 684	SNP	Ciclev10029965m.g	Curk et al. (submitted), PlosOne	C. ret	0.5172	1
8p3998031	scaffold_8	3 998 031	SNP	Ciclev10029698m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
8p3998069	scaffold_8	3 998 069	SNP	Ciclev10029698m.g	Curk et al. (submitted), PlosOne	C. med	0.5556	1
8p44391	scaffold_8	44 391	SNP	Ciclev10028759m.g	Curk et al. (submitted), PlosOne	C. med	1	1
9p13931153	scaffold_9	13 931 153	SNP	Ciclev10006583m.g	Curk et al. (submitted), PlosOne	C. med	0.6842	1
9p13931363	scaffold_9	13 931 363	SNP	Ciclev10006583m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
9p25060404	scaffold_9	25 060 404	SNP	Ciclev10006230m.g	Curk et al. (submitted), PlosOne	C. med	0.7778	1
9p25060473	scaffold_9	25 060 473	SNP	Ciclev10006230m.g	Curk et al. (submitted), PlosOne	C. mic	1	1
9p31143176	scaffold_9	31 143 176	SNP	Ciclev10006644m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. mic	1	1
ACO-C601	scaffold_1	22 395 958	SNP	Ciclev10007338m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	1	1
ACO-P353	scaffold_1	22 396 206	SNP	Ciclev10007338m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. max	0.375	N
AOC-C593	scaffold_6	25 003 653	SNP	Ciclev10011556m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	1	1
AOC-M290	scaffold_6	25 003 350	SNP	Ciclev10011556m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.913	1
ATMR-M728	scaffold_3	45 711 958	SNP	Ciclev10018482m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	1	1
CCC1-M85	scaffold_2	31 582 491	SNP	Ciclev10014154m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.5172	1
CCC1-P727	scaffold_2	31 581 849	SNP	Ciclev10014154m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. max	0.8305	1
CHI-M598	scaffold_4	3 336 062	SNP	Ciclev10032697m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	N	0.519	N
CHS-M183	scaffold_3	50 308 712	SNP	Ciclev10020530m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.4194	N
CHS-P57	scaffold_3	50 308 586	SNP	Ciclev10020530m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. max	1	1
CiC0599-01	scaffold_1	27 107 358	SNP	ET093125	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.1892	N

CiC1380-05	scaffold_5	10 987 419	SNP	ET072553	Ollitrault et al (2012) BMC Genomics 2012, 13:13	N	0.1905	N
CiC1444-03	scaffold_7	562 777	SNP	ET073216 / Ciclev10027576m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. max	1	1
CiC1459-02	scaffold_3	43 582 915	SNP	ET073328 / Ciclev10023341m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	1	1
CiC1749-05	scaffold_8	24 429 013	SNP	ET097636	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.4667	N
CiC2110-01	scaffold_1	3 241 154	SNP	ET099643	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.1	N
CiC2128-01	scaffold_6	19 376 782	SNP	ET111354 / Ciclev10011048m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.76	1
CiC2417-04	scaffold_5	41 973 357	SNP	ET101382 / Ciclev10002683m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. max	1	1
CiC2518-02	scaffold_9	15 424 556	SNP	ET101955 / Ciclev10007074m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	N	0.6232	N
CiC3056-02	scaffold_6	22 087 277	SNP	ET075329	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. max	0.6573	1
CiC3674-02	scaffold_7	2 639 448	SNP	ET079224 / Ciclev10024678m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	1	1
CiC3712-01	scaffold_2	27 937 128	SNP	ET079481 / Ciclev10014584m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.76	1
CiC3740-02	scaffold_4	18 979 564	SNP	ET079647	Ollitrault et al (2012) BMC Genomics 2012, 13:13	N	0.3009	N
CiC4240-04	scaffold_4	2 315 343	SNP	ET106812	Ollitrault et al (2012) BMC Genomics 2012, 13:13	N	0.1997	N
CiC4581-01	scaffold_1	17 994 845	SNP	ET109034	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.2571	N
CiC4681-02	scaffold_3	21 787 625	SNP	ET109640	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.0476	N
CiC4876-07	scaffold_9	144 478	SNP	ET080580 / Ciclev10005180m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.913	1
CiC5087-01	scaffold_9	1 955 754	SNP	ET111514	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.0811	N
CiC5164-02	Scaffold_8	4 873 415	SNP	ET111943	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. max	0.5307	1
CiC5796-12	scaffold_3	41 554 117	SNP	ET082752 / Ciclev10018633m.g	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. max	1	1
CiC5842-02	scaffold_5	37 109 141	SNP	ET083106	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	0.3333	N
CiC5950-02	scaffold_1	25 122 651	SNP	ET083949	Ollitrault et al (2012) BMC Genomics 2012, 13:13	C. ret	1	1
CiC6213-07	scaffold_4	24 870 555	SNP	ET085253	Ollitrault et al (2012) BMC Genomics 2012, 13:13	N	0.5429	N
DFR-M240	scaffold_5	41 195 542	SNP	Ciclev10001757m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	1	N
DXS-C545	scaffold_7	5 043 674	SNP	Ciclev10024949m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	0.7778	1
DXS-M618	scaffold_7	5 043 747	SNP	Ciclev10024949m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.5714	1
EMA-M30	scaffold_1	7 406 802	SNP	Ciclev10007810m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.3333	N
F3H-C341	scaffold_2	8 056 212	SNP	Ciclev10018289m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	N	0.28	N
F3H-M309	scaffold_2	8 056 244	SNP	Ciclev10018289m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.1	N
F3H-P30	scaffold_2	8 056 523	SNP	Ciclev10018289m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. max	0.9048	1

FLS-M400	scaffold_7	6 013 417	SNP	Ciclev10026028m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.3333	N
FLS-P129	scaffold_7	6 013 688	SNP	Ciclev10026028m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. max	0.5172	1
HYB-C433	scaffold_9	29 490 458	SNP	Ciclev10005481m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	1	1
HYB-M62	scaffold_9	29 490 826	SNP	Ciclev10005481m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.375	N
LAPX-M238	scaffold_6	11 706 007	SNP	Ciclev10012554m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.5172	1
LCY2-M379	scaffold_8	19 260 867	SNP	Ciclev10028245m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.4194	N
LCYB-P736	scaffold_9	22 729 397	SNP	Ciclev10004730m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. max	0.375	N
MDH-M519	scaffold_3	7 266 173	SNP	Ciclev10020378m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.6296	1
MDH-MP69	scaffold_3	7 266 623	SNP	Ciclev10020378m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	N	0.5429	N
NCED3-M535	scaffold_3	29 352 884	SNP	Ciclev10019364m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	1	1
PEPC-M316	scaffold_2	11 091 624	SNP	Ciclev10014164m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	1	1
PKF-C64	scaffold_2	35 686 712	SNP	Ciclev10014760m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	1	1
PKF-M186	scaffold_2	35 686 592	SNP	Ciclev10014760m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	N	0.5429	N
PSY-C461	scaffold_6	21 392 858	SNP	Ciclev10011841m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	1	1
PSY-M30	scaffold_6	21 393 272	SNP	Ciclev10011841m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.76	1
SOS1-M50	scaffold_2	26 306 845	SNP	Ciclev10015265m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	1	1
TRPA-M593	scaffold_2	35 896 837	SNP	Ciclev10014770m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. ret	0.8333	1
TSC-C80	scaffold_1	27 714 027	SNP	Ciclev10007428m.g	Garcia-Lor et al. (2013). APS 1(4):1200406.	C. med	1	1
CAC23	Scaffold_3	209 895	SSR	Ciclev10020034m.g	Kijas et al. TAG (1997) 94:701-706.	N	0.5296	N
CI02F07	scaffold_8	15 057 840	SSR	AJ567406	Froelicher et al. (2008) Mol. Eco.res. 8 (1): 119-122.	N	0.4589	N
CI03D12a	Scaffold_3	4 183 645	SSR	FR677577	Aleza et al 2011. Ann Bot 108(1): 37–50.	N	0.4405	N
mCrCIR01C06	scaffold_6	24 791 517	SSR	FR692356	Cuenca et al. (2011). Heredity 107, 462–470	N	0.2002	N
mCrCIR01F04a	scaffold_8	1 063 542	SSR	AM489736	Froelicher et al. (2008) Mol. Eco.res. 8 (1): 119-122.	N	0.3322	N
mCrCIR03B07	scaffold_7	11 545 990	SSR	FR677573	Cuenca et al. (2011). Heredity 107, 462–470	N	0.2868	N
mCrCIR03C08	scaffold_2	27 340 389	SSR	FR677576	Cuenca et al. (2011). Heredity 107, 462–470	N	0.1839	N
mCrCIR04H06	scaffold_2	8 097 269	SSR	FR677579	Cuenca et al. (2011). Heredity 107, 462–470	N	0.4606	N
mCrCIR07D06	scaffold_4	7 294 237	SSR	FR677581	Cuenca et al. (2011). Heredity 107, 462–470	N	0.3999	N
Mest001	scaffold_1	20 040 617	SSR	DY262452	F. Luro unpublished markers	N	0.386	N
Mest015	scaffold_8	24 850 430	SSR	FC912829	F. Luro unpublished markers	N	0.5511	N
Mest046	scaffold_2	33 532 310	SSR	DY266487	F. Luro unpublished markers	N	0.4477	N

Mest088	scaffold_5	36 035 011	SSR	DY271576	F. Luro unpublished markers	N	0.2746	N
Mest107	scaffold_7	211 621	SSR	DY274062	F. Luro unpublished markers	N	0.6377	N
Mest131	scaffold_3	50 550 369	SSR	DY276912 / Ciclev10018930m.g	F. Luro unpublished markers	N	0.4928	N
Mest192	scaffold_6	17 475 542	SSR	DY283129	F. Luro unpublished markers	N	0.4236	N
Mest488	scaffold_6	21 253 670	SSR	DY297637	F. Luro unpublished markers	N	0.409	N
TAA1	scaffold_6	24 896 450	SSR	Ciclev10011968m.g	Kijas et al. TAG (1997) 94:701-706.	N	0.427	N
TAA41	Scaffold_2	35 861 148	SSR		Kijas et al. TAG (1997) 94:701-706.	N	0.1899	N

