DISS. ETH NO. 27778

CLOSING IN ON GENOMICS-ASSISTED BREEDING IN APPLE

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

MICHAELA JUNG

Ing., Technical University in Zvolen

born on 14.09.1991

citizen of

Slovak Republic

accepted on the recommendation of

Prof. Dr. Bruno Studer, examiner

Dr. Andrea Patocchi, co-examiner

Dr. Hélène Muranty, co-examiner

Prof. Dr. Fred van Eeuwijk, co-examiner

2021

Table of contents

Summary

Apple (*Malus domestica* Borkh.) is globally the third most important fruit crop. It is used for cooking, in beverages or to be eaten raw. Changing consumer preferences, climate and pressure of pests create a demand for new varieties. However, breeding of an apple variety is a decades-long process. To improve breeding efficiency, the traditional phenotypic selection can be complemented with marker-assisted and/or genomic selection. For the application of genomic selection, genomic and phenotypic data of a reference population are used to train a genomic prediction model. This model is then deployed to make predictions of genomic-estimated breeding values for genotyped breeding material. As the breeding material is not phenotyped, the costs for phenotyping are saved. It was estimated that genomic selection can increase the rate of genetic gain.

Recent development of genomic tools in apple opened pathways towards genomic selection in this crop. Reference genomes permitted the development of single nucleotide polymorphism (SNP) arrays, which subsequently allowed for genotyping of hundreds of apple genotypes at a very high marker density. The availability of genotypic information for a large collection of genotypes was the basis for the choice of the plant material to form the apple reference population (apple REFPOP), which in turn became the cornerstone for this thesis.

The general introduction (**Chapter 1**) of the thesis outlines the history of plant breeding, the development of genetic markers, genetic mapping, and marker-assisted selection. Furthermore, it describes genomic selection and genome-wide association studies (GWAS), which were the two main statistical approaches for the analysis of genomic and phenotypic data applied in the thesis. A description of the past and present of the cultivated apple is followed by a characterization of genomic tools available for this species. Finally, the research aims of the thesis are presented.

In the first article (**Chapter 2**), the apple REFPOP is described and analyzed for its applicability to genomics-assisted breeding. The population was designed as a diverse collection of accessions originating from European apple collections and progeny from recent European breeding programs. Six identical sets, each containing at least two replications of all 534 genotypes, were planted in 2016 at six locations across Europe and so presumably became the largest multinational experimental design in apple worldwide. Here, imputation was used to consolidate a high-density dataset of 303,239 SNP markers for all apple REFPOP genotypes with an imputation accuracy of 0.95. The genomic dataset was then applied to estimate population characteristics of the apple REFPOP, revealing low linkage disequilibrium and weak population structure. An initial phenotypic dataset of two phenology traits, i.e., floral emergence and harvest date, was established from measurements performed at the apple REFPOP locations during one season. The full set of phenotypic and genomic data was used to verify if the known marker-trait associations could be rediscovered performing a GWAS, and the analysis revealed associations in several known genomic regions. The same dataset was then used to perform genomic predictions obtaining average genomic prediction accuracies of 0.57 and 0.75 for floral emergence and harvest date, respectively. Additionally, statistical analyses with decreasing SNP densities showed various marker-trait associations depending on the studied trait and a plateau of genomic prediction accuracy at 10,000 SNPs regardless of

the trait. Results reported in this chapter deem the apple REFPOP suitable for genomics-assisted breeding.

In the second article (**Chapter 3**), the phenotypic dataset was expanded to 30 traits related to fruit quality, phenology, vigor, and productivity measured across the apple REFPOP locations and up to three seasons. Both phenotypic and genomic data built a dataset of an unprecedented size in apple. These data facilitated characterization of genetic architecture for all traits when reporting 59 stable and 277 location-specific marker-trait associations using GWAS. The obtained marker-trait associations were compared with loci reported in 41 reviewed QTL mapping studies and GWAS spanning more than two decades, showing that 69% of the reported associations were novel. With the goal to capture and identify different relationships between markers and phenotypes and to find the best modelling solution for individual traits, the dataset was further exploited when fitting single-environment univariate, single-environment multivariate, multi-environment univariate, and multi-environment multivariate genomic prediction models for the 30 traits and reporting average genomic prediction accuracy of 0.18–0.88, 0.27–0.78, 0.21–0.86 and 0.13–0.85, respectively. Based on the obtained results, strategies for future implementation of genomic prediction tools in apple breeding were discussed.

The general discussion (**Chapter 4**) revisits the analyses reported in both the first and the second article to critically evaluate applied methodologies and suggest their future improvements. This chapter identified possibilities for more efficient phenotyping and genotyping and ways to improve genomic prediction accuracy. It also addressed the puzzle of the genotype by environment interaction, remaining challenges of the genome-wide association studies and some future applications of the apple REFPOP towards the utilization of genomic selection in apple.

Zusammenfassung

Der Apfel (*Malus domestica* Borkh.) ist weltweit die drittwichtigste Obstart. Er wird zum Kochen, in Getränken oder zum Rohverzehr verwendet. Sich ändernde Verbraucherpräferenzen, Klima und Schädlingsdruck schaffen einen Bedarf an neuen Sorten. Die Züchtung einer Apfelsorte ist jedoch ein jahrzehntelanger Prozess. Um die Züchtungseffizienz zu verbessern, kann die traditionelle phänotypische Selektion durch markergestützte und/oder genomische Selektion ergänzt werden. Bei der Anwendung der genomischen Selektion werden genomische und phänotypische Daten einer Referenzpopulation verwendet, um ein genomisches Vorhersagemodell zu trainieren. Dieses Modell wird daraufhin eingesetzt, um Vorhersagen von genomisch geschätzten Zuchtwerten für das Zuchtmaterial zu treffen. Hierbei wird das Zuchtmaterial nur genotypisiert, wodurch sich die Kosten für die Phänotypisierung reduzieren.

Jüngste Entwicklungen genomischer Werkzeuge eröffneten neue Wege der genomischen Selektion im Apfel. So erlaubten Referenzgenome die Entwicklung des Einzelnukleotid-Polymorphismus (SNP), der in der Folge die Genotypisierung hunderter Apfelgenotypen mit einer sehr hohen Markerdichte ermöglichte. Die Verfügbarkeit dieser genotypischen Informationen hat die Grundlage für die Auswahl der Apfelgenotypen der Apfel-Referenzpopulation (Apfel REFPOP) gebildet, welche wiederum der Grundstein dieser Arbeit ist.

Die allgemeine Einleitung (**Kapitel 1**) der Arbeit skizziert die Geschichte der Pflanzenzüchtung, die Entwicklung der genetischen Marker, die genetische Kartierung und die markergestützte Selektion. Darüber hinaus werden die genomische Selektion und genomweite Assoziationsstudien (GWAS) beschrieben, welche in dieser Arbeit die zwei wichtigsten statistischen Ansätze für die Analyse genomischer und phänotypischer Daten darstellen. Nach einer Beschreibung der Vergangenheit und Gegenwart des Kulturapfels folgt eine Charakterisierung der für diese Art verfügbaren genomischen Werkzeuge. Schließlich werden die Forschungsziele der Arbeit vorgestellt.

Im ersten Artikel (**Kapitel 2**) wird die Apfel REFPOP beschrieben und auf ihre Anwendbarkeit für die Genom-gestützte Züchtung hin analysiert. Die Population wurde als eine vielfältige Sammlung von Akzessionen aus europäischen Apfelsammlungen und Nachkommen aus aktuellen europäischen Zuchtprogrammen konzipiert. Sechs identische Sets, die jeweils mindestens zwei Replikationen aller 534 Genotypen enthielten, wurden 2016 an sechs Standorten in Europa gepflanzt. So wurden sie meines Wissens zum weltweit größten multinationalen Versuchsdesign beim Apfel. Eine Imputation hat es ermöglicht, einen dichten Datensatz von 303.239 SNP-Markern mit einer Imputationsgenauigkeit von 0.95 für alle Genotypen der REFPOP zu konsolidieren. Dieser genomische Datensatz wurde dann zur Schätzung der Populationscharakteristika der Apfel REFPOP verwendet, wobei sich ein geringes Kopplungsungleichgewicht und eine schwache Populationsstruktur zeigte. Ein erster phänotypischer Datensatz zweier phänologischer Merkmale, sprich Blühbeginn und Erntetermin, wurde aus Messungen erstellt, die an den Apfel REFPOP-Standorten während einer Saison durchgeführt wurden. Es wurden GWAS und genomische Vorhersageanalysen durchgeführt, um zu überprüfen, ob die Apfel REFPOP für Anwendungen der genomgestützten Züchtung gut geeignet ist. Der vollständige Satz an phänotypischen und genomischen Daten wurde verwendet, um

zu überprüfen, ob die bekannten Marker-Merkmal-Assoziationen durch eine GWAS wiederentdeckt werden konnten. Die Analyse zeigte Assoziationen in mehreren bekannten genomischen Regionen. Derselbe Datensatz wurde dann verwendet, um genomische Vorhersagen durchzuführen, wobei eine durchschnittliche genomische Vorhersagegenauigkeit von 0,57 bzw. 0,75 für den Blütenaufgang bzw. das Erntedatum erzielt wurde. Zusätzlich zeigten statistische Analysen mit abnehmender SNP-Dichte verschiedene Marker-Merkmal-Assoziationen in Abhängigkeit vom untersuchten Merkmal. Zudem zeigte sich ein Merkmal–unabhängiges Plateau der genomischen Vorhersagegenauigkeit bei 10.000 SNPs. Die in diesem Kapitel beschriebenen Ergebnisse sprechen für die Eignung der Apfel REFPOP für die genomgestützte Züchtung.

Im zweiten Artikel (**Kapitel 3**) wurde der phänotypische Datensatz auf 30 Merkmale erweitert, die sich auf Fruchtqualität, Phänologie, Wuchsstärke und Produktivität beziehen und bis zu drei Jahren hinweg an den Apfel REFPOP-Standorten gemessen wurden. Sowohl die phänotypischen als auch die genomischen Daten bildeten für den Apfel einen Datensatz von noch nie dagewesener Größe. Diese Daten erleichterten die Charakterisierung der genetischen Architektur für alle Merkmale, indem sie 59 stabile und 277 standortspezifische Marker-Merkmal-Assoziationen mittels GWAS aufzeigten. Die erhaltenen Marker-Merkmal-Assoziationen wurden mit Loci, welche in 41 QTL-Kartierungsstudien und GWAS über mehr als zwei Jahrzehnte berichtet wurden, verglichen und zeigten, dass 69% der berichteten Assoziationen neuartig sind. Mit dem Ziel, verschiedene Beziehungen zwischen Markern und Phänotypen zu erfassen und zu identifizieren und die beste Modellierungslösung für einzelne Merkmale zu finden, wurde der Datensatz weiter ausgewertet. Hierbei wurden univariate Einzel-Umwelt-, multivariate Einzel-Umwelt-, univariate Multi-Umwelt- und multivariate Multi-Umwelt-Modelle zur genomischen Vorhersage für die 30 Merkmale angepasst und eine durchschnittliche genomische Vorhersagegenauigkeit von 0,18-0,88, 0,27-0,78, 0,21-0,86 bzw. 0,13-0,85 berichtet. Basierend auf den erzielten Ergebnissen wurden Strategien für die zukünftige Implementierung von Werkzeugen zur genomischen Vorhersage in der Apfelzüchtung diskutiert.

In der allgemeinen Diskussion (**Kapitel 4**) werden die im ersten und zweiten Artikel berichteten Analysen erneut betrachtet, um die angewandten Methoden kritisch zu bewerten und deren weitere Entwicklung vorzuschlagen. In diesem Kapitel wurden Möglichkeiten für eine effizientere Phänotypisierung und Genotypisierung sowie Wege zur Verbesserung der genomischen Vorhersagegenauigkeit aufgezeigt. Zusätzlich wurde sich mit dem Rätsel der Genotyp-Umwelt-Interaktion, den verbleibenden Herausforderungen der genomweiten Assoziationsstudien und einigen zukünftigen Anwendungen der Apfel REFPOP für die Nutzung der genomischen Selektion beim Apfel auseinandergesetzt.

1 General introduction

1.1 History of plant breeding

The origin of plant breeding dates back to the beginnings of cultivation and ultimate domestication of many wild plant species, which started about 13.000 to 11.000 years ago[1]. Natural selection under conditions of cultivation accompanied by artificial selection[1], which is a conscious choice of plants featuring traits favored by humans, gave rise to countless crop species and landraces over thousands of years.

The way to modern plant breeding was paved since the nineteenth century, when Darwin used artificial selection after domestication as a starting line to introduce his theory of natural selection[2]. In the same period, Mendel established his laws of inheritance on an example of simple traits in pea plants[3]. Understanding of complex traits was shaped by the biometric views of Pearson, Galton, and others, without explicit knowledge about inheritance. The biometricians entered an intellectual battle against the Mendelians, which culminated in 1918, when Fisher formulated the hypothesis of cumulative Mendelian factors[4]. These factors essentially represented many individual genes, which together determine the quantitative nature of complex phenotypic traits. As Darwinian and Mendelian evolutionary principles became firmly established early in the twentieth century, plant breeding transformed into a scientific process[1]. Nevertheless, even after the double-helix structure of DNA was described in 1953, only phenotypes and pedigree records were used for breeding of quantitative traits until suitable genetic markers were brought to light.

1.2 Genetic markers

Following the development of techniques to detect a restriction fragment length polymorphism (RFLP), the usefulness of genetic markers for plant breeding was recognized in the late 1980s[5]. The RFLPs identified differences in genetic information carried by individuals using restriction enzymes, which cut DNA molecules into detectable restriction fragments. Based on the game-changing discovery of the polymerase chain reaction (PCR), other marker types such as the random amplification of polymorphic DNA (RAPD), sequence characterized amplified region (SCAR), cleaved amplified polymorphic sequence (CAPS), amplified fragment length polymorphism (AFLP) or simple sequence repeat (SSR) were developed. Especially the reproducible and highly polymorphic SSRs have been implemented in a variety of applications in plant genetics and breeding such as genetic mapping, DNA fingerprinting or study of genetic diversity[6]. As a further improvement, multiplexing allowed for simultaneous use of multiple SSRs in the same PCR reaction. With the advancement of the sequencing technologies, these low-marker-density approaches could be largely replaced with a direct analysis of sequence differences between individuals using single nucleotide polymorphism (SNP) markers or insertions/deletions (indels). Development of SNP arrays and genotyping-by-sequencing allowed for genotyping of individuals with thousands of markers at the same time[7-9].

1.3 Genetic mapping

With the advent of genetic markers, the resolution of quantitative traits into Mendelian factors became possible using genetic linkage mapping[10]. For its application, a biparental mapping population is generally developed from a cross between two parents chosen to promote variation for the trait of interest in their progeny. The phenotypes are assessed, and genetic markers are scored in the mapping population. Provided the markers are evenly spread over the genome, a better mapping resolution is reached with greater number of markers[11]. Statistical methods are finally used to identify genetic markers that are significantly associated with phenotypic variation. The inference of the relationship between genotype and phenotype can consist in mapping of single genes responsible for simple traits or quantitative trait locus (QTL) mapping of several large-effect loci affecting variation of oligogenic quantitative traits.

1.4 Marker-assisted selection

Genetic markers significantly associated with phenotypic variation are the prerequisite for marker-assisted selection (MAS). In this selection process, genetic markers associated with traits of interest are used to choose individuals carrying the favorable allele(s). MAS performed at seedling stage eliminates the phenotyping for the traits associated with markers used, thus saving labor and/or maintenance costs[12]. MAS can be efficient in case the marker was associated with a trait that can be phenotypically assessed only after years of cultivation (e.g., fruit-related traits in fruit trees, wood quality in forest trees). The method brought benefits such as sidestepping inoculation with pathogens when breeding for pathogen resistance[13]. Nevertheless, the range of MAS applications remained mostly restricted to simple and oligogenic traits.

1.5 Genomic selection

Entering the genomics era, the density of genetic markers available across plant genomes increased to thousands. In 2001, a method of genomic selection for capturing phenotypic variation of quantitative traits underlain by many genes of small effect was presented by Meuwissen et al.[14]. In genomic selection, a training population is commonly screened for hundreds to thousands of phenotypes and many thousands of genetic markers. Both phenotypes and genotypes enter a genomic prediction model to finally obtain genomic-estimated breeding values (GEBVs) for new breeding material. This allows to make selections on breeding material for different traits using the marker information alone, thus eliminating costly phenotyping of polygenic traits (Figure 1).

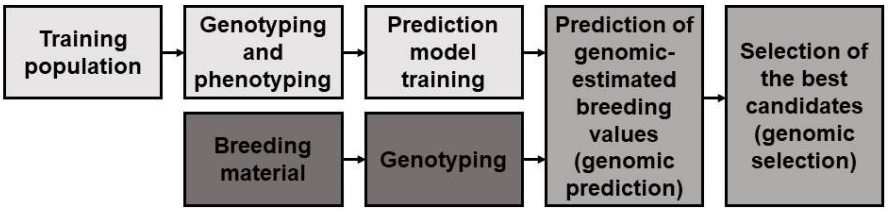


Figure 1: Traditional visualization of the process of genomic selection adapted from Heffner et al.[15]

1.5.1 Prediction accuracy

To measure correctness of a prediction algorithm, prediction accuracy can be estimated as a correlation between the measured breeding values (i.e., phenotypes) and the GEBVs. To obtain prediction accuracy estimates, model validation techniques such as cross-validation are deployed to show how the statistical method generalizes to an independent dataset. In an example of a five-fold cross-validation, all available genotyped and phenotyped individuals are split into two parts. First, a training set of 80% of the individuals is used to estimate marker effects. Second, the remaining 20% of individuals form a test set, which is used to make predictions (i.e., GEBVs) and compare these with the true breeding values via correlation. Even when the test set is closely genetically related to the training set, as the test set is not involved in model training, its errors are independent and any correlation between predicted and realized values is due to prediction of true effects[16]. In the five-fold cross-validation, sampling of individuals into the training and test sets is repeated five times. Sampling without replacement ensures that all individuals contribute to the comparison of predicted and realized values.

Caution is advised in the estimation of prediction accuracy obtained from cross validation under population structure, especially when the training population involves genotypes only distantly related to the breeding population[17]. In such a case, estimated genomic prediction accuracies might not represent the accuracies of a commercial breeding population[17]. Both population structure and relatedness within and between the training set and the breeding material may also affect the size of the training population necessary for accurate predictions. In earlier simulations, a small training population was sufficient for predictions in closely related biparental populations, the opposite was true for distantly related individuals[18].

Another challenge connected with genomic prediction accuracy is the marker density. Dense genome-wide marker coverage leading to a linkage disequilibrium between each causal locus with at least one marker is needed for accurate predictions[15]. An associated factor is the genetic architecture of a trait, where the number of loci underlying the trait affect heritability and predictability of the trait, the model accuracy decreasing with an increasing trait complexity[19]. Depending on trait architecture, different marker densities, feature (i.e., marker) selection methods and genomic prediction models may contribute to optimizing prediction performance[19,20].