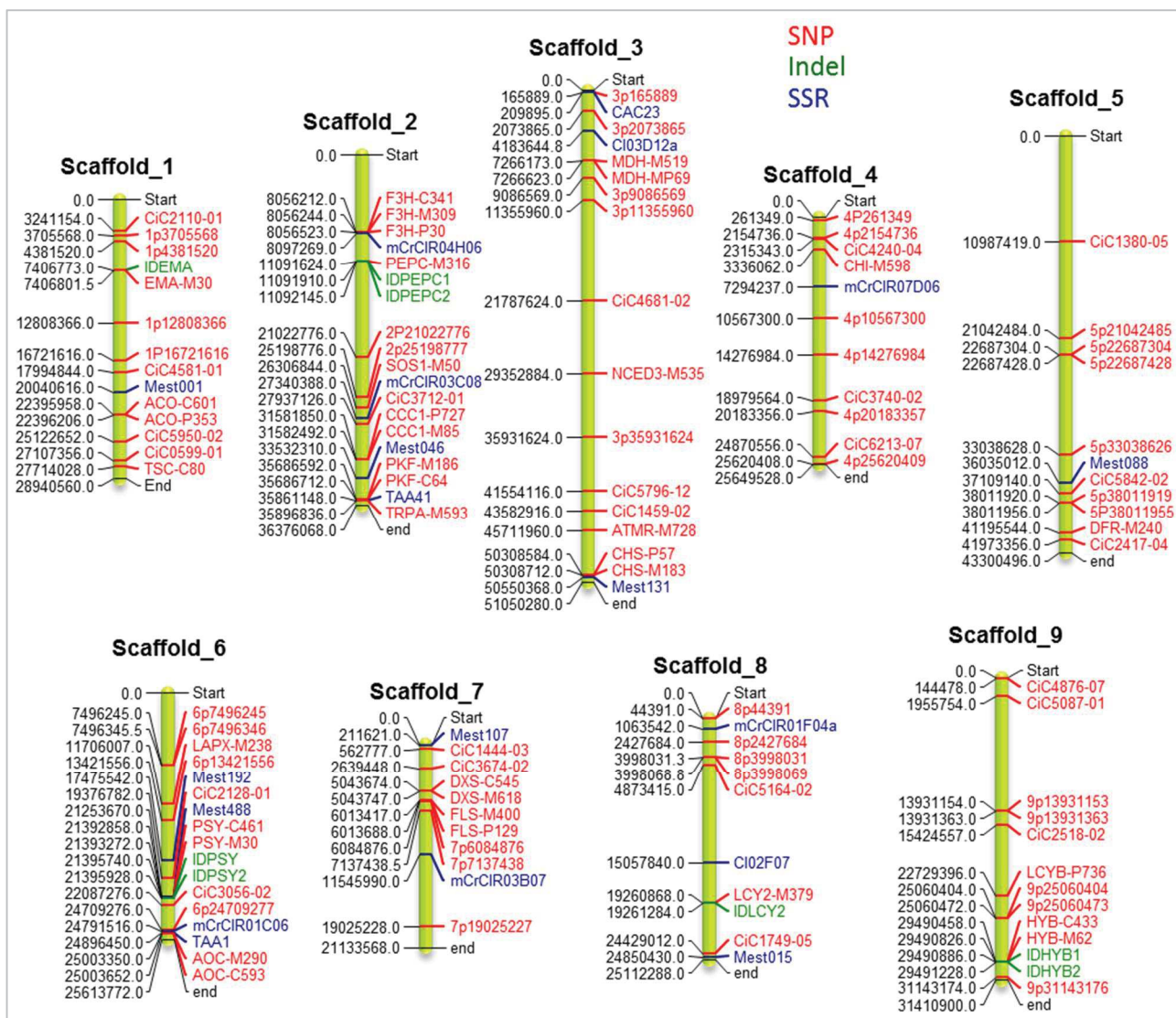


Figure S.III.1. Markers distribution over the nine chromosomes





Figure S.III.2a. Mytotype Close up

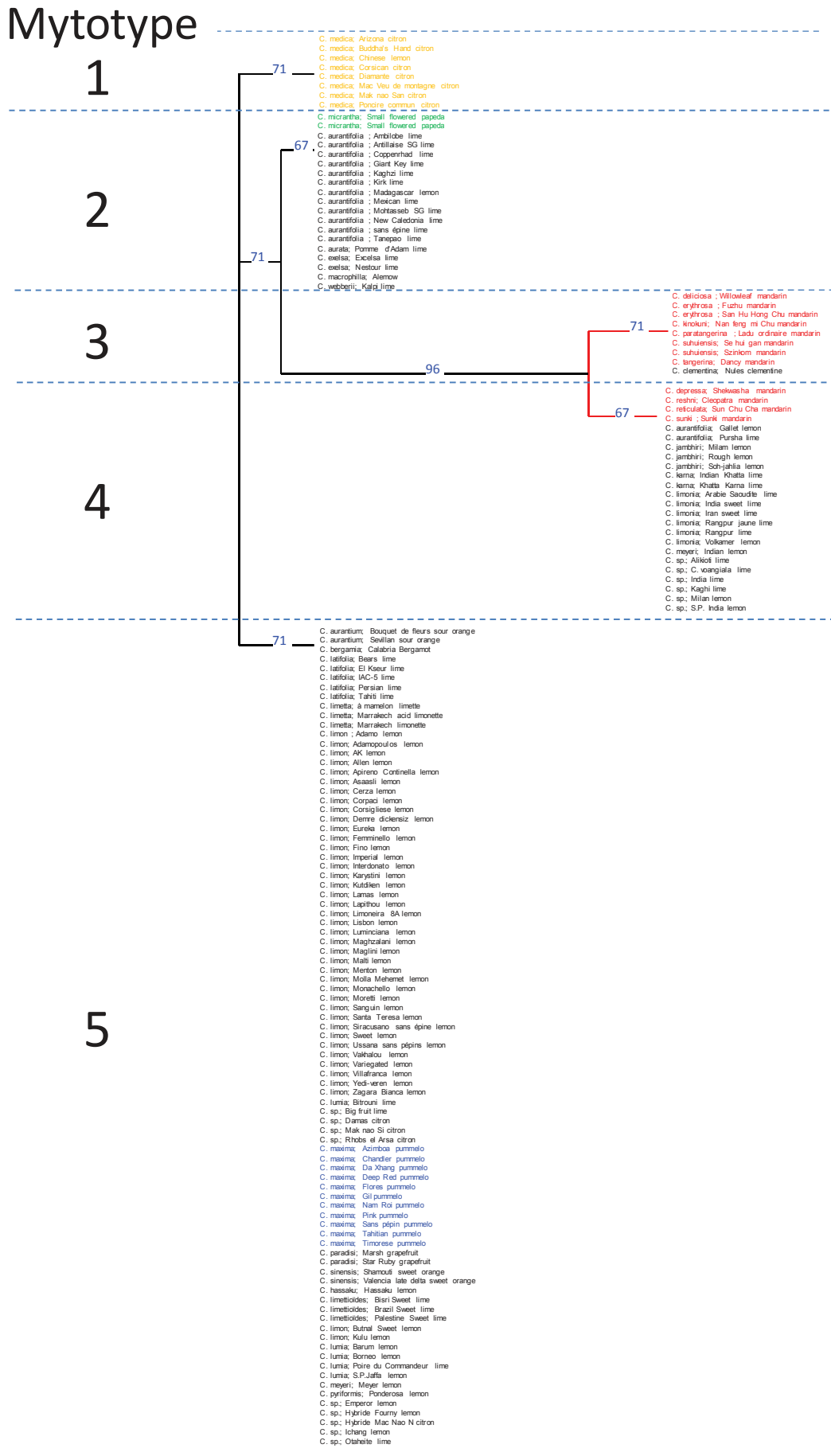


Figure S.III.2b. Chlorotype Close up

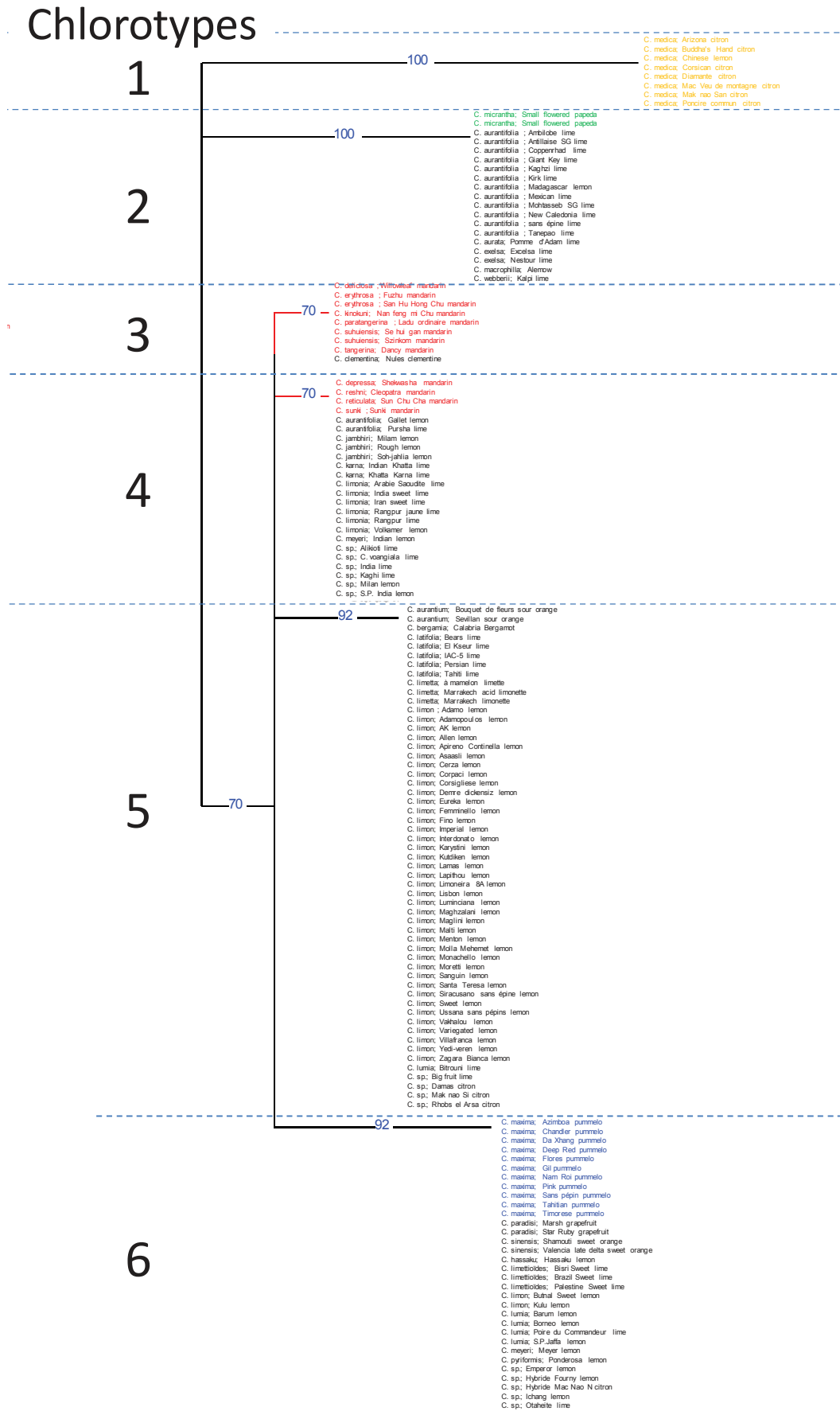


Table S.III.3. amplicon fragment size patterns of the 8 cytoplasmic markers for the six identified cytoplasmic types

Cytoplasmic type	Mitotype	nad2	nad7	rrn5	Chlorotype	ccpm2	ccmp5	ntcp7	ccmp6	ntcp9
1	1	242	160	261	1	213	99	200	127	266
2	2	252	160	261	2	203	94	189	136	269
3	3	244	152	256	3	197	94	182	129	259
4	4	252	152	256	4	197	94	182	129	262
5	5	261	160	261	5	197	84	182	129	269
6					197	94	182	131	271	