1.5.2 Genomic prediction models

In genomic prediction models, all available marker information is incorporated simultaneously. This is an advantage over least-squares analyses, a standard approach in regression analysis, where the large number of predictors (i.e., markers) and a comparably low sample size (i.e., the size of the training population), also called the problem of the 'large p and small n', lead to an insufficient number of degrees of freedom to fit all marker effects at once. If only the biggest effects are included in the least-squares model instead, these become overestimated and lead to a strong decrease in prediction accuracy[14].

With the aim of incorporating the whole marker information without variable selection and to optimize model performance, numerous parametric, semi-parametric and non-parametric statistical methods were proposed for genomic prediction[21]. As one of the first

parametric tools, the best linear unbiased prediction (BLUP) was proposed by Meuwissen et al.[14]. The BLUP together with a mixed model formulation originated in the work of Henderson from 1949 and has been widely expanded ever since[22,23]. The BLUP is equivalent to ridge regression (RR) and also known as the 'RR-BLUP'[24]. Although unrealistically, it assumes that every marker explains an equal amount of variance, marker effects follow a normal distribution and thus equal shrinkage applies for each marker effect[14,21]. The same assumptions apply to the Bayesian ridge regression but the level of shrinkage is estimated using a Bayesian hierarchical model with a Gaussian prior[21]. Preceding the development of genomic selection, the BLUP was traditionally used with a pedigree relationship matrix, which describes covariances between genotypes based on their ancestry[23]. Meuwissen et al.[14] founded their BLUP on a design matrix composed of single markers where the model first estimated and then summed all marker effects for every individual. In the genomic BLUP or 'G-BLUP', VanRaden[25] replaced the traditional pedigree relationship matrix by a genomic relationship matrix to facilitate efficient incorporation of the rapidly expanding numbers of markers. The Bayesian least absolute shrinkage and selection operator, the 'Bayesian LASSO', includes a marker-specific shrinkage of regression coefficients using a double-exponential prior to accommodate the large number of genetic markers[26]. In BayesA and BayesB, the variance of each marker can vary based on a combination of information from a prior distribution of the variances and a distribution of the data[14]. BayesB uses a prior π with a high probability that a marker has a zero effect (original $\pi = 0.95$), because in reality there are many markers with no genetic variance and a few markers associated with segregating QTL[14]. BayesA does not have a prior density peak at zero variance, but the probability that a marker has zero effect is infinitesimal ($\pi = 0$)[14]. BayesCπ treats the prior probability π that a genetic marker has zero effect as an unknown parameter that can be estimated[27]. The semi-parametric reproducing kernel Hilbert spaces (RKHS) regression relaxes the assumptions about the linear form of the marker–phenotype relationship and refrains from using parametric distributions for marker effects[28]. Because it does not assume linearity, the RKHS regression might better capture non-additive effects[21]. Random forest[29] (RF) is a non-parametric solution to the problem of large p and small n. The RF and other machine learning methods such as artificial neural networks may be able to capture non-linear relationships between markers and phenotypes better than their parametric counterparts[21]. The majority of genomic prediction models consider additive effects associated with SNP markers only, but extensions of additive models which partition variance in statistical components due to additivity, dominance, and epistasis are also available[30].

Comparing various methods, Heslot et al.[21] observed similar accuracy across models, but the level of overfitting, computation time and the distribution of marker effects varied between models. When Howard et al.[22] compared the performance of parametric and non-parametric methods in a combination with additive and non-additive epistatic genetic architecture of traits, only incremental distinctions were observed among parametric models and their performance was better with additive genetic architecture, while the non-parametric models succeeded under epistasis. Genomic prediction based on additive genetic effects can accelerate genetic gain and non-additive effects offer further opportunities for improvement[31].

1.5.3 Multi-environment testing and prediction

If the same plant genotypes are propagated and grown across wider areas, various responses to different environments from the same genotype may be expected due to genotype be environment interaction effects. For genotype by environment interaction, loci underlying traits vary across environments. Within a breeding program, the genotypes might be tested over a series of environments to estimate the interaction effects. The data collected from such multi-environment trial can be used to estimate variance components for genotypes, genotype by environment interaction and error, which gives a partitioning of the phenotypic variance among genotype means and can be used to estimate trait heritability over all tested environments[32]. As the magnitude of the genotype by environment interaction increases relative to the genotypic component, both heritability and response to selection decrease[32].

Many of the available genomic prediction models originate in animal breeding, where environmental effects usually play a minor role. As the environmental effects cannot be considered by the traditional prediction models *per se*, the analysis of the multi-environment trials can be split into two steps. In the first step, mixed models are used to account for environmental variables and produce mean phenotypic values for individuals. Only in the second step, the breeding values are calculated jointly for all environments using genomic prediction. As a complement to this two-step methodology, one-step approaches provide with the possibility to estimate the effect of environment and genotype by environment interaction at the same time as the breeding values. These models can be used to predict performance of untested individuals in known environments[33]. The genotype by environment interaction still poses a challenge for the implementation of genomic selection in plant breeding[31].

1.5.4 Multivariate prediction

In breeding programs phenotyping for multiple traits, correlations may occur between the measured traits. To take advantage of the information that one trait carries about another, multivariate genomic prediction on multiple traits can be used instead of traditional single-trait (univariate) models[34]. Multivariate genomic prediction is able to increase prediction accuracy when applied to a combination of low- and high-heritability traits and if missing data is present in the dataset of traits[34]. It may also be useful for correlated traits when the measurement of some traits is costly[35]. Additionally, the approach can be applied to the estimates of a single trait from different environments that are treated as distinct traits by the model[36]. The multivariate models became another useful element in the genomic prediction toolbox.

1.5.5 Implementation of genomic prediction

Already in 2001, Meuwissen et al.[14] concluded from their simulations that genomic selection could substantially increase the rate of genetic gain in animals and plants, especially if combined with reproductive schemes to shorten the generation interval such as the early use of genetically superior individuals for further breeding. Over the years, additional simulations have confirmed that genomic selection would substantially enhance gains per unit time, but empirical validation remained outstanding[15,37]. Genomic selection was implemented for the first time in dairy cattle in 2008. During the first seven years since its introduction in the United States dairy cattle breeding, genomic selection

led to rates of genetic gain per year that were doubled for yield traits and even tripled for lowly heritable traits[38]. Although there are limited reports about the actual impact of genomic selection on genetic gain in plants[31], the beginning of the last decade has brought numerous implementations of genomic prediction in annual crops such as maize[39] or wheat[40,41] as well as in perennials such as loblolly pine[42] or apple[43]. Maize became a prime example for the application of genomic selection in commercial breeding schemes, where the method contributed to an increase in yield under drought-stress conditions of the United States corn belt[44].

1.6 Genome-wide association studies

If genomic selection is sometimes called the MAS on a genome wide scale[16], then genome-wide association studies (GWAS) are analogous to QTL mapping. In contrast to QTL mapping, which explores a single mapping population with a limited number of segregating alleles, GWAS investigate thousands to millions of markers across natural populations or diverse individuals. The goal of GWAS is to understand genetic variation underlying traits of interest by identifying genomic regions associated with these traits. Established early in the twenty-first century in human genetics[45], GWAS have found a plethora of associations in humans and become the most widely used approach to relate genetic variation to phenotypic diversity[46]. In crops, more than 1,000 GWAS have revealed a substantial number of genotype–phenotype associations[47].

Despite its general success, a largely discussed problem of GWAS lies in the spurious associations, i.e., false positives, produced for traits correlated with population structure[48]. Several milestones in method development were reached when incorporating population structure, cryptic relationships among individuals within sub-populations as well as increasing computing efficiency and statistical power until advanced methods such as Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK), which is optimized for millions of markers and individuals, could be developed[49].

1.7 Cultivated apple

The cultivated apple (*Malus domestica* Borkh.) belongs to the Rosaceae family, genus *Malus*. It is an allopolyploid, but behaves like a diploid, although triploid genotypes are not uncommon[50]. Its genome estimated to 709–719 Mb[51,52], which has undergone a rather recent whole-genome duplication[53,54], is composed of 17 chromosomes. The species is mostly self-incompatible and thus highly heterozygous[55], and it has a juvenile phase of at least four to five years[56].

M. domestica Borkh. originated in the populations of its wild ancestor *M. sieversii* Ldb. in the Tian Shan mountains of Central Asia[57]. After its domestication, the apple has been dispersed from Central Asia to Europe along the Silk Road[57,58]. This process promoted an intensive hybridization and introgression with the European crabapple (*M. sylvestris* Mill.) and a less pronounced gene flow from the Caucasian crabapple (*M. orientalis* Uglitz.)[57]. Grafting facilitated propagation and conservation of superior genotypes. The cultivated apple has developed into a panmictic, well-separated species[57], the third most produced fruit crop worldwide (Figure 2)[59].