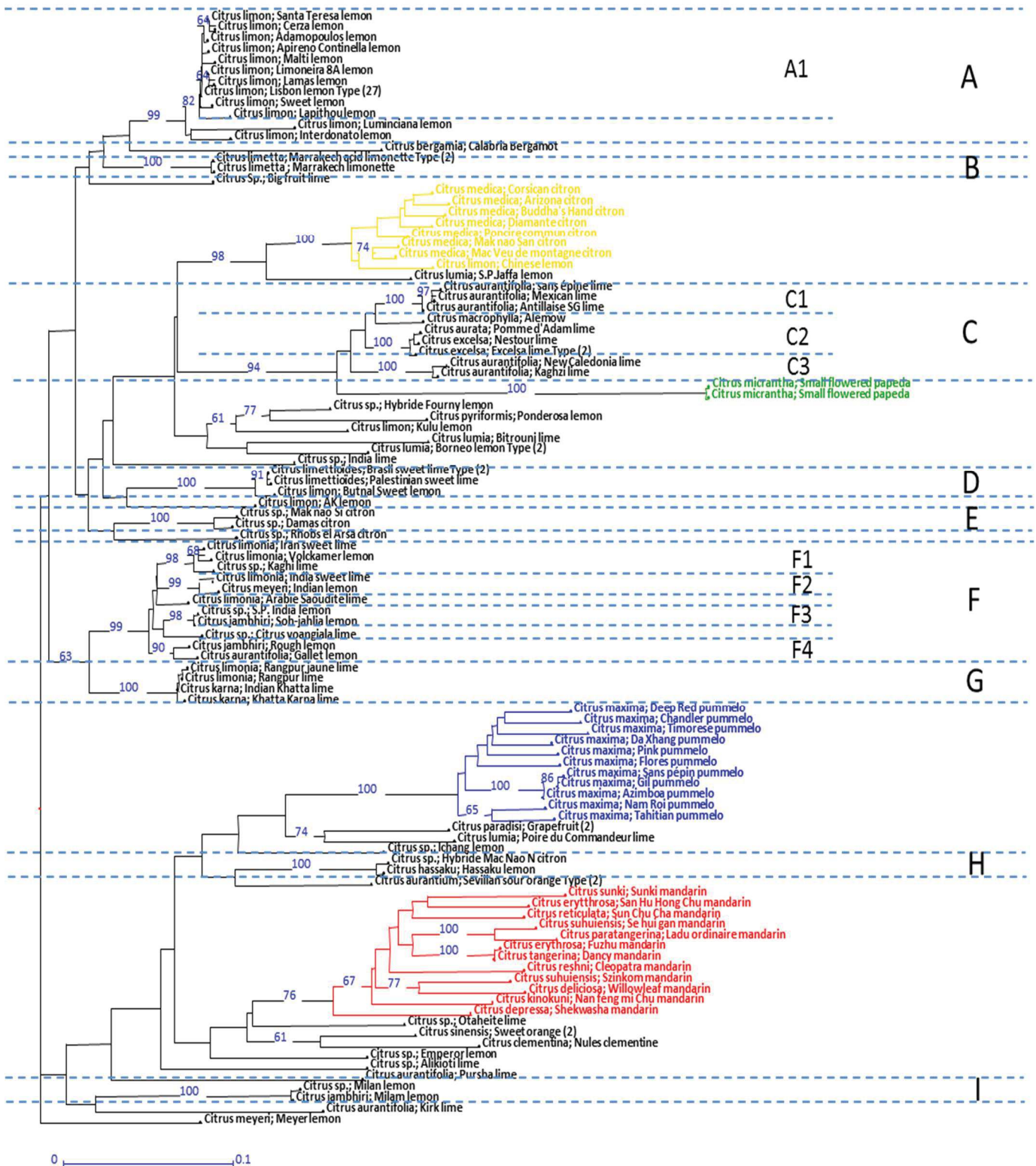


Figure S.III.3. NJ performed on the 103 MLGs with 123 molecular markers

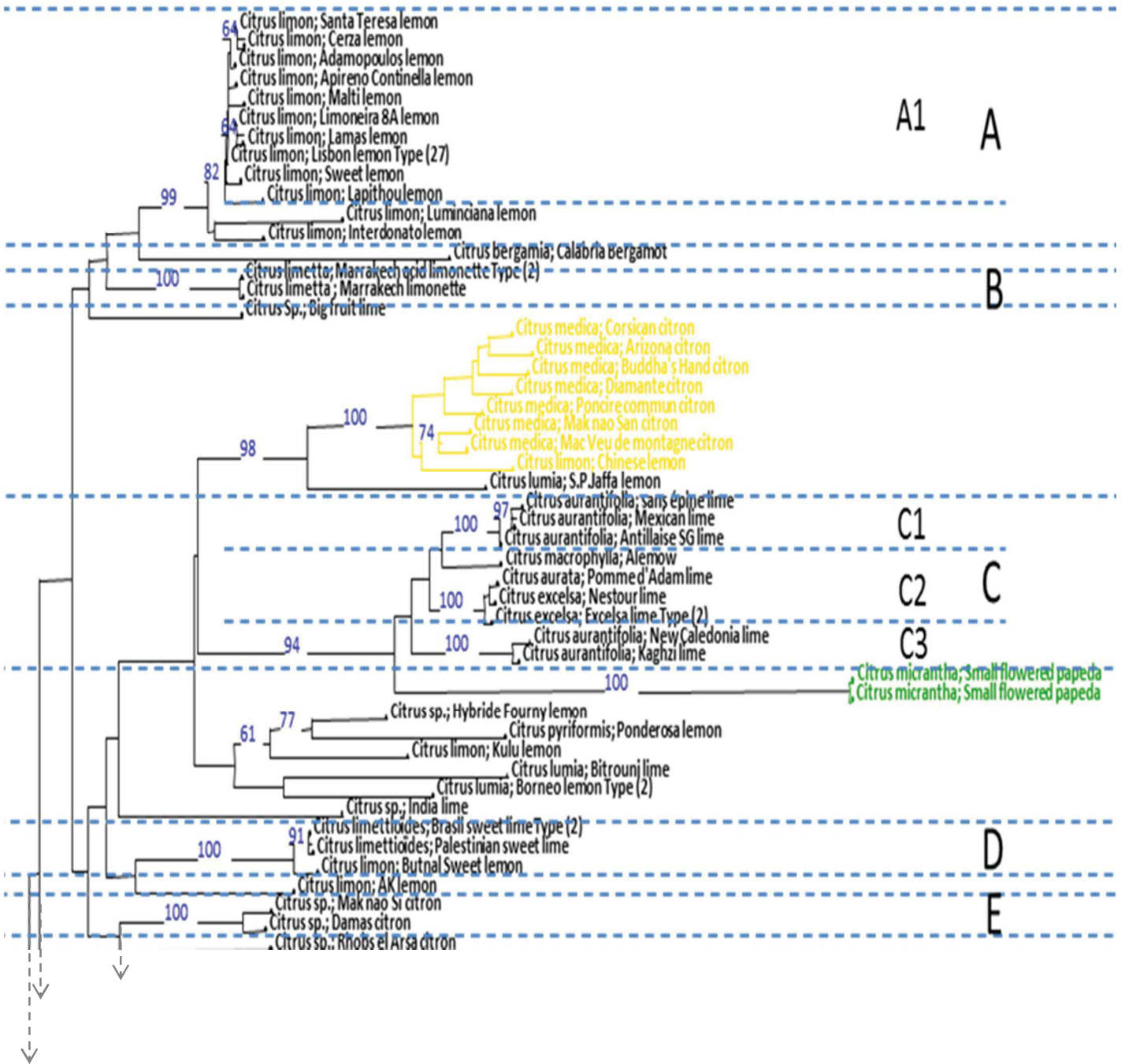


Figure S.III.3a. Close up 1

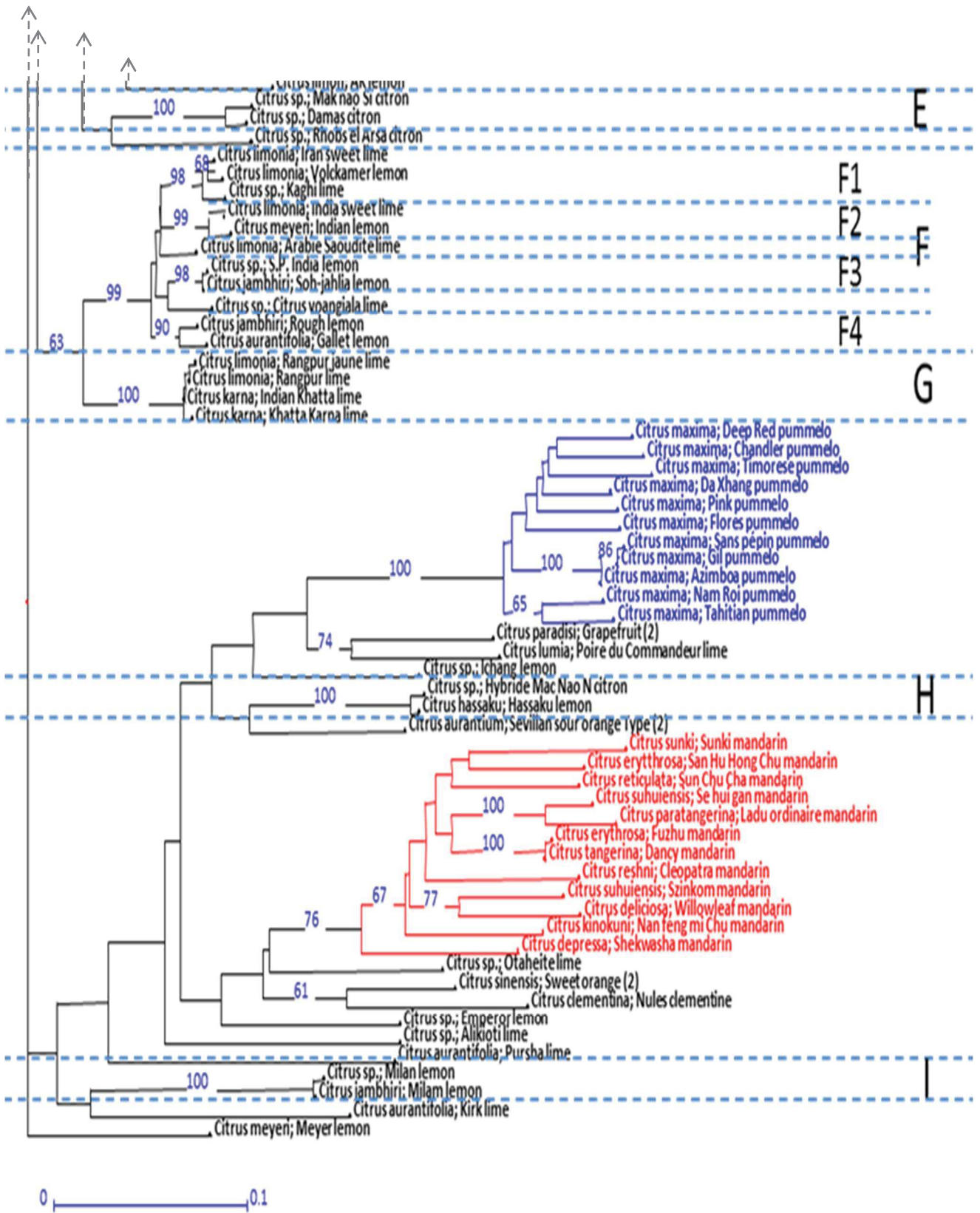


Figure S.III.3b. Close up 2

Table S.III.4. Estimation of the best number of subpopulation in structure analysis

Population structure was inferred using the Structure program version 2.3.4 (Pritchard Lab, 2014). No *a priori* population structure was defined. The linkage model option was used, with allele frequencies correlated and computed probability of the data for K estimating. Analyses were made with K value (number of subpopulations) varying from 1–10. The statistics used to select the correct K value were as in Evanno *et al.* (2005): the mean likelihood,  $L(K)$ ; the mean difference between successive likelihood values of K,  $L'(K)$ ; the absolute value of this difference,  $|L''(K)|$ ; and  $\Delta K$ , which is the mean of the absolute values of  $L''(K)$  divided by the standard deviation of  $L(K)$ . The likelihood distribution  $L(K)$  and  $\Delta K$  were the main values used to choose the optimal K value of the population. Analysis of delta K indicated that optimal results were obtained with  $K = 4$  [Additional file 5]. Minimal variability in estimated frequencies was observed between the ten runs for  $K = 4$ .

	MEDIA	DESVEST	$L'(K)$	$ L''(K) $	$\Delta K$
K=1	-7729.13	7.74913			
K=2	-6427.08	48.8093	1302.05	322.15	6.60018182
K=3	-5447.18	185.105	979.9	286.32	1.5468003
K=4	-4753.6	6.46168	693.58	571.02	88.3701808
K=5	-4631.04	175.014	122.56	39.27	0.22438233
K=6	-4547.75	10.3861	83.29	94.24	9.07364081
K=7	-4558.7	24.4782	-10.95	3.51	0.1433929
K=8	-4566.14	19.4367	-7.44	18.8	0.96724292
K=9	-4554.78	16.8503	11.36	32.84	1.94892882
K=10	-4576.26	55.8905	-21.48	21.48	0.38432292

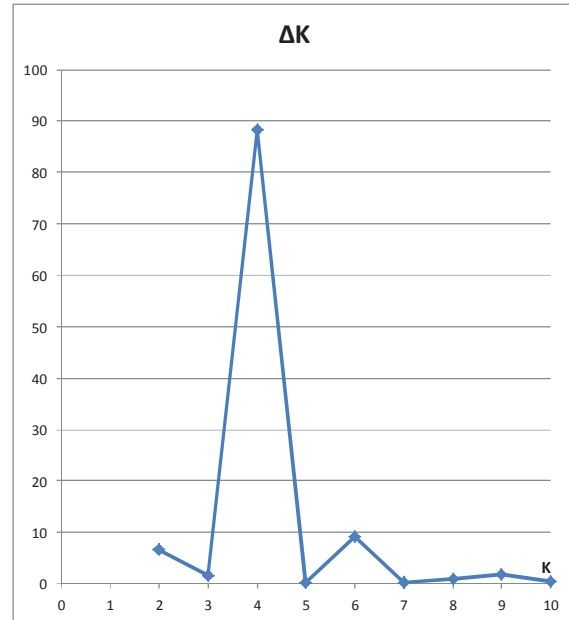


Table S.III.5. Means and confidence interval of the contribution of the four basic taxa (from 10 permuted and aligned independent Structure run cluster outputs)