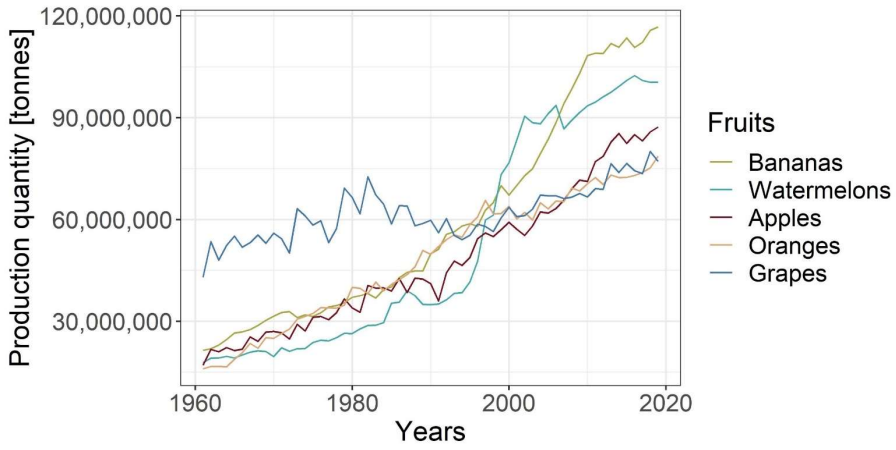


Figure 2: Production quantity across years worldwide for the five currently most produced primary fruit crops[59]

Thousands of apple cultivars were described and are conserved in repositories worldwide[60]. Despite the large genetic variation described in the European germplasm[61], the recently reconstructed pedigrees of some of these heritage cultivars showed extensive relatedness among the studied germplasm with a strong influence of a small number of cultivars through the history of apple breeding[62,63]. The study revealed a few major founders, namely two Renaissance cultivars 'Reinette Franche' and 'Margil' and an eighteenth century Russian cultivar 'Alexander'[62]. In the United States germplasm collection, the nineteenth century cultivars 'Golden Delicious' and 'Red Delicious' showed numerous first-degree relatives, consistent with their repeated use in breeding[63]. In Europe, the discrepancy between the strong relatedness of cultivars and their high genetic diversity may be explained by little inbreeding in the germplasm and the fact that crosses often, especially since the nineteenth century, took place between cultivars from different regions in Europe[62].

On a long term, vegetative propagation is used to conserve cultivars of interest. However, plant pathogens evolve rapidly due to their short life cycles and cause epidemics since the dawn of agriculture. Monocultures typical for intensive crop production further accelerate the spread of virulent strains. Only a fraction of the existing apple cultivars is grown commercially and the majority of these require high doses of pesticides to prevent crop failure[60,64]. Genetic improvement in apple is needed to cope with pests and diseases as well as future climatic constraints and to meet consumer demands.

1.8 Genomic tools in apple

Pathogen resistance combined with high yield and fruit quality is generally the aim of an apple breeding program. MAS provided breeders with the first tools to test genetic background of their breeding material at a juvenile growth stage. Protocols were established for the implementation of MAS in breeding programs using several markers[65]. The markers have been developed for traits such as the resistance to fire blight, apple scab, powdery mildew or a few fruit quality traits[65,66]. In the Genome Database for Rosaceae[67], 1,595 QTL (from that 151 QTL for various disease resistances) were reported for *M. domestica* Borkh. as of February 2021 (https://www.rosaceae.org/search/qtl). Nevertheless, MAS remained impractical for numerous traits related to yield and fruit quality.

The first genome sequence of a diploid apple cultivar 'Golden Delicious'[54] aided the development of an Illumina Infinium® 8K SNP array[68]. In the following decade, further apple genome sequences were released, namely those of the 'Golden Delicious' doubled-haploid GDDH13[69], of the anther-derived homozygous genotype HFTH1 of Chinese parentage[51], and of the cultivars 'Gala'[70] and 'Gala Galaxy'[52]. More recent SNP arrays (Illumina Infinium® 20K SNP array[9] and Affymetrix Axiom® Apple 480K SNP array[8]) or the genotyping-by-sequencing (GBS) approach[7] facilitated genotyping at medium to high densities.

These genomic tools enabled GWAS for several apple phenology traits, fruit quality traits, a productivity trait (preharvest fruit drop), volatiles, polyphenols, and also diseases (apple scab and apple *Marssonina* blotch, Table 1). Accession collections have been mostly deployed in GWAS, but progeny[71,72] or a combination of progeny and accessions[73] have also been used. Genotyping density has ranged between very low (less than 300 SNPs) and very high (7,218K SNPs, Table 1). GWAS of genomic datasets beyond 100K SNPs have been limited to flowering and harvest time, fruit firmness and skin color[58,74,75]. GAPIT[76], TASSEL[77] and MLMM[78] have been the most popular software used. Repeatedly, the GWAS have been based on less than 100 genotypes and marker-trait associations have often been discovered for some but not all studied traits (Table 1).

A similar spectrum of traits as in GWAS has also been explored in apple genomic prediction studies (see Table 2 for a full comparison), but many key traits related to productivity and fruit quality could not be predicted thus far. Accessions, progeny, or their combination have been used for training of genomic prediction models almost equally often. The genotyping density did not exceed 100K SNPs in any of the reviewed studies. RR-BLUP and G-BLUP have been the most often used genomic prediction models, and only one model comparison of just two models was performed[43]. Low to high prediction accuracies have been estimated using cross-validation or validation with full-sib progeny. The number of genotypes has ranged between 136 and 1,120. The past genomic prediction analyses have been carried out at a local scale with only one study estimating the effect of G×E[79]. Multivariate genomic prediction has not been applied in any of the reviewed studies.

1.9 Research aims of the thesis

The thesis is based on the apple reference population (apple REFPOP), which was established to overcome shortages of the past study designs in apple. The apple REFPOP has been genotyped at high marker density and phenotyped for numerous quantitative traits with the aim to

1) characterize the genetic architecture of traits with major importance for apple breeding,

2) develop genomic prediction models as a foundation for the application of genomic selection in breeding programs.

To this end, the measured genotypes and phenotypes were analyzed together in GWAS to identify loci associated with the studied traits. Furthermore, genomic prediction models were fitted to the genomic and phenotypic data to estimate prediction accuracy for the quantitative traits.

The objective of the first article (Chapter 2) was to test suitability of the apple REFPOP for GWAS and genomic prediction with an initial phenotypic dataset of two phenology traits, i.e., floral emergence and harvest date, which were measured at the apple REFPOP locations for one year. In the second article (Chapter 3), the phenotypic dataset was expanded to 30 traits related to fruit quality, phenology, vigor, and productivity measured at up to six locations and up to three years. The objective of this large-scale phenotyping was to identify marker-trait associations using GWAS and so to characterize genetic architecture of the studied traits. Furthermore, the data were deployed in the single-environment univariate, single-environment multivariate, multi-environment univariate, and multi-environment multivariate genomic prediction with the goal to capture and predict different relationships between markers and phenotypes for the various traits.

References

1 Allard, R. W. *Principles of plant breeding*. (John Wiley & Sons, 1999).

2 Darwin, C. *On the Origin of Species*. (J. Murray, 1859).

3 Mendel, G. *Versuche über Pflanzen-Hybriden*. (Naturforschender Verein, 1866).

4 Fisher, R. A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh* **52**, 399-433, doi:10.1017/S0080456800012163 (1918).

5 Tanksley, S. D., Young, N. D., Paterson, A. H. & Bonierbale, M. W. RFLP mapping in plant breeding: New tools for an old science. *Bio/Technology* **7**, 257-264, doi:10.1038/nbt0389-257 (1989).

6 Gupta, P. K. & Varshney, R. K. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica* **113**, 163-185, doi:10.1023/A:1003910819967 (2000).

7 Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE* **6**, e19379, doi:10.1371/journal.pone.0019379 (2011).

8 Bianco, L. *et al.* Development and validation of the Axiom®Apple480K SNP genotyping array. *The Plant Journal* **86**, 62-74, doi:10.1111/tpj.13145 (2016).

9 Bianco, L. *et al.* Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLOS ONE* **9**, e110377, doi:10.1371/journal.pone.0110377 (2014).

10 Paterson, A. H. *et al.* Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721-726, doi:10.1038/335721a0 (1988).

11 Mauricio, R. Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics* **2**, 370-381, doi:10.1038/35072085 (2001).

12 Wannemuehler, S. D. *et al.* A cost–benefit analysis of DNA informed apple breeding. *HortScience horts* **54**, 1998, doi:10.21273/hortsci14173-19 (2019).

13 Khan, M. A. *et al.* Development of molecular markers linked to the 'Fiesta' linkage group 7 major QTL for fire blight resistance and their application for marker-assisted selection. *Genome* **50**, 568-577, doi:10.1139/G07-033 (2007).

14 Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819 (2001).

15 Heffner, E. L., Sorrells, M. E. & Jannink, J.-L. Genomic selection for crop improvement. *Crop Science* **49**, 1-12, doi:10.2135/cropsci2008.08.0512 (2009).

16 Meuwissen, T. Genomic selection: marker assisted selection on a genome wide scale. *Journal of Animal Breeding and Genetics* **124**, 321-322, doi:10.1111/j.1439-0388.2007.00708.x (2007).

17 Werner, C. R. *et al.* How population structure impacts genomic selection accuracy in cross-validation: Implications for practical breeding. *Frontiers in Plant Science* **11**, doi:10.3389/fpls.2020.592977 (2020).

18 Hickey, J. M. *et al.* Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science* **54**, 1476-1488, doi:10.2135/cropsci2013.03.0195 (2014).

19 Clark, S. A., Hickey, J. M. & van der Werf, J. H. J. Different models of genetic variation and their effect on genomic evaluation. *Genetics Selection Evolution* **43**, 18, doi:10.1186/1297-9686-43-18 (2011).

20 Bermingham, M. L. *et al.* Application of high-dimensional feature selection: Evaluation for genomic prediction in man. *Scientific Reports* **5**, 10312, doi:10.1038/srep10312 (2015).

21 Heslot, N., Yang, H.-P., Sorrells, M. E. & Jannink, J.-L. Genomic selection in plant breeding: A comparison of models. *Crop Science* **52**, 146-160, doi:10.2135/cropsci2011.06.0297 (2012).

22 Howard, R., Carriquiry, A. L. & Beavis, W. D. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes|Genomes|Genetics* **4**, 1027-1046, doi:10.1534/g3.114.010298 (2014).

23 Henderson, C. R. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447, doi:10.2307/2529430 (1975).

24 Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* **4**, 250-255, doi:10.3835/plantgenome2011.08.0024 (2011).