Varieties	Missing data Missing Data (%)	Citron	Mandarin	C. micrantha	pummelo
Calabria Bergamot	2	0.300±0.009	0.271±0.008	0.009±0.002	0.418±0.009
Sevillan sour orange Type (2)	1	0.008±0.001	0.482±0.004	0.008±0.000	0.501±0.005
Arizona citron	1	0.977±0.001	0.007±0.000	0.007±0.000	0.007±0.000
Corsican citron	0	0.976±0.001	0.007±0.000	0.007±0.000	0.007±0.000
Damas citron	2	0.424±0.011	0.249±0.005	0.008±0.001	0.317±0.007
Diamante citron	0	0.974±0.002	0.008±0.000	0.008±0.001	0.007±0.000
Mac Veu de montagne citron	1	0.951±0.001	0.014±0.001	0.008±0.000	0.025±0.001
Buddha's Hand citron	1	0.978±0.000	0.007±0.000	0.007±0.000	0.006±0.000
Mak nao San citron	4	0.949±0.003	0.016±0.001	0.009±0.001	0.024±0.001
Mak nao Si citron	5	0.518±0.006	0.143±0.011	0.007±0.001	0.330±0.013
Poncire commun citron	5	0.949±0.002	0.01±0.000	0.020±0.001	0.020±0.001
Rhobs el Arsa citron	2	0.511±0.010	0.295±0.010	0.015±0.002	0.177±0.008
Adamopoulos lemon	0	0.476±0.012	0.316±0.007	0.007±0.001	0.2±0.008
AK lemon	1	0.498±0.011	0.347±0.012	0.008±0.001	0.146±0.006
Apireno Continella lemon	4	0.459±0.011	0.347±0.012	0.007±0.001	0.185±0.008
Butnal Sweet lemon	4	0.483±0.011	0.347±0.010	0.008±0.001	0.160±0.006
Cerza lemon	1	0.481±0.015	0.315±0.014	0.007±0.001	0.195±0.008
Chinese lemon	4	0.950±0.001	0.015±0.000	0.010±0.000	0.023±0.000
Borneo lemon Type (2)	0	0.368±0.006	0.009±0.001	0.121±0.010	0.500±0.014
Emperor lemon	1	0.04±0.006	0.520±0.005	0.090±0.008	0.348±0.009
Sweet lemon	8	0.486±0.007	0.320±0.010	0.008±0.002	0.184±0.011
Gallet lemon	2	0.498±0.005	0.484±0.005	0.008±0.001	0.008±0.001
Hassaku lemon	2	0.031±0.002	0.450±0.004	0.012±0.001	0.505±0.005
Ichang lemon	4	0.081±0.004	0.345±0.005	0.025±0.003	0.548±0.006
Indian lemon	2	0.454±0.008	0.525±0.010	0.009±0.001	0.011±0.001
Interdonato lemon	4	0.544±0.009	0.265±0.007	0.006±0.001	0.183±0.007
Lisbon lemon Type (27)	1	0.486±0.012	0.310±0.011	0.007±0.001	0.196±0.008
Kulu lemon	1	0.559±0.008	0.041±0.002	0.007±0.001	0.391±0.006
Lamas lemon	0	0.486±0.006	0.314±0.012	0.008±0.001	0.190±0.010
Lapithou lemon	2	0.510±0.008	0.285±0.008	0.008±0.001	0.196±0.009
Limoneira 8A lemon	0	0.478±0.010	0.308±0.011	0.008±0.002	0.204±0.008
Luminciana lemon	2	0.565±0.012	0.238±0.008	0.010±0.000	0.186±0.007
Malti lemon	0	0.484±0.010	0.314±0.008	0.007±0.001	0.194±0.007
Meyer lemon	6	0.436±0.013	0.338±0.011	0.008±0.002	0.215±0.006
Milam lemon	2	0.349±0.008	0.454±0.006	0.009±0.001	0.186±0.006
Milan lemon	1	0.364±0.009	0.448±0.008	0.006±0.001	0.181±0.008
Ponderosa lemon	1	0.588±0.008	0.008±0.001	0.011±0.001	0.391±0.008
S.P. Jaffa lemon	0	0.641±0.007	0.008±0.001	0.008±0.001	0.342±0.006
Santa Teresa lemon	0	0.489±0.009	0.307±0.009	0.008±0.001	0.193±0.010
Soh-jahlia lemon	0	0.498±0.004	0.484±0.004	0.009±0.001	0.008±0.000
Alemow	0	0.475±0.007	0.008±0.001	0.507±0.007	0.008±0.000
Small flowered papeda	0	0.007±0.000	0.007±0.000	0.977±0.001	0.007±0.000
Small flowered papeda	0	0.007±0.000	0.006±0.000	0.978±0.001	0.007±0.000
Citrus voangiala lime	0	0.498±0.010	0.485±0.010	0.006±0.000	0.008±0.001
Volckamer lemon	2	0.468±0.005	0.513±0.005	0.009±0.001	0.008±0.001
Nules clementine	0	0.007±0.000	0.656±0.005	0.009±0.001	0.327±0.004
Hybride Fourny lemon	2	0.493±0.011	0.009±0.001	0.008±0.001	0.489±0.010
Hybride Mac Nao N citron	2	0.023±0.002	0.436±0.004	0.064±0.004	0.475±0.008
India lime	1	0.380±0.007	0.262±0.007	0.072±0.013	0.283±0.013
Big fruit lime	1	0.449±0.008	0.264±0.009	0.008±0.002	0.278±0.008
Alikioti lime	2	0.095±0.005	0.401±0.009	0.010±0.001	0.493±0.010
Antillaise SG lime	4	0.485±0.008	0.008±0.001	0.498±0.008	0.007±0.000
Arabie Saoudite lime	2	0.487±0.005	0.494±0.005	0.008±0.000	0.008±0.000
Bitrouni lime	2	0.313±0.005	0.009±0.000	0.012±0.001	0.664±0.005
Palestinian sweet lime	2	0.462±0.012	0.364±0.014	0.008±0.001	0.165±0.006
Iran sweet lime	0	0.492±0.006	0.491±0.006	0.007±0.001	0.008±0.001
Brasil sweet lime Type (2)	2	0.476±0.010	0.350±0.009	0.008±0.001	0.164±0.005
India sweet lime	2	0.504±0.007	0.477±0.007	0.009±0.000	0.008±0.001
Excelsa lime Type (2)	2	0.489±0.008	0.009±0.001	0.493±0.007	0.008±0.001
Kaghi lime	2	0.493±0.006	0.487±0.006	0.009±0.001	0.009±0.001
Kaghzi lime	0	0.474±0.005	0.007±0.001	0.509±0.006	0.008±0.001
Indian Khatta lime	2	0.489±0.007	0.490±0.008	0.009±0.000	0.010±0.001
Khatta Karna lime	2	0.492±0.008	0.488±0.007	0.009±0.002	0.009±0.001
Kirk lime	6	0.276±0.012	0.352±0.015	0.138±0.007	0.232±0.010
Mexican lime	1	0.484±0.003	0.008±0.000	0.498±0.003	0.009±0.000



Nestour lime	0	0.519±0.005	0.009±0.001	0.463±0.006	0.008±0.001
New Caledonia lime	0	0.468±0.005	0.006±0.000	0.517±0.005	0.007±0.001
Otaheite lime	4	0.011±0.001	0.654±0.006	0.008±0.001	0.326±0.007
Pursha lime	1	0.311±0.005	0.653±0.006	0.006±0.000	0.027±0.003
Rangpur lime	0	0.490±0.005	0.489±0.006	0.009±0.001	0.010±0.001
Rangpur jaune lime	1	0.499±0.005	0.478±0.005	0.010±0.001	0.010±0.002
sans épine lime	1	0.477±0.009	0.008±0.001	0.504±0.007	0.008±0.002
Marrakech limonette	2	0.499±0.007	0.366±0.010	0.009±0.001	0.123±0.007
Marrakech acid limonette Type (2)	2	0.521±0.012	0.353±0.008	0.007±0.001	0.118±0.008
Cleopatra mandarin	1	0.008±0.000	0.977±0.001	0.007±0.000	0.007±0.000
Willowleaf mandarin	0	0.008±0.000	0.868±0.002	0.008±0.000	0.115±0.003
Dancy mandarin	0	0.007±0.000	0.879±0.006	0.008±0.001	0.104±0.006
Fuzhu mandarin	1	0.007±0.000	0.882±0.005	0.008±0.000	0.101±0.005
Ladu ordinaire mandarin	1	0.007±0.000	0.935±0.003	0.008±0.001	0.048±0.002
Nan feng mi Chu mandarin	1	0.009±0.000	0.889±0.003	0.009±0.000	0.091±0.003
San Hu Hong Chu mandarin	1	0.007±0.000	0.971±0.001	0.008±0.000	0.012±0.001
Se hui gan mandarin	1	0.008±0.000	0.935±0.002	0.007±0.000	0.049±0.002
Shekwasha mandarin	1	0.038±0.002	0.92±0.002	0.010±0.001	0.030±0.001
Sun Chu Cha mandarin	0	0.007±0.000	0.975±0.001	0.007±0.000	0.009±0.000
Sunki mandarin	1	0.007±0.000	0.971±0.001	0.008±0.000	0.012±0.001
Szinkom mandarin	0	0.006±0.000	0.934±0.002	0.007±0.000	0.051±0.002
Sweet orange (2)	0	0.008±0.001	0.526±0.008	0.008±0.001	0.456±0.008
Azimboa pummelo	1	0.006±0.000	0.008±0.000	0.007±0.000	0.976±0.000
Chandler pummelo	5	0.007±0.000	0.009±0.000	0.008±0.001	0.974±0.002
Da Xhang pummelo	1	0.007±0.000	0.008±0.001	0.008±0.000	0.975±0.001
Flores pummelo	2	0.008±0.000	0.008±0.000	0.008±0.000	0.975±0.001
Tahitian pummelo	4	0.007±0.000	0.007±0.001	0.009±0.000	0.975±0.002
Timorese pummelo	1	0.007±0.000	0.008±0.000	0.007±0.000	0.976±0.001
Deep Red pummelo	0	0.007±0.000	0.007±0.000	0.009±0.001	0.975±0.001
Gil pummelo	1	0.007±0.000	0.008±0.000	0.008±0.000	0.977±0.001
Nam Roi pummelo	5	0.008±0.001	0.009±0.000	0.008±0.000	0.974±0.002
Pink pummelo	0	0.007±0.000	0.008±0.000	0.007±0.000	0.976±0.001
Sans pépin pummelo	1	0.008±0.000	0.008±0.000	0.008±0.000	0.974±0.001
Poire du Commandeur lime	2	0.007±0.001	0.348±0.007	0.007±0.001	0.636±0.009
Grapefruit (2)	1	0.006±0.001	0.401±0.006	0.007±0.000	0.583±0.005
Pomme d'Adam lime	1	0.525±0.005	0.008±0.001	0.457±0.006	0.008±0.000
Rough lemon	0	0.495±0.006	0.487±0.007	0.008±0.001	0.008±0.000
S.P. India lemon	0	0.498±0.006	0.485±0.006	0.008±0.000	0.007±0.001

Table S.III.6. congruency of molecular data (123 SNP, Indel, SSR markers) with several hypothesis of lime and lemon origins

Variety name (Tanaka classification)	Phylogenetic model Swingle and Reece classification	Female parent	Male parent	N.C.	CT
<b>Diploid limes and lemons</b>					
Lisbon lemon (MLGCit)	C. aurantium x C. medica (C. maximaxC. reticulata) x (C. medica)	Sour orange	Citron (pop)	99.2%	Ok
Marrakechacid limonetteMLG	C. aurantium x C. medica (C. maximaxC. reticulata) x (C. medica) C. sinensis x C. medica (C. maximaxC. reticulata) x (C. medica)	Sour orange Sweet orange	Citron (pop) Citron (pop)	98.3% 83.5%	Ok No
Volkamer lemon ( <i>C. limonia</i> )	C. reticulata x C. medica	Cleopatra_mandarin	Citron (pop)	86.7%	Ok
		Willowleaf_mandarin	Citron (pop)	90.8%	No
		Dancy_mandarin	Citron (pop)	94.2%	No
		Fuzhu_mandarin	Citron (pop)	93.3%	No
		Ladu_ordinaire_mandarin	Citron (pop)	92.5%	No
		Nan_feng_mi_Chu_mandarin	Citron (pop)	87.5%	No
		San_Hu_Hong_Chu_mandarin	Citron (pop)	87.5%	No
		Se_hui_gan_mandarin	Citron (pop)	90.8%	No
		Shekwasha_mandarin	Citron (pop)	90.8%	Ok
		Sun_Chu_Cha_mandarin	Citron (pop)	94.2%	Ok
		Sunki_mandarin	Citron (pop)	85.8%	Ok
Szinkom_mandarin	Citron (pop)	91.7%	No		
All Mandarin (Pop)	Citron (pop)	97.5%			
Rangpur lime ( <i>C. limonia</i> )	C. reticulata x C. medica	Cleopatra_mandarin	Citron (pop)	85.2%	Ok
		Willowleaf_mandarin	Citron (pop)	87.7%	No
		Dancy_mandarin	Citron (pop)	89.3%	No
		Fuzhu_mandarin	Citron (pop)	88.5%	No
		Ladu_ordinaire_mandarin	Citron (pop)	85.2%	No
		Nan_feng_mi_Chu_mandarin	Citron (pop)	87.7%	No
		San_Hu_Hong_Chu_mandarin	Citron (pop)	84.4%	No
		Se_hui_gan_mandarin	Citron (pop)	85.2%	No
		Shekwasha_mandarin	Citron (pop)	86.1%	Ok
		Sun_Chu_Cha_mandarin	Citron (pop)	89.3%	Ok
		Sunki_mandarin	Citron (pop)	85.2%	Ok
Szinkom_mandarin	Citron (pop)	86.8%	No		
All Mandarin (Pop)	Citron (pop)	96.7%	OK acid Mandarin CT		
Rough lemon ( <i>C. jambhiri</i> )	C. reticulata x C. medica	Cleopatra_mandarin	Citron (pop)	86.2%	Ok
		Willowleaf_mandarin	Citron (pop)	88.6%	No
		Dancy_mandarin	Citron (pop)	95.1%	No
		Fuzhu_mandarin	Citron (pop)	94.3%	No
		Ladu_ordinaire_mandarin	Citron (pop)	89.4%	No
		Nan_feng_mi_Chu_mandarin	Citron (pop)	87.8%	No
		San_Hu_Hong_Chu_mandarin	Citron (pop)	87.0%	No
		Se_hui_gan_mandarin	Citron (pop)	87.0%	No
		Shekwasha_mandarin	Citron (pop)	90.2%	Ok
		Sun_Chu_Cha_mandarin	Citron (pop)	95.1%	Ok
		Sunki_mandarin	Citron (pop)	86.2%	Ok
Szinkom_mandarin	Citron (pop)	91.1%	No		
All Mandarin (Pop)	Citron (pop)	97.6%			
Mexican lime ( <i>C. aurantifolia</i> )	C. micrantha x C. medica	C. micrantha	Citron (pop)	95.0%	Ok
Excelsa lime ( <i>C. excelsa</i> )	C. micrantha x C. medica	C. micrantha	Citron (pop)	88.2%	Ok
Alemow ( <i>C. macrophylla</i> )	C. micrantha x C. medica	C. micrantha	Citron (pop)	89.3%	Ok
Variety name (Tanaka classification)	Phylogenetic model Swingle and Reece classification	Female parent	Male parent	N.C.	CT

	C. micrantha x C. medica	C. micrantha	Citron (pop)	86.1%	Ok
New caledonia lime (C. aurantifolia)	C. aurantifolia x C. medica	Mexican lime	Citron (pop)	93.4%	Ok
	(C. micrantha x C. medica) x C. medica				
	C. aurantifolia x C. aurantifolia (C. micrantha x C. medica) x (C. micrantha x C. medica)	Mexican lime	Mexican lime	91.8%	
Kirk lime (C. aurantifolia)	C. aurantifolia x C. aurantium (C. micrantha x C. medica) x (C. maxima x C. reticulata)	Mexican lime	Sour orange	82.8%	OK
	C. aurantifolia x C. sinensis (C. micrantha x C. medica) x (C. maxima x C. reticulata)	Mexican lime	Sweet orange	89.7%	OK
	C. aurantifolia x (C. maxima x C. reticulata)	Mexican lime	Mand/pamp (pop)	93.1%	OK
	(C. micrantha x C. medica) x (C. maxima x C. reticulata)				
Calabria Bergamot orange (C. bergamia)	C. limon x C. aurantium	Lisbon lemon	Sour orange	96.7%	OK
	C. limon x C. medica	Lisbon lemon	Citron (Pop)	74.2%	OK
	C. limon x C. limon	Lisbon lemon	Lisbon lemon	85.0%	OK
	C. limon x C. aurantium	Marrakechacid limonetteMLG	Sour orange	90.0%	OK
	C. limon x C. medica	Marrakechacid limonetteMLG	Citron (Pop)	72.5%	OK
	C. limon x C. limon	Marrakechacid limonetteMLG	Lisbon lemon	87.5%	OK
	C. aurantium x C. aurantifolia	Sour orange	Palestinian Sweet lime	88.3%	OK
Meyer lemon (C. meyeri)	C. sinensis x C. medica	Sweet orange	Citron (pop)	86.4%	Ok
	(C. maxima x C. reticulata) x C. medica	Poire du commendeur	Citron (pop)	77.1%	Ok
		Hassaku lemon	Citron (pop)	83.1%	Ok
		Otaheite lime	Citron (pop)	84.7%	Ok
		Pummelo/Mandarin (pop)	Citron (pop)	95.8%	Ok
Palestine Sweet lime (C. limettioides)	C. sinensis x C. medica	Sweet orange	Citron (pop)	85.0%	Ok
	(C. maxima x C. reticulata) x C. medica	Poire du commendeur	Citron (pop)	80.0%	Ok
		Hassaku lemon	Citron (pop)	83.3%	Ok
		Otaheite lime	Citron (pop)	85.0%	Ok
		Pummelo/Mandarin (pop)	Citron (pop)	98.3%	Ok
Poire du Commendeur	C. maxima x C. medica	Pummelo (pop)	Mandarin (Pop)	94.2%	Ok
Fourny citron hybrid	C. maxima x C. medica	Pummelo (pop)	Citron (pop)	96.7%	Ok
<b>Tetraploid lime</b>					
Giant Key (C. aurantifolia)	doubled C. aurantifolia	Dobling of Mexican lime chromozome stock		100.0%	Ok
<b>Triploid limes</b>					
Bears lime (C. latifolia)		Lemon Lisbonne x gamete	Mexican lime 2x gamete	99.1%	Ok
El Kseur lime (C. latifolia)		Lemon Lisbonne x gamete	Mexican lime 2x gamete	99.1%	Ok
IAC-5 lime (C. latifolia)	C. limon x C. aurantifolia	Lemon Lisbonne x gamete	Mexican lime 2x gamete	99.1%	Ok
Persian lime (C. latifolia)	(C. maxima x C. reticulata) x (C. micrantha x C. medica)	Lemon Lisbonne x gamete	Mexican lime 2x gamete	99.1%	Ok
Tahiti lime (C. latifolia)		Lemon Lisbonne x gamete	Mexican lime 2x gamete	99.1%	Ok
Ambilobe lime (C. aurantifolia)		Mexican lime 2x gamete	Citron (pop) n gamete	96.3%	Ok
Coppenrhad lime (C. aurantifolia)		Mexican lime 2x gamete	Citron (pop) n gamete	97.4%	Ok
Madagascar lemon (C. aurantifolia)	C. aurantifolia x C. medica	Mexican lime 2x gamete	Citron (pop) n gamete	96.3%	Ok
Mohasseb lime (C. aurantifolia)	(C. micrantha x C. medica) x C. medica	Mexican lime 2x gamete	Citron (pop) n gamete	98.8%	Ok
Tanepao lime (C. aurantifolia)		Mexican lime 2x gamete	Citron (pop) n gamete	96.4%	Ok

# Annexe V

---

## Supplementary information Chapitre IV

## Eureka lemon

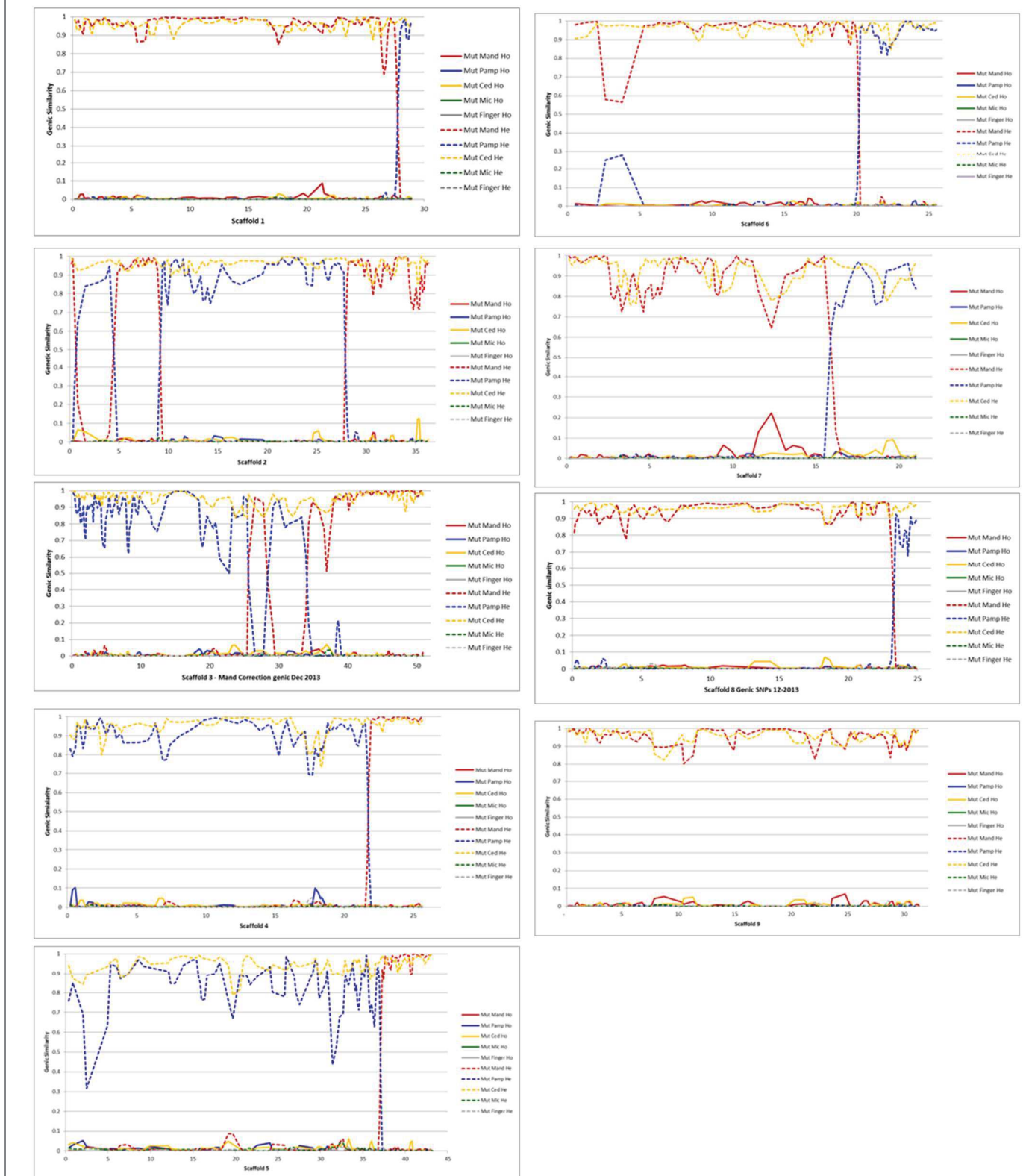


Figure S.IV.1a. 'Eureka' lemon phylogenomic structure of their 9 chromosomes by the pattern of homozygosity and heterozygosity frequency for the five sets of ancestral taxa diagnostic SNPs along each chromosome.

## Rough lemon



Figure S.IV.1b. 'Rough' lemon phylogenomic structure of their 9 chromosomes by the pattern of homozygosity and heterozygosity frequency for the five sets of ancestral taxa diagnostic SNPs along each chromosome.

## Volkamer lemon



Figure S.IV.1c. 'Volkamer' lemon phylogenomic structure of their 9 chromosomes by the pattern of homozygosity and heterozygosity frequency for the five sets of ancestral taxa diagnostic SNPs along each chromosome.

## Rangpur lime



Figure S.IV.1d. 'Rangpur' lime phylogenomic structure of their 9 chromosomes by the pattern of homozygosity and heterozygosity frequency for the five sets of ancestral taxa diagnostic SNPs along each chromosome.