25 VanRaden, P. M. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414-4423, doi:10.3168/jds.2007-0980 (2008).

26 de los Campos, G. *et al.* Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* **182**, 375-385, doi:10.1534/genetics.109.101501 (2009).

27 Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186, doi:10.1186/1471-2105-12-186 (2011).

28 Gianola, D., Fernando, R. L. & Stella, A. Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* **173**, 1761-1776, doi:10.1534/genetics.105.049510 (2006).

29 Breiman, L. Random forests. *Machine Learning* **45**, 5-32, doi:10.1023/A:1010933404324 (2001).

30 Varona, L., Legarra, A., Toro, M. A. & Vitezica, Z. G. Non-additive effects in genomic selection. *Frontiers in Genetics* **9**, doi:10.3389/fgene.2018.00078 (2018).

31 Voss-Fels, K. P., Cooper, M. & Hayes, B. J. Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics* **132**, 669-686, doi:10.1007/s00122-018-3270-8 (2019).

32 Cooper, M. & DeLacy, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* **88**, 561-572, doi:10.1007/BF01240919 (1994).

33 Hardner, C. M. *et al.* Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. *Horticulture Research* **6**, 6, doi:10.1038/s41438-018-0081-7 (2019).

34 Jia, Y. & Jannink, J.-L. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* **192**, 1513-1522, doi:10.1534/genetics.112.144246 (2012).

35 Lado, B. *et al.* Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theoretical and Applied Genetics* **131**, 2719-2731, doi:10.1007/s00122-018-3186-3 (2018).

36 Gianola, D. & Fernando, R. L. A multiple-trait Bayesian LASSO for genome-enabled analysis and prediction of complex traits. *Genetics* **214**, 305-331, doi:10.1534/genetics.119.302934 (2020).

37 Wong, C. K. & Bernardo, R. Genomewide selection in oil palm: Increasing selection gain per unit time and cost with small populations. *Theor Appl Genet* **116**, 815-824, doi:10.1007/s00122-008-0715-5 (2008).

38 García-Ruiz, A. *et al.* Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences* **113**, E3995-E4004, doi:10.1073/pnas.1519061113 (2016).

39 Zhao, Y. *et al.* Accuracy of genomic selection in European maize elite breeding populations. *Theoretical and Applied Genetics* **124**, 769-776, doi:10.1007/s00122-011-1745-y (2012).

40 Rutkoski, J. *et al.* Evaluation of genomic prediction methods for fusarium head blight resistance in wheat. *The Plant Genome* **5**, doi:10.3835/plantgenome2012.02.0001 (2012).

41 Heffner, E. L., Jannink, J.-L., Iwata, H., Souza, E. & Sorrells, M. E. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science* **51**, 2597-2606, doi:10.2135/cropsci2011.05.0253 (2011).

42 Zapata-Valenzuela, J. *et al.* SNP markers trace familial linkages in a cloned population of *Pinus taeda*—prospects for genomic selection. *Tree Genetics & Genomes* **8**, 1307-1318, doi:10.1007/s11295-012-0516-5 (2012).

43 Kumar, S. *et al.* Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). *PLOS ONE* **7**, e36674, doi:10.1371/journal.pone.0036674 (2012).

44 Gaffney, J. *et al.* Industry-scale evaluation of maize hybrids selected for increased yield in drought-stress conditions of the US corn belt. *Crop Science* **55**, 1608-1618, doi:10.2135/cropsci2014.09.0654 (2015).

45 Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95-108, doi:10.1038/nrg1521 (2005).

46 Frazer, K. A., Murray, S. S., Schork, N. J. & Topol, E. J. Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* **10**, 241-251, doi:10.1038/nrg2554 (2009).

47 Liu, H.-J. & Yan, J. Crop genome-wide association study: A harvest of biological relevance. *The Plant Journal* **97**, 8-18, doi:10.1111/tpj.14139 (2019).

48 Larsson, S. J., Lipka, A. E. & Buckler, E. S. Lessons from Dwarf8 on the strengths and weaknesses of structured association mapping. *PLOS Genetics* **9**, e1003246, doi:10.1371/journal.pgen.1003246 (2013).

49 Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**, doi:10.1093/gigascience/giy154 (2018).

50 Brown, S. in *Fruit Breeding* (eds Marisa Luisa Badenes & David H. Byrne) 329-367 (Springer US, 2012).

51 Zhang, L. *et al.* A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications* **10**, 1494, doi:10.1038/s41467-019-09518-x (2019).

52 Broggini, G. A. L. *et al.* Chromosome-scale de novo diploid assembly of the apple cultivar 'Gala Galaxy'. *bioRxiv*, 2020.2004.2025.058891, doi:10.1101/2020.04.25.058891 (2020).

53 Han, Y. *et al.* Integration of physical and genetic maps in apple confirms whole-genome and segmental duplications in the apple genome. *Journal of Experimental Botany* **62**, 5117-5130, doi:10.1093/jxb/err215 (2011).

54 Velasco, R. *et al.* The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42**, 833-839, doi:10.1038/ng.654 (2010).
55 Lassois, L. *et al.* Genetic diversity, population structure, parentage analysis, and construction of core collections in the French apple germplasm based on SSR markers. *Plant Molecular Biology Reporter* **34**, 827-844, doi:10.1007/s11105-015-0966-7 (2016).
56 Visser, T. Juvenile phase and growth of apple and pear seedlings. *Euphytica* **13**, 119-129, doi:10.1007/BF00033299 (1964).
57 Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I. & Gladieux, P. The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**, 57-65, doi:10.1016/j.tig.2013.10.002 (2014).
58 Duan, N. *et al.* Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications* **8**, 249, doi:10.1038/s41467-017-00336-7 (2017).
59 FAOSTAT (Food and Agriculture Organization of the United Nations, 2019).
60 Way, R. D. *et al.* Apples (*Malus*). *Acta Horticulturae*, 3-46, doi:10.17660/ActaHortic.1991.290.1 (1991).
61 Urrestarazu, J. *et al.* Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology* **16**, 130, doi:10.1186/s12870-016-0818-0 (2016).
62 Muranty, H. *et al.* Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biology* **20**, 2, doi:10.1186/s12870-019-2171-6 (2020).
63 Migicovsky, Z. *et al.* Genomic consequences of apple improvement. *Horticulture Research* **8**, 9, doi:10.1038/s41438-020-00441-7 (2021).
64 Simon, S., Brun, L., Guinaudeau, J. & Sauphanor, B. Pesticide use in current and innovative apple orchard systems. *Agronomy for Sustainable Development* **31**, 541-555, doi:10.1007/s13593-011-0003-7 (2011).
65 Baumgartner, I. O. *et al.* Development of SNP-based assays for disease resistance and fruit quality traits in apple (*Malus* × *domestica* Borkh.) and validation in breeding pilot studies. *Tree Genetics & Genomes* **12**, 35, doi:10.1007/s11295-016-0994-y (2016).
66 Jänsch, M. *et al.* Identification of SNPs linked to eight apple disease resistance loci. *Molecular Breeding* **35**, 45, doi:10.1007/s11032-015-0242-4 (2015).
67 Jung, S. *et al.* 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. *Nucleic Acids Res* **47**, D1137-D1145, doi:10.1093/nar/gky1000 (2019).
68 Chagné, D. *et al.* Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLOS ONE* **7**, e31745, doi:10.1371/journal.pone.0031745 (2012).
69 Daccord, N. *et al.* High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49**, 1099-1106, doi:10.1038/ng.3886 (2017).
70 Sun, X. *et al.* Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics* **52**, 1423-1432, doi:10.1038/s41588-020-00723-9 (2020).
71 Kumar, S. *et al.* Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* **14**, 393, doi:10.1186/1471-2164-14-393 (2013).
72 Kumar, S., Raulier, P., Chagné, D. & Whitworth, C. Molecular-level and trait-level differentiation between the cultivated apple (*Malus* × *domestica* Borkh.) and its

main progenitor Malus sieversii. *Plant Genetic Resources* **12**, 330-340, doi:10.1017/S1479262114000136 (2014).

73 Minamikawa, M. F. *et al.* Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. *Horticulture Research* **8**, 49, doi:10.1038/s41438-021-00485-3 (2021).

74 Hu, Y. *et al.* ERF4 affects fruit firmness through TPL4 by reducing ethylene production. *The Plant Journal* **103**, 937-950, doi:10.1111/tpj.14884 (2020).

75 Urrestarazu, J. *et al.* Genome-wide association mapping of flowering and ripening periods in apple. *Frontiers in Plant Science* **8**, 1923, doi:10.3389/fpls.2017.01923 (2017).

76 Lipka, A. E. *et al.* GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **28**, 2397-2399, doi:10.1093/bioinformatics/bts444 (2012).

77 Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635, doi:10.1093/bioinformatics/btm308 (2007).

78 Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**, 825-830, doi:10.1038/ng.2314 (2012).

79 Kumar, S. *et al.* Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. *G3 Genes|Genomes|Genetics* **5**, 2711-2718, doi:10.1534/g3.115.021105 (2015).