## Mexican lime



Figure S.IV.1e. 'Mexican' lime phylogenomic structures of their 9 chromosomes by the pattern of homozygosity and heterozygosity frequency for the five sets of ancestral taxa diagnostic SNPs along each chromosome.

## Alemow



Figure S.IV.1f. 'Alemow' phylogenomic structures of their 9 chromosomes by the pattern of homozygosity and heterozygosity frequency for the five sets of ancestral taxa diagnostic SNPs along each chromosome.

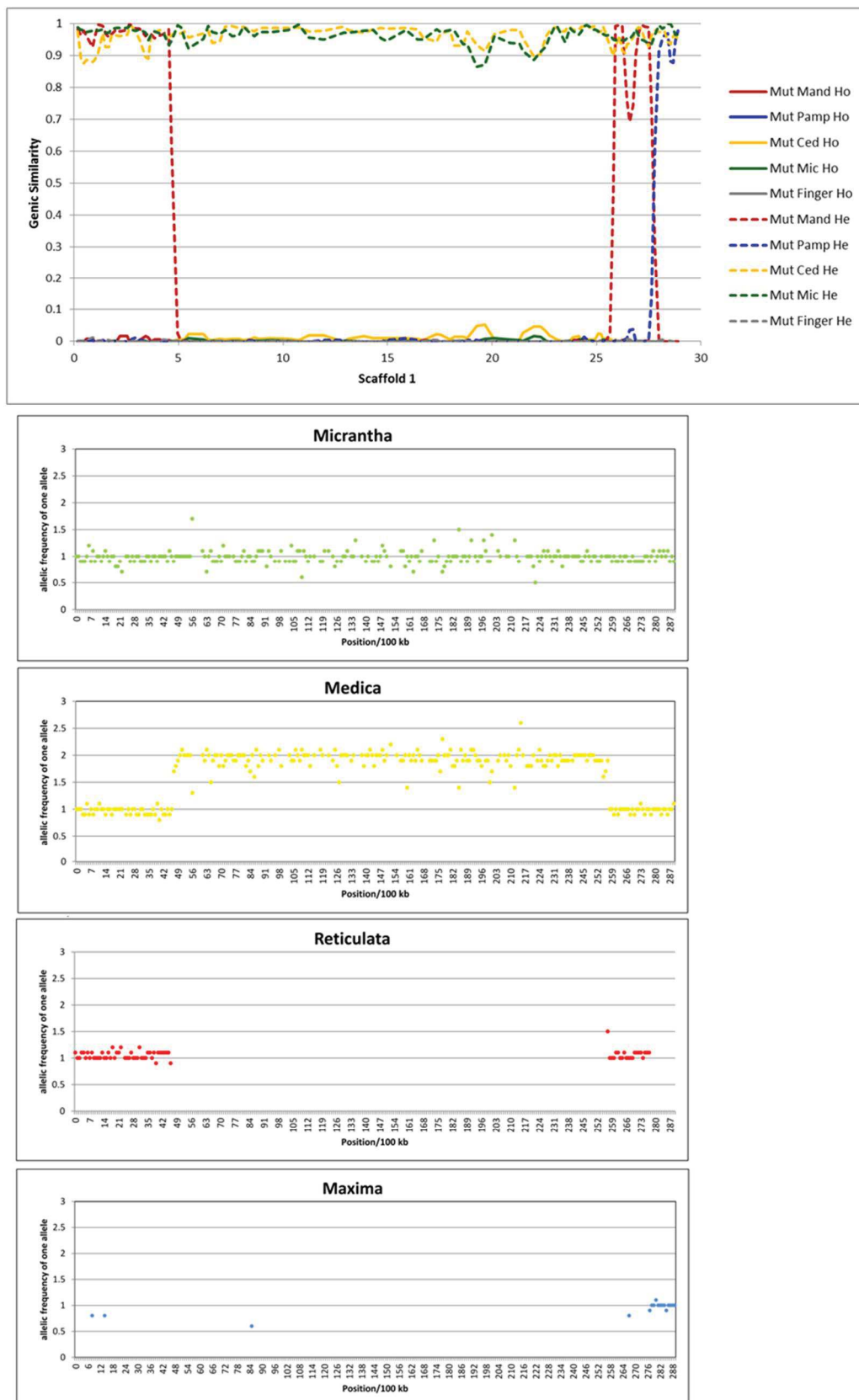


Figure S.IV.2a. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 1. Frequency of homozygosity and heterozygosity for the five sets of ancestral taxa diagnostic alleles and relative numbers of reads in the genomic regions in heterozygosity between two taxa (*C. medica* and *C. micrantha* for the concerned chromosomes).

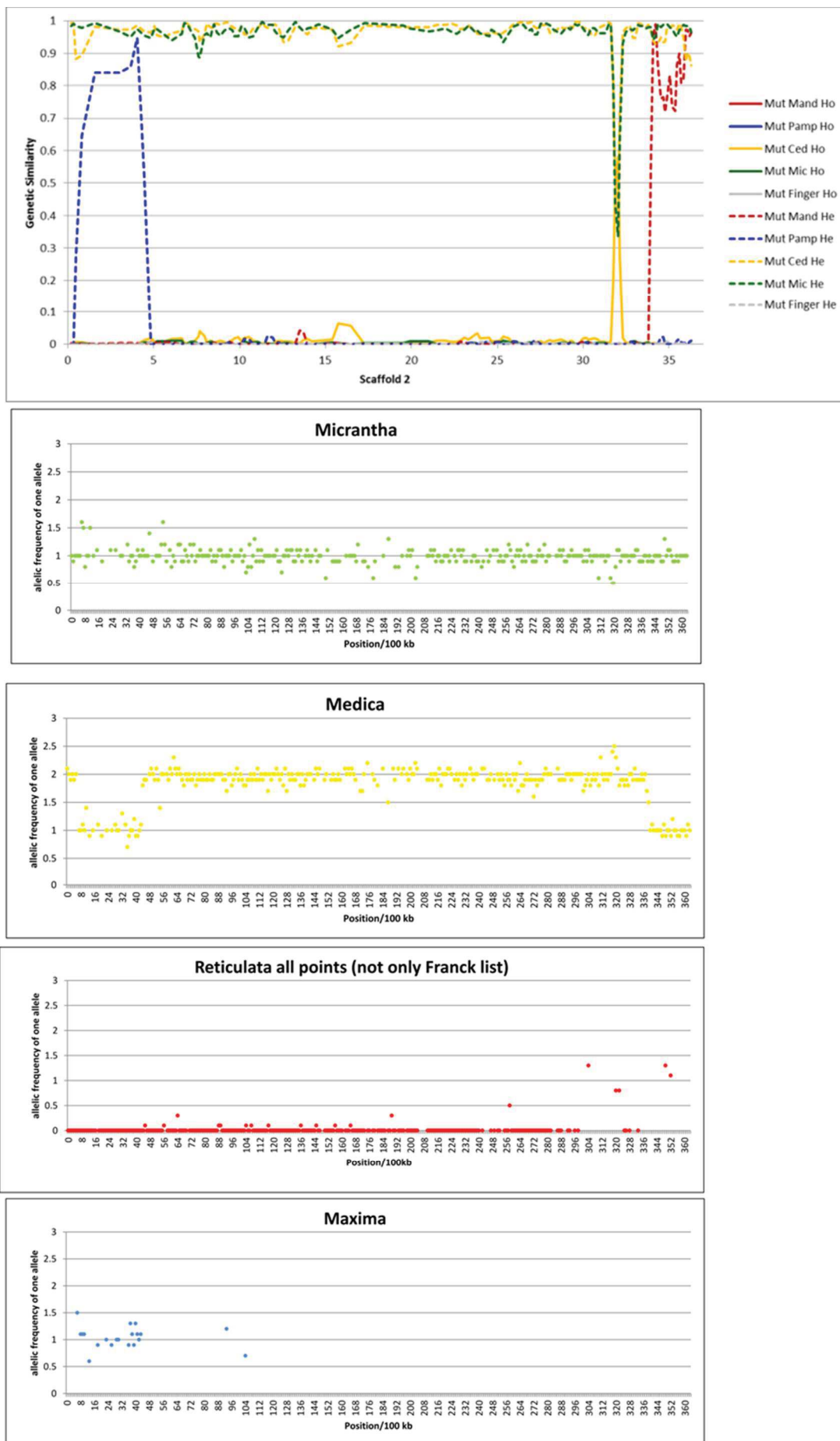


Figure S.IV.2b. Phylogenomic structure of the triploid 'Bears' lime for chromosome 2.

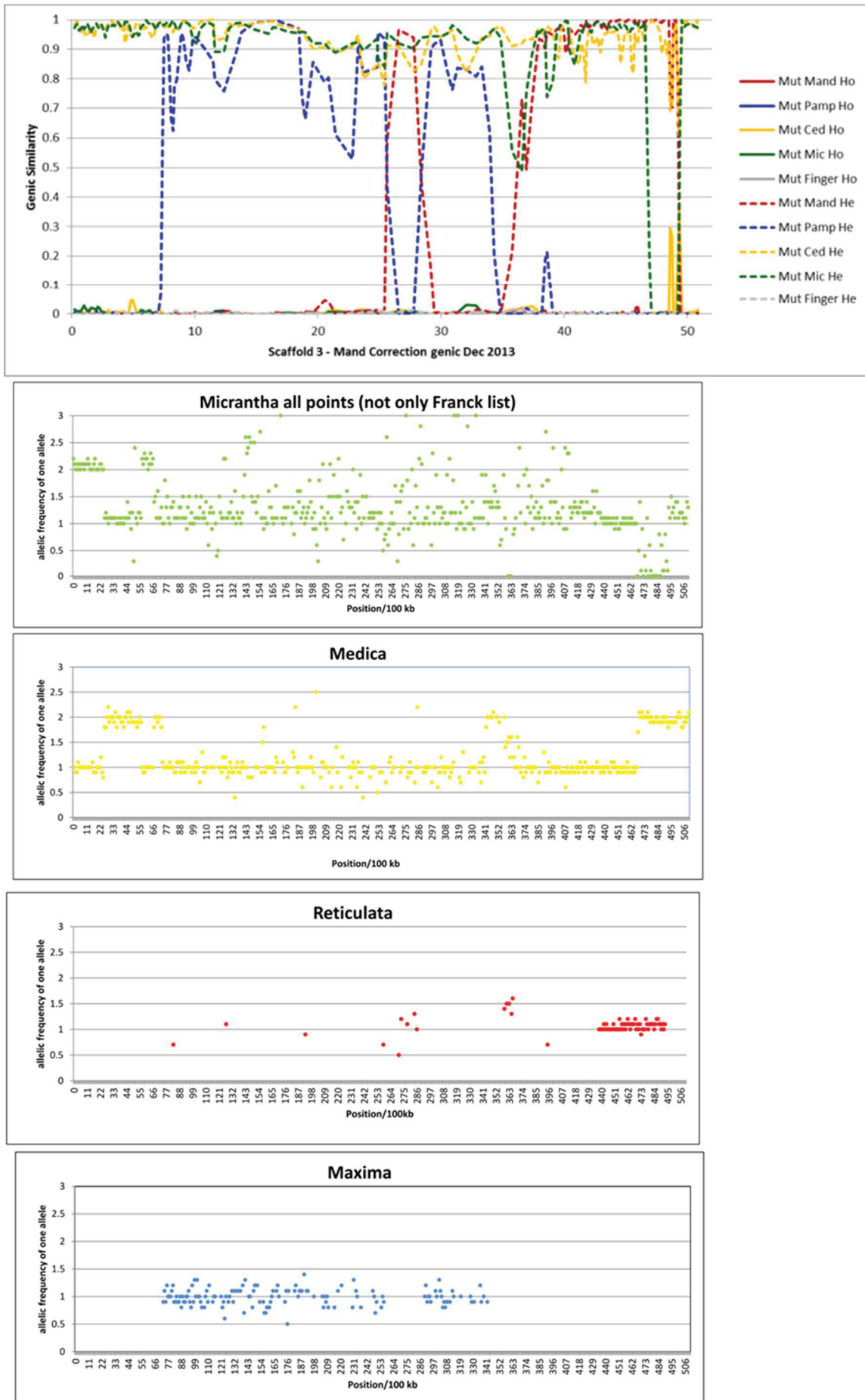


Figure S.IV.2c. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 3.

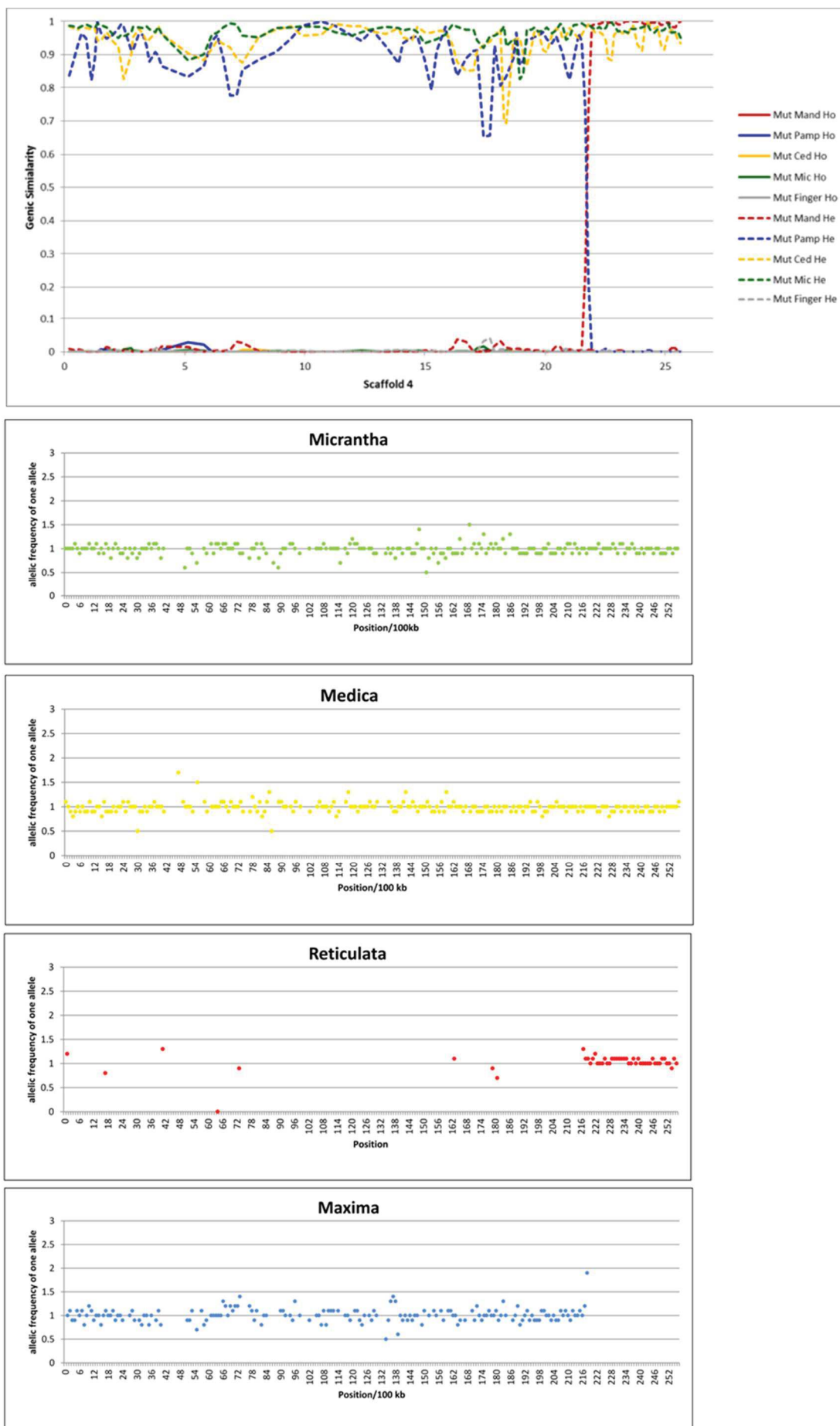


Figure S.IV.2d. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 4.

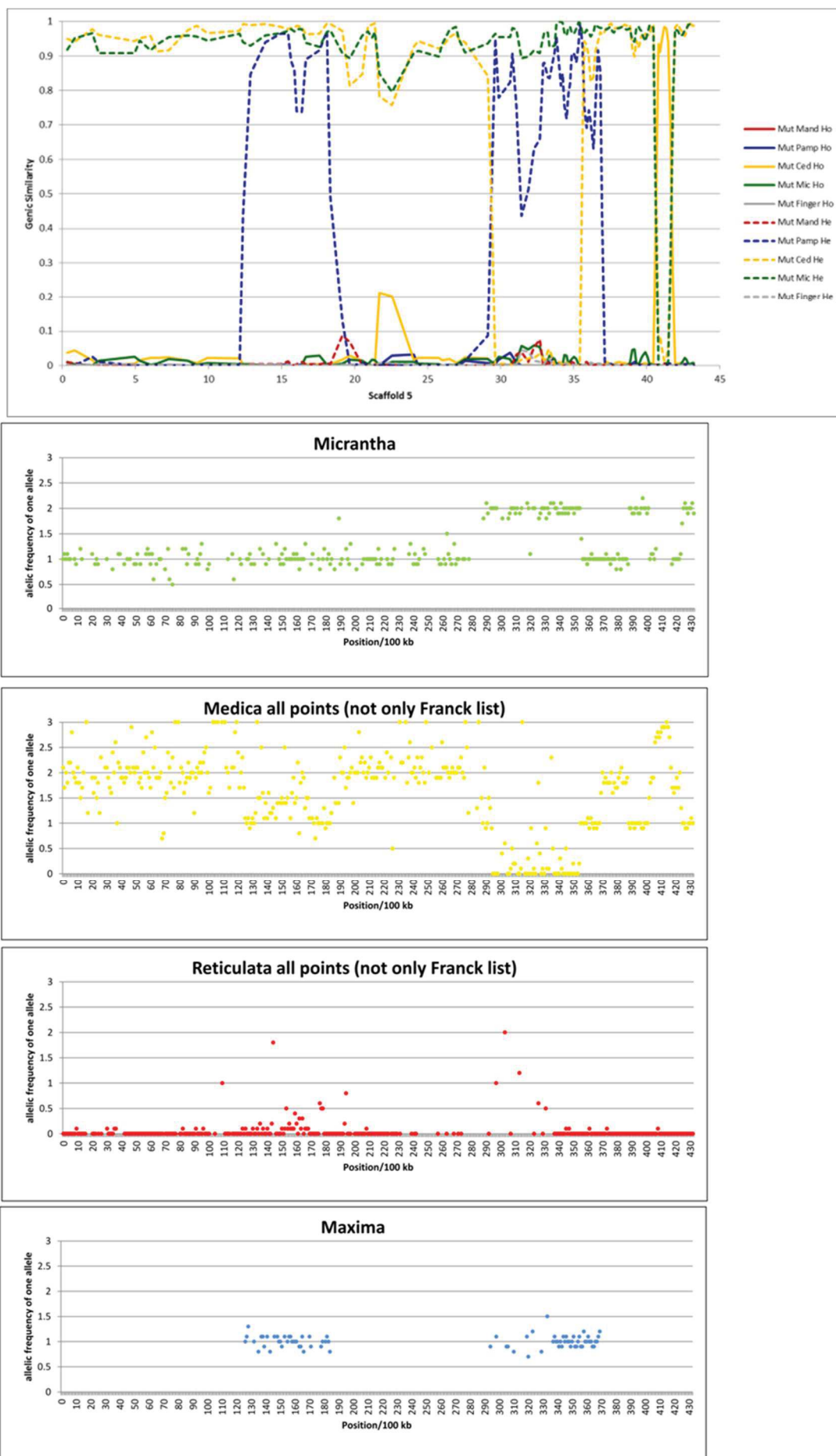


Figure S.IV.2e. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 5.

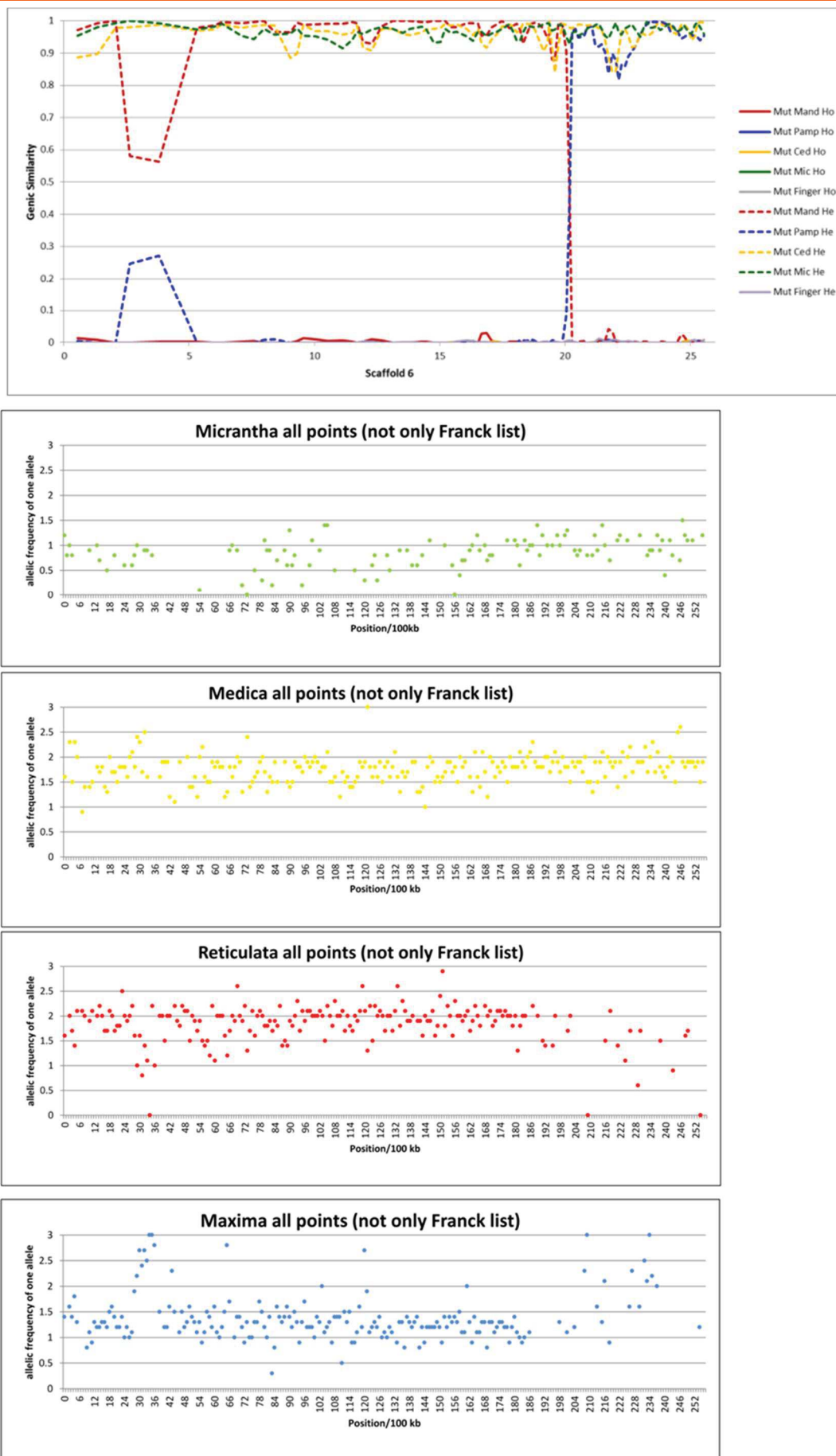


Figure S.IV.2f. Phylogenomic structure of the triploid 'Bears' lime for chromosome 6.