Table 1: An overview of genome-wide association studies in apple

Publication	Year	Plant material	SNP genotyping	Traits	Model or software	Number of genotypes	Score*
Kumar et al.[1]	2013	progeny	8K (8K array)	fruit quality	MLM in GAPIT	1,200	6/6
Kumar et al.[2]	2014	progeny	4K (8K array)	fruit quality	MLM in GAPIT	115	2/8
Lozano et al.[3]	2014	progeny	0.3K (SNP assay)	fruit quality	MLM in GAPIT	94	0/7
Migicovsky et al.[4]	2016	accessions	8K (GBS)	floral, phenology, fruit and vegetative	EMMAX	689	4/36
Amyotte et al.[5]	2017	accessions	52K (GBS)	fruit quality	MLM and GLM in TASSEL	85	7/22
Duan et al.[6]	2017	accessions	7,218K (deep sequencing)	fruit quality	LMM in FaST-LMM	75	1/1
Lee et al.[7]	2017	accessions	13K (GBS)	fruit quality	GAPIT	237	2/6
Moriya et al.[8]	2017	accessions	16K (20K array)	fruit quality	rrBLUP	82	1/1
Moriya et al.[9]	2017	accessions	(20K array)	fruit quality	MLMM	160	1/1
Urrestarazu et al.[10]	2017	accessions	275K (480K array)	phenology	MLMM	1,168	2/2
McClure et al.[11]	2018	accessions	55K (GBS)	fruit quality, apple scab	WeightedMLM in TASSEL, CARAT	172	4/10
Larsen et al.[12]	2019	accessions	16K (GBS)	fruit quality, volatiles, phenology	MLM and GLM in TASSEL	110–177	10/107
McClure et al.[13]	2019	accessions	99K (GBS)	polyphenols	MLMM	136	9/20
Hu et al.[14]	2020	accessions	7,218K (deep sequencing)	fruit quality	EMMAX	61	1/1
Noh et al.[15]	2020	accessions	43K (GBS)	apple *Marssonina* blotch	mrMLM	192	1/1
Kunihisa et al.[16]	2021	progeny and accessions separately	12K (20K array)	fruit quality	rrBLUP	468	5/5
Migicovsky et al.[17]	2021	accessions	33K (GBS)	fruit quality	TASSEL, MLMM	520–596	1/3
Minamikawa et al.[18]	2021	progeny and accessions	12K (20K array)	productivity, fruit quality, phenology	BayesB	844	22/27

* number of traits with detected significant marker-trait associations / all analyzed traits

1 Kumar, S. *et al.* Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* **14**, 393, doi:10.1186/1471-2164-14-393 (2013).

2 Kumar, S., Raulier, P., Chagné, D. & Whitworth, C. Molecular-level and trait-level differentiation between the cultivated apple (*Malus* × *domestica* Borkh.) and its main progenitor Malus sieversii. *Plant Genetic Resources* **12**, 330-340, doi:10.1017/S1479262114000136 (2014).

3 Lozano, L. *et al.* Feasibility of genome-wide association analysis using a small single nucleotide polymorphism panel in an apple breeding population segregating for fruit skin color. *Journal of the American Society for Horticultural Science J. Amer. Soc. Hort. Sci.* **139**, 619, doi:10.21273/jashs.139.6.619 (2014).

4 Migicovsky, Z. *et al.* Genome to phenome mapping in apple using historical data. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0113 (2016).

5 Amyotte, B., Bowen, A. J., Banks, T., Rajcan, I. & Somers, D. J. Mapping the sensory perception of apple using descriptive sensory evaluation in a genome wide association study. *PLOS ONE* **12**, e0171710, doi:10.1371/journal.pone.0171710 (2017).

6 Duan, N. *et al.* Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications* **8**, 249, doi:10.1038/s41467-017-00336-7 (2017).

7 Lee, S. J. *et al.* Identification of potential gene-associated major traits using GBS-GWAS for Korean apple germplasm collections. *Plant Breeding* **136**, 977-986, doi:10.1111/pbr.12544 (2017).

8 Moriya, S. *et al.* Identification of QTLs for flesh mealiness in apple (*Malus* × *domestica* Borkh.). *The Horticulture Journal* **86**, 159-170, doi:10.2503/hortj.MI-156 (2017).

9 Moriya, S. *et al.* Allelic composition of MdMYB1 drives red skin color intensity in apple (*Malus* × *domestica* Borkh.) and its application to breeding. *Euphytica* **213**, 78, doi:10.1007/s10681-017-1864-x (2017).

10 Urrestarazu, J. *et al.* Genome-wide association mapping of flowering and ripening periods in apple. *Frontiers in Plant Science* **8**, 1923, doi:10.3389/fpls.2017.01923 (2017).

11 McClure, K. A. *et al.* A genome-wide association study of apple quality and scab resistance. *The Plant Genome* **11**, 170075, doi:10.3835/plantgenome2017.08.0075 (2018).

12 Larsen, B. *et al.* Genome-wide association studies in apple reveal loci for aroma volatiles, sugar composition, and harvest date. *The Plant Genome* **12**, 180104, doi:10.3835/plantgenome2018.12.0104 (2019).

13 McClure, K. A. *et al.* Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Horticulture Research* **6**, 107, doi:10.1038/s41438-019-0190-y (2019).

14 Hu, Y. *et al.* ERF4 affects fruit firmness through TPL4 by reducing ethylene production. *The Plant Journal* **103**, 937-950, doi:10.1111/tpj.14884 (2020).

15 Noh, J., Do, Y. S., Kim, G. H. & Choi, C. A genome-wide association study for the detection of genes related to apple *Marssonina* Blotch disease resistance in apples. *Scientia Horticulturae* **262**, 108986, doi:10.1016/j.scienta.2019.108986 (2020).

16 Kunihisa, M. *et al.* Genome-wide association study for apple flesh browning: detection, validation, and physiological roles of QTLs. *Tree Genetics & Genomes* **17**, 11, doi:10.1007/s11295-021-01492-0 (2021).

17 Migicovsky, Z. *et al.* Genomic consequences of apple improvement. *Horticulture Research* **8**, 9, doi:10.1038/s41438-020-00441-7 (2021).

18 Minamikawa, M. F. *et al.* Tracing founder haplotypes of Japanese apple varieties: Application in genomic prediction and genome-wide association study. *Horticulture Research* **8**, 49, doi:10.1038/s41438-021-00485-3 (2021).

Table 2: An overview of genomic prediction studies in apple

Publication	Year	Plant material	SNP genotyping	Traits	Model	Validation	Accuracy	Number of genotypes
Kumar et al.[1]	2012	7 full-sib progeny	8K (8K array)	fruit quality	RR-BLUP, Bayesian LASSO	cross-validation	high	1,120
Kumar et al.[2]	2015	17 full-sib progeny	3K (8K array)	fruit quality	G-BLUP	leave-one-progeny-out cross-validation	low to high	247
Muranty et al.[3]	2015	20 full-sib progeny	8K (20K array)	productivity, fruit quality	BayesCπ	5 full-sib progeny	low to moderate	977
Migicovsky et al.[4]	2016	accessions	8K (GBS)	floral, phenology, fruit, vegetative	RR-BLUP	cross-validation	low to moderate	689
McClure et al.[5]	2018	accessions	75K (GBS)	fruit quality, apple scab	RR-BLUP	cross-validation	low to moderate, high for apple scab	172
McClure et al.[6]	2019	accessions	99K (GBS)	fruit polyphenols	RR-BLUP	cross-validation	low to moderate	136
Roth et al.[7]	2020	accessions, 6 full-sib progeny	8K (20K array)	fruit texture	RR-BLUP	cross-validation, 6 full-sib progeny	low to high	537
Kumar et al.[8]	2020	accessions	6K (GBS)	fruit quality	G-BLUP	cross-validation	moderate to high	274
Minamikawa et al.[9]	2021	accessions, 16 full-sib progeny	12K (20K array)	productivity, fruit quality, phenology	BayesB	leave-one-progeny-out cross-validation	low to moderate	844

1 Kumar, S. *et al.* Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). *PLOS ONE* **7**, e36674, doi:10.1371/journal.pone.0036674 (2012).

2 Kumar, S. *et al.* Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. *G3 Genes|Genomes|Genetics* **5**, 2711-2718, doi:10.1534/g3.115.021105 (2015).

3 Muranty, H. *et al.* Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research* **2**, 15060, doi:10.1038/hortres.2015.60 (2015).

4 Migicovsky, Z. *et al.* Genome to phenome mapping in apple using historical data. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0113 (2016).

5 McClure, K. A. *et al.* A genome-wide association study of apple quality and scab resistance. *The Plant Genome* **11**, 170075, doi:10.3835/plantgenome2017.08.0075 (2018).

6 McClure, K. A. *et al.* Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Horticulture Research* **6**, 107, doi:10.1038/s41438-019-0190-y (2019).

7 Roth, M. *et al.* Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. *Horticulture Research* **7**, 148, doi:10.1038/s41438-020-00370-5 (2020).

8 Kumar, S., Hilario, E., Deng, C. H. & Molloy, C. Turbocharging introgression breeding of perennial fruit crops: A case study on apple. *Horticulture Research* **7**, 47, doi:10.1038/s41438-020-0270-z (2020).

9 Minamikawa, M. F. *et al.* Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. *Horticulture Research* **8**, 49, doi:10.1038/s41438-021-00485-3 (2021).

2 The apple REFPOP – a reference population for genomics-assisted breeding in apple

Jung, Michaela[1,2]; Roth, Morgane[2,3]; Aranzana, Maria José[4,5]; Auwerkerken, Annemarie[6]; Bink, Marco[7,8]; Denancé, Caroline[9]; Dujak, Christian[5]; Durel, Charles-Eric[9]; Font i Forcada, Carolina[4]; Cantin, Celia M.[4,10]; Guerra, Walter[11]; Howard, Nicholas P.[12,13]; Keller, Beat[1,2]; Lewandowski, Mariusz[14]; Ordidge, Matthew[15]; Rymenants, Marijn[6,16]; Sanin, Nadia[11]; Studer, Bruno[1]; Zurawicz, Edward[14]; Laurens, François[9]; Patocchi, Andrea[2]; Muranty, Hélène[9]

[1]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, 8092 Zurich, Switzerland;
[2]Breeding Research group, Agroscope, 8820 Wädenswil, Switzerland;
[3]Current address: GAFL, INRAE, 84140 Montfavet, France;
[4]IRTA (Institut de Recerca i Tecnologia Agroalimentàries), 08140 Caldes de Montbui, Barcelona, Spain;
[5]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193 Bellaterra, Barcelona, Spain;
[6]Better3fruit N.V., 3202 Rillaar, Belgium;
[7]Biometris, Wageningen University and Research, 6708 PB Wageningen, The Netherlands;
[8]Hendrix Genetics Research, Technology and Services B.V., PO Box 114 5830AC Boxmeer, The Netherlands;
[9]IRHS, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071 Beaucouzé, France;
[10]ARAID (Fundación Aragonesa para la Investigación y el Desarrollo), 50018 Zaragoza, Spain;
[11]Research Centre Laimburg, 39040 Auer, Italy;
[12]Department of Horticultural Science, University of Minnesota, St. Paul, MN 55108, USA;
[13]Institute of Biology and Environmental Sciences, University of Oldenburg, 26129 Oldenburg, Germany;
[14]Research Institute of Horticulture, 96-100 Skierniewice, Poland;
[15]University of Reading, School of Agriculture, Policy and Development, Whiteknights, RG6 6AR Reading, United Kingdom;
[16]Laboratory for Plant Genetics and Crop Improvement, KU Leuven, B-3001, Leuven, Belgium.

Material from: 'JUNG et al., THE APPLE REFPOP – A REFERENCE POPULATION FOR GENOMICS-ASSISTED BREEDING IN APPLE, HORTICULTURE RESEARCH, published 2020, Springer Nature, doi: https://doi.org/10.1038/s41438-020-00408-8'

Abstract

Breeding of apple is a long-term and costly process due to the time and space requirements for screening selection candidates. Genomics-assisted breeding utilizes genomic and phenotypic information to increase the selection efficiency in breeding programs, and measurements of phenotypes in different environments can facilitate the application of the approach under various climatic conditions. Here we present an apple reference population: the apple REFPOP, a large collection formed of 534 genotypes planted in six European countries, as a unique tool to accelerate apple breeding. The population consisted of 269 accessions and 265 progeny from 27 parental combinations, representing the diversity in cultivated apple and current European breeding material, respectively. A high-density genome-wide dataset of 303,239 SNPs was produced as a combined output of two SNP arrays of different densities using marker imputation with an imputation accuracy of 0.95. Based on the genotypic data, linkage disequilibrium was low and population structure was weak. Two well-studied phenological traits of horticultural importance were measured. We found marker-trait associations in several previously identified genomic regions and maximum predictive abilities of 0.57 and 0.75 for floral emergence and harvest date, respectively. With decreasing SNP density, the detection of significant marker-trait associations varied depending on trait architecture. Regardless of the trait, 10,000 SNPs sufficed to maximize genomic prediction ability. We confirm the suitability of the apple REFPOP design for genomics-assisted breeding, especially for breeding programs using related germplasm, and emphasize the advantages of a coordinated and multinational effort for customizing apple breeding methods in the genomics era.

2.1 Introduction

Apple (*Malus* × *domestica* Borkh.) is one of the most economically valuable fruit crops in temperate regions[1]. Thousands of cultivars are grown in national and private repositories around the globe. Extensive genetic variation described in the European apple germplasm illustrates the available genetic diversity among cultivars[2,3]. Only a fraction of the existing apple cultivars is commercially used. Although ongoing breeding programs worldwide aim to create new cultivars adapted to consumer demands and changing climate, these goals could be difficult to reach within the narrow elite genetic pool of modern breeding material[4].

Since the advent of genomics, genotyping tools have begun to produce affordable genome-wide marker data. Large datasets are being analyzed to explore genotype-phenotype relationships in genome-wide association studies (GWAS) and to allow genomic prediction. Particularly, genomic prediction[5] has revolutionized breeding and more than doubled genetic progress of major livestock such as cattle[6]. The method relies upon models fitted to broad datasets of genotypes and phenotypes from a training population. The aim is to predict the agronomic performance of breeding material related to the training set based on the marker information alone.

For the application of genomic prediction in fruit trees, apple became a model species due to its economic importance and the range of available research resources[7]. Genomic prediction in apple was tested for the first time by Kumar et al. in 2012[8]. However, this and further studies have been based on a limited number of genetic markers and/or been carried out at a local scale[9–14].

The prediction accuracy of genomic prediction models generally increases with the number of markers used to genotype the training population, reaching a plateau depending on the architecture of the trait, the number of individuals in the training population, the size of the genome and linkage disequilibrium[15]. Low population structure[2] and rapid linkage disequilibrium decay in highly diverse apple germplasm[3] underlie the need for dense SNP marker datasets for GWAS. Considerable progress has been made in the development of genomic resources in apple[16]. Creation of low and medium density SNP arrays, such as the Illumina Infinium® 20K SNP genotyping array (20K array)[17] was followed by the establishment of the Affymetrix Axiom® Apple 480K SNP genotyping array (480K array) with more than 480 thousand SNPs[18]. Using this SNP array for apple, markers significantly associated with phenological traits in several germplasm collections have been successfully discovered by GWAS[3]. Although the high marker density of the 480K array may lead to higher prediction accuracies than achieved before, no genomic prediction study was conducted using this array so far. Commercial apple breeding programs still cannot afford large-scale genotyping of their germplasm with expensive tools. Therefore, a balance between genotyping density, costs and predictive ability should be found for the application of genomic prediction in breeding.

Here, we present an apple reference population: the apple REFPOP. The population has been replicated across six environments in Europe and designed for comparing two management practices, which will allow for a thorough and unique study of the effects of genotype, environment and management as well as their interactions on apple phenotypes. The main objectives of this study were to consolidate the high-density SNP

marker dataset for all apple REFPOP genotypes, apply the SNP marker dataset when describing population characteristics of the apple REFPOP and prove suitability of the apple REPOP design for genomics-assisted breeding. Success of the genomics-assisted breeding may depend on characteristics such as marker density, trait architecture or size of the training population[10,15]. For the first time in apple, these aspects could be tested with (i) GWAS and genomic prediction using the high-density 480K array marker dataset, (ii) a comparison of the effects of SNP density on GWAS and genomic prediction and (iii) the prediction precision analysis. Our further aim was to discuss the use of the apple REFPOP for genomic prediction in multiple environments, across multiple traits and multi-management practices and to facilitate the improvement of apple breeding using the established apple REFPOP.

2.2 Materials and methods

2.2.1 Composition of apple REFPOP plant material

The apple REFPOP collection composed of accessions and progeny was designed and established by the collaborators of the FruitBreedomics project[7]. The accession group consisted of old and modern diploid accessions representing a wide range of genetic diversity in apple. Simple sequence repeat data obtained by Fernández-Fernández[37], Lassois et al.[21] and Urrestarazu et al.[2] were used to allocate unique genotype code (so-called MUNQ, for *Malus* UNiQue genotype code as described by Muranty et al.[19]) to the accessions included in these studies, which resulted in 1292 unique genotypes available for further choice. When possible, passport data of the accessions belonging to each unique genotype were used to identify its country or region of origin. To form the accession group, a subset of the unique genotypes was created. All possible unique genotypes were chosen when less than 15 genotypes were available per each country or region of origin. For origins represented by a larger number of accessions, priority was given to the genotypes already analyzed with the Axiom®Apple480K array[18]. Additional genotypes of each origin were chosen randomly so that the overall number of selected genotypes per origin was proportional to the number of genotypes of this origin in the whole collection of unique genotypes. Unique genotypes previously analyzed with the 480K array but of unknown origin were also included. Additionally, five accessions ('Red Winter', 'O53T136', 'Priscilla-NL', 'P7 R4A4' and X6398) considered as founders in the progeny group pedigree were added to the accession group. Most accessions were chosen from the apple germplasm collection of the National Fruit Collection, Brogdale, United Kingdom in order to simplify collection of the plant material. The budwood of the founders stemmed from two sources: France's National Research Institute for Agriculture, Food and Environment (INRAE), Angers, France and Wageningen University and Research (WUR), Wageningen, The Netherlands. Availability of budwood from trees affected the numbers of chosen accessions. Therefore, accessions often had to be either (i) replaced with a different accession of the same MUNQ or (ii) excluded in case no other accession corresponding to the unique genotype was available or could provide enough budwood. Additionally, the triploid accession 'Biesterfelder Renette' (MUNQ 1106.1) and the accession 'Karinable' (MUNQ 7828) with no available SNP data were excluded from the analysis, although planted in the orchards.

The progeny group of the apple REFPOP included 27 full-sib parental combinations previously used in the European project FruitBreedomics[7]. These full-sib parental combinations originated from eight different breeding programs and they were obtained from 32 parents while 13 of the parents were included in the accession group. For most parental combinations, genotypic information was available ahead of this study, with different genotyping density depending on the parental combination. Twenty-two parental combinations had been genotyped with the 20K array[7,17,38]. In addition, in the frame of a pilot study of genomic selection[9], three other parental combinations were genotyped with a custom 512 SNPs array using the TaqMan OpenArray technology covering the whole genome at a very low density. Finally, for two parental combinations, whole genome data were not available.

A subset of ten individuals per parental combination was chosen to form the progeny group of the REFPOP. For the 25 parental combinations for whom the whole genome marker data was available, ten individuals were chosen using a genetic distance sampling strategy[39]. In all cases, the ten individuals formed the center of each of the clusters defined by genetic distance. Where budwood was not available in sufficient numbers, they were either (i) replaced by individuals closest to the center of the same cluster or (ii) excluded from the progeny group leading to fewer than ten genotypes per parental combination being chosen (as in the case of parental combinations X6679 × X6808 and X6679 × X6417). From the parental combination 'Jonathan' × 'Prima', eleven progeny were included in the REFPOP. One progeny of the 'Dalinette' × X6681 parental combination (NOVADI/0830) was found as triploid using 20K genotypic data and thus excluded from the analysis, although planted in the orchards. For the two parental combinations without whole genome marker data available, individuals were chosen randomly.

2.2.2 Multiplication of plant material and planting design

In 2015, budwood from each apple genotype was collected and grafted onto 'M9' rootstocks. The grafting was performed in three different nurseries, i.e., at (i) INRAE Angers, France, (ii) Better3Fruit, Rillaar, Belgium and (iii) Consorzio Italiano Vivaisti, San Giuseppe, Italy. The following year, grafted trees were planted across six contrasting environments, each located in (i) Rillaar, Belgium, (ii) Angers, France, (iii) Laimburg, Italy, (iv) Skierniewice, Poland, (v) Lleida, Spain and (vi) Wädenswil, Switzerland. The environments ranged across several biogeographical regions in Europe, i.e., the Mediterranean in Spain, Atlantic in France and Belgium, Alpine in Italy and Continental biogeographical region in Switzerland and Poland[40]. Depending on the environment, planting distance within and between tree rows ranged from 0.9 to 1.3 m and 3.2 to 3.6 m, respectively. All orchards were divided into two parts: the first part was to be managed with the common agricultural practice of the country, the second was to receive alternative management conditions (e.g., low pesticide or water input). So far, both parts have been managed in the same way. The first part consisted of two randomized complete blocks, totaling together 1,068 trees, each block containing one replicate per genotype. The second part also consisted of two randomized complete blocks, each containing one representative of approx. one third of the genotypes (on average 184 genotypes). Every environment shared about half of these genotypes (92 on average) with one of the other countries (country pairs: Belgium – Italy, France – Switzerland, Poland – Spain). Additionally, each environment shared on average 34 genotypes with three further countries. The blocks within parts were used to ensure that two replicates of the same genotype were not planted at a close proximity to each other, but were not intended to block for environmental effects. The two parts were physically separated by a row of trees comprised of additional representatives of genotypes included in the apple REFPOP as well as other material. Since the alternative management regimes were not applied in the initial years, the trees in the barrier, as well as trees in what would later be the alternative management scenario may have contributed phenotypic values for traits in 2018. Three control genotypes 'Gala', 'Golden Delicious' and Modì® 'CIVG198' were each replicated 48 and 18 times in the first and the second part of the orchard, respectively.

2.2.3 Molecular marker genotyping

A SNP marker dataset for the apple REFPOP was produced using two overlapping SNP arrays of different resolution. The genotypic data generated with the 480K array for a total of 1,356 unique genotypes were retrieved from previous studies[3,19]. The SNP marker dataset included all apple REFPOP accessions (including 13 parents of the apple REFPOP progeny group) and additional unique genotypes later used as a reference set for marker imputation. The applied filtering strategy differed from the original one described by Bianco et al.[18] in discarding the quality prediction based on metrics of the SNP clusters and making use of the pedigree, which was reconstructed in a recent study using all genotyped diploid unique genotypes[19]. For more details about SNP filtering, see Supplementary Methods 1. The SNP positions consistent with the apple reference genome based on the doubled haploid GDDH13 v1.1[26] were used. Markers unassigned to the 17 apple chromosomes were excluded, resulting in a dataset of 303,239 biallelic SNPs.

Genomic data for the apple REFPOP progeny group were generated using the 20K array. For 210 progeny from 22 parental combinations, the data were already available[7,17,38]. The 49 remaining progeny from parental combinations X338 × 'Braeburn', ACW 11303 × ACW 18522, ACW 13652 × ACW 11567, 'Dalinette' × X6681 and X6398 × 'Pinova' were genotyped with the 20K array within the framework of this study. All 20K array SNPs were filtered and the allele data for the remaining SNPs were curated to ensure the data made logical marker inheritance and co-segregation patterns following the methods and principles described by Vanderzande et al.[41] The set of 20K-array-generated markers was aligned to the 480K array marker set and 7060 of the 20K array SNPs were retained for further analysis.

2.2.4 Marker imputation

Due to the difference in marker resolution of the SNP arrays that were used to generate the SNP marker datasets, marker imputation was performed to provide high-density SNP marker information across the whole apple REFPOP. Genotypes included in the apple REFPOP and additional genotypes involved only in the imputation were used either as (i) reference for the imputation or (ii) imputation set or (iii) validation set (Table 1). First, a reference set of 480K array data was formed from 1,356 accessions and six progeny of the cross 'Fuji' × 'Pinova' that had been previously used for validation of the 480K array and the analysis of apple pedigrees[18,19]. Second, the imputation set was formed of the remaining 259 progeny from 27 parental combinations (i.e., including the remaining four progeny from the parental combination 'Fuji' × 'Pinova' that were not included in the validation set). Third, 40 additional progeny of the parental combination 'Fuji' × 'Pinova' and 46 progeny of 'Golden Delicious' × 'Renetta Grigia di Torriana' which had all been genotyped using the 480K array[18], but none of which had been chosen for inclusion in the apple REFPOP, were designated as a validation set.

The imputation was performed with the localized haplotype clustering implemented in the software Beagle 4.0 using pedigree information[42]. In the first step, reference genotypes (Table 1, reference set of data) along with the recently inferred pedigrees[19] were supplied to the program for the inference of haplotype phase and minor marker imputation in the reference set. To prepare data for the second step of imputation, SNP density of the validation set (480K array data, see also Table 1, validation set of data) was

decreased to the density of the imputation set (20K array data, see also Table 1, imputation set of data) in order to spike the imputation set with known samples of the validation set. In the second step of the imputation, phased reference genotypes along with the pedigrees were used to impute missing marker values in the extended imputation set (i.e., both imputation set and reduced-density validation set samples). Imputation accuracy in the validation set was then evaluated by computing the Pearson correlation between the imputed and original high-density SNP genotypic values in the validation set controls.

Table 1: Overview of the unique genotypes used in the genotype imputation.

Number of unique genotypes	Set of data	Population type	SNP array resolution
269	Reference	Accessions, apple REFPOP	480K
1,087	Reference	Accessions, additional material	480K
6*	Reference	1 parental combination, apple REFPOP	480K
259	Imputation	27 parental combinations, apple REFPOP	20K
86**	Validation	2 parental combinations, additional material	480K

* 6 progeny of 'Fuji' × 'Pinova'
** 40 progeny of 'Fuji' × 'Pinova' and 46 progeny of 'Golden Delicious' × 'Renetta Grigia di Torriana'

2.2.5 Genomic data analyses

Linkage disequilibrium

Linkage disequilibrium statistics were calculated as a square of the correlation coefficient (r^2) between pairs of SNPs on each chromosome with the R package snpStats[43]. To reduce computational time, SNPs of each chromosome were sampled randomly to include one tenth of the markers per chromosome in the r^2 calculation, which resulted in 4.6×10^8 marker combinations. A loess smoother ($\alpha = 0.5$) was fitted to 100,000 randomly chosen r^2 values across the whole span of chromosomes. Additionally, the loess smoother ($\alpha = 0.5$) was calculated for all obtained r^2 values for pairs of SNPs within a 5 kb distance. To determine the distance between SNPs at which the r^2 dropped below 0.2, average r^2 was calculated at 100 kb, 5 kb, 1kb and 0.1 kb (100 bp) as a mean of all r^2 within a window of 100 bp around each of the values.

Population structure

The neighbor-joining method as implemented in the R package ape[44] was used to estimate and visualize an unrooted neighbor-joining tree. Principal component analysis (PCA) with supplementary individuals was performed with the R package FactoMineR[45]. Tenfold cross-validation for the number of populations $K = \{1, 2, \ldots, 20\}$ was performed with ADMIXTURE 1.3, a program for estimating ancestry in unrelated individuals[46]. ADMIXTURE was used with default settings and a subset of markers filtered according to the program's manual. To avoid spurious effects of high linkage disequilibrium between adjacent markers, SNPs were removed in sliding windows of 50 SNPs advanced by 10 SNPs when squared correlation was greater than 0.1 for pairs of variants, leading to a subset of 12,374 SNPs. The correct K value was identified at the lowest cross-validation error. As structure is best estimated among unrelated (or weakly related) individuals, the neighbor-joining tree and ADMIXTURE were estimated for the accession group only. For

the population structure analyses, the European accessions were divided into several broad regions of origin to compensate for the uncertainty around the exact origin of old varieties[2].

2.2.6 Phenotype scoring

Two phenotypic traits were evaluated at the six plantation sites in 2018. Floral emergence of each tree was recorded as the date when 10% of flowers opened[47]. Harvest date was measured as the date when more than 50% of the fruits reached full physiological maturity, as determined by iodine coloration or expert knowledge[47]. Both traits were evaluated for each of the replicate trees individually. Deviation from the phenotyping protocol for harvest date led to exclusion of harvest date recorded in Poland. After measurements, dates were converted to counts of days starting at the beginning of the year in which they were measured.

2.2.7 Phenotypic data analyses

Raw phenotypic values were corrected for spatial heterogeneity individually within environments to obtain the adjusted phenotypic values of each tree. The corrected as well as uncorrected (raw) phenotypic values were used to estimate the individual-location clonal mean heritability. The adjusted phenotypic values of each tree were further used to fit a mixed model including the effects of genotype, environment and their interaction. The variance of each effect was calculated from the fitted mixed model to estimate multi-location clonal mean heritability and the fraction of phenotypic variation associated with the effects. Finally, the adjusted phenotypic values of each tree were used to obtain phenotypic least square means of genotypes across environments, i.e., a single mean phenotypic value for each genotype.

Correction of spatial heterogeneity

To account for spatial variation of the complete block design, e.g., due to different soil composition or water availability in the orchards, and to predict adjusted phenotypic values of each tree, spatial heterogeneity in the phenotypic data was modeled separately for each environment and trait