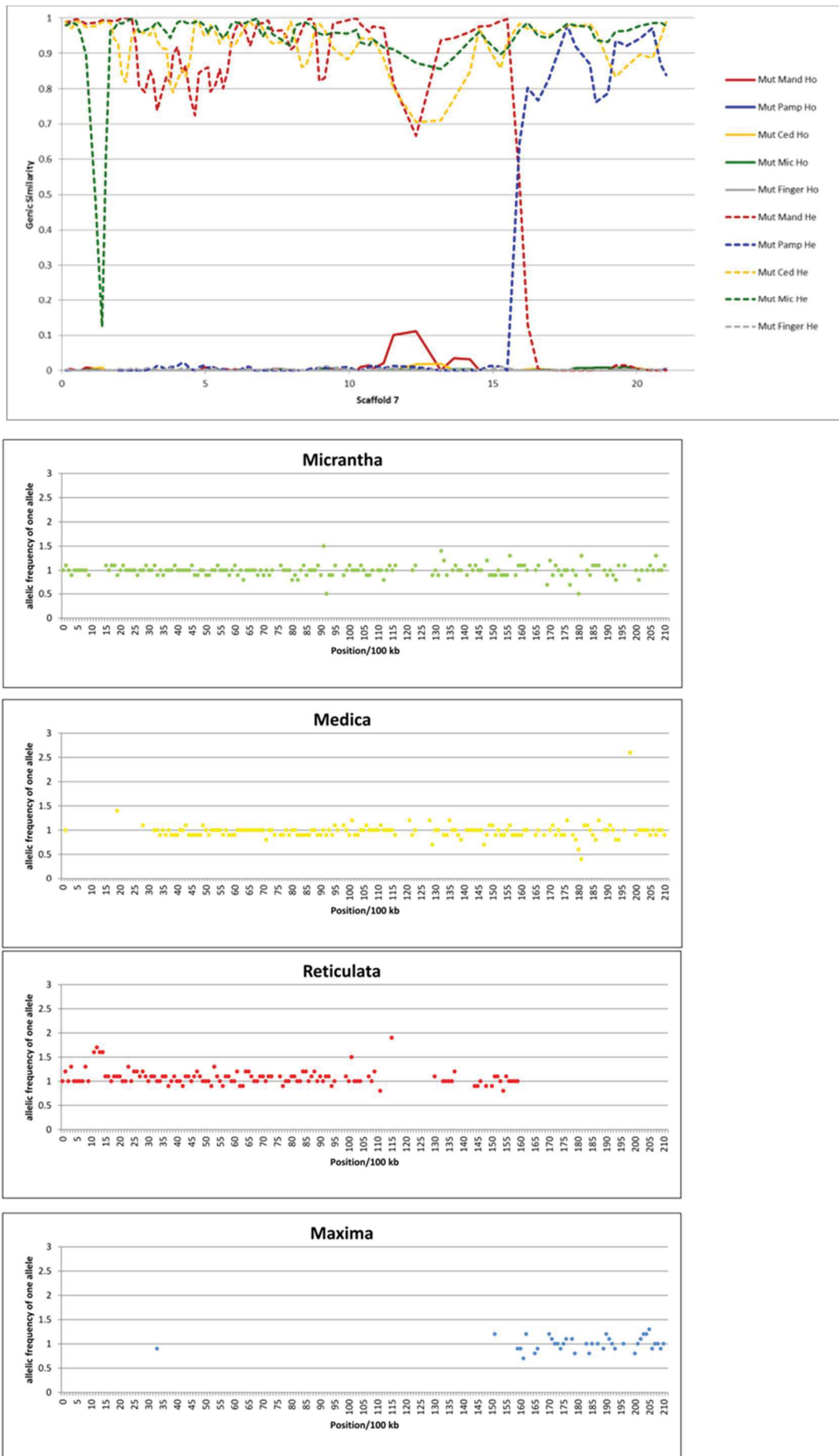


Figure S.IV.2g. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 7.

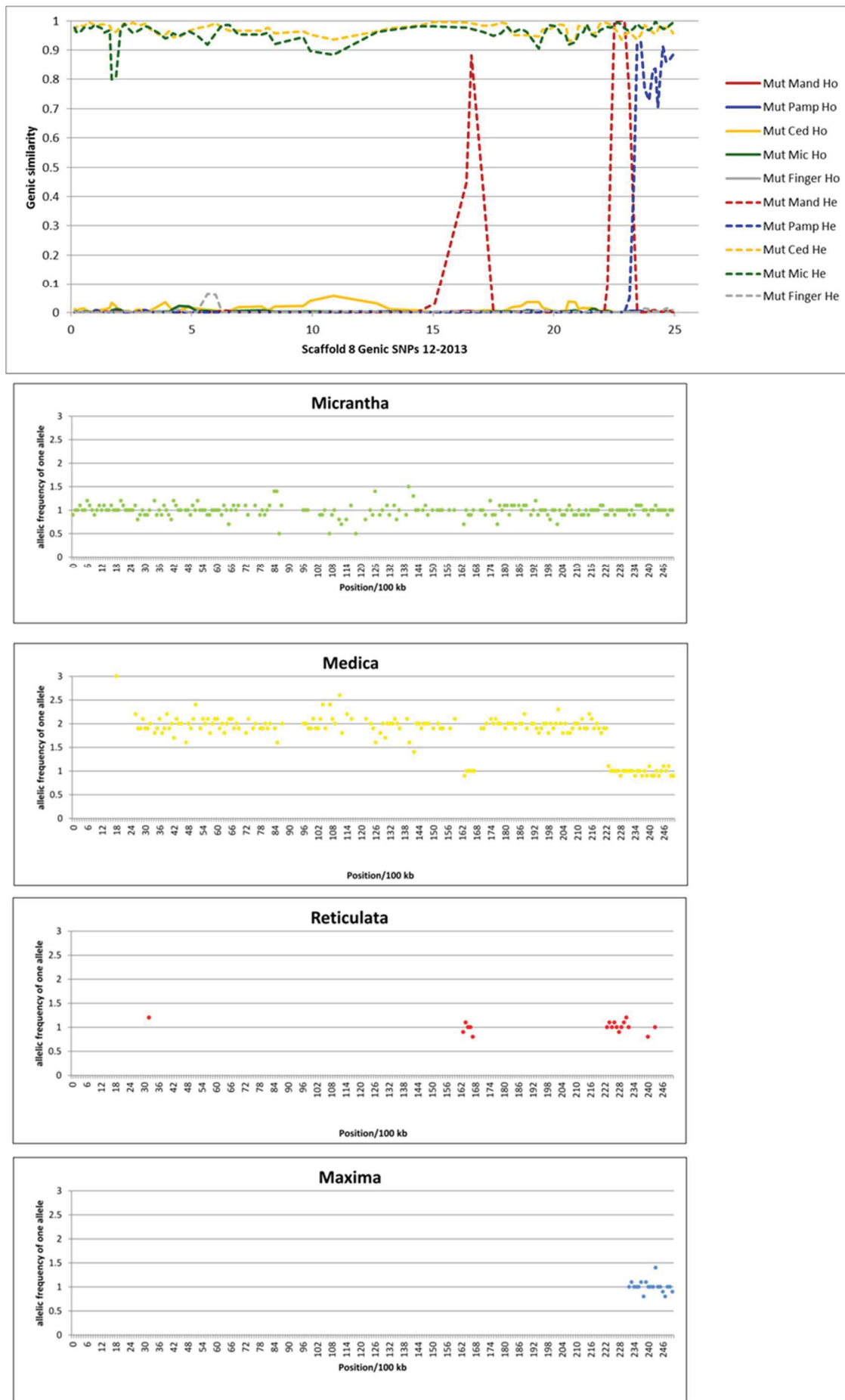


Figure S.IV.2h. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 8.

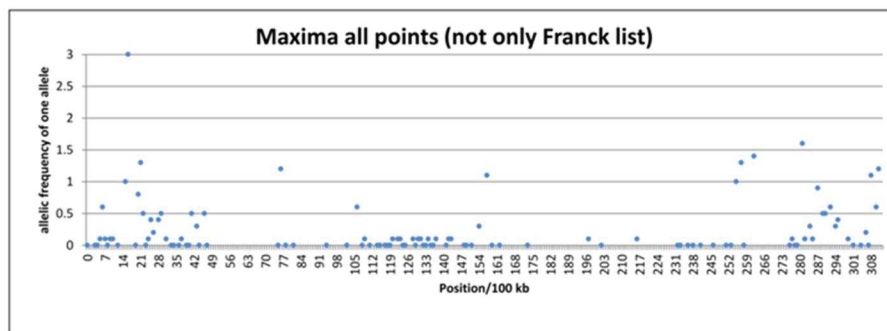
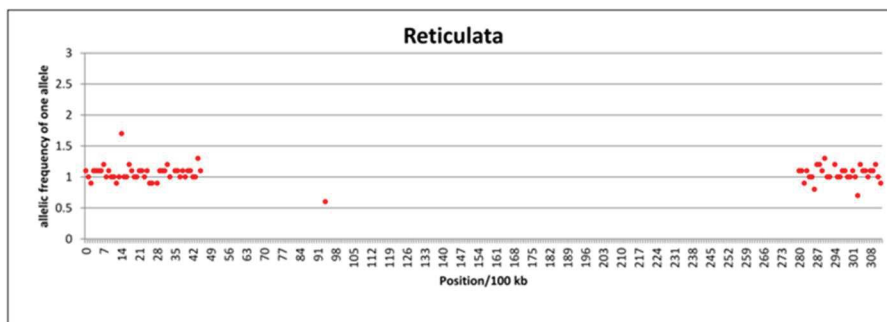
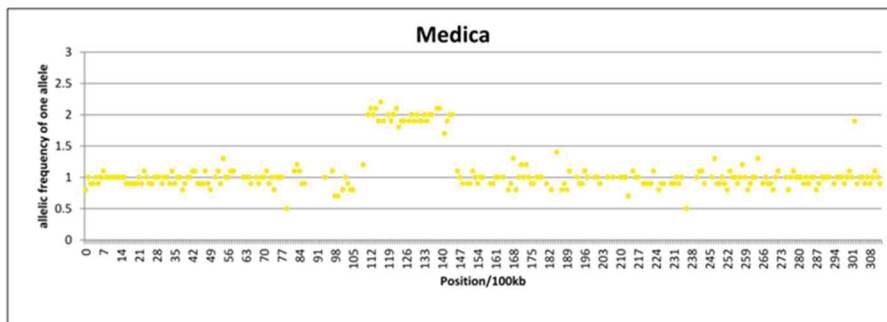
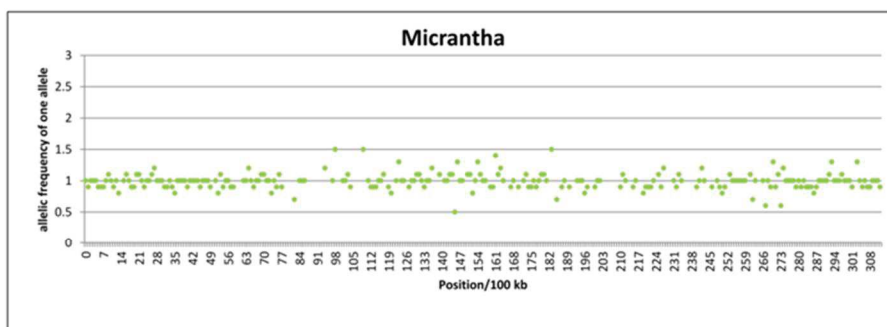
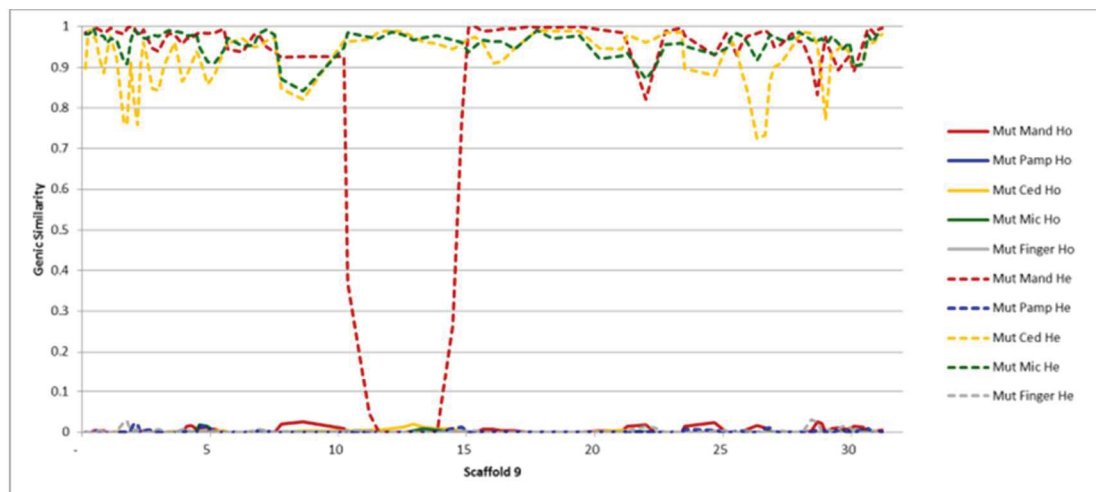


Figure S.IV.2i. Phylogenomic structure of the triploid ‘Bears’ lime for chromosome 9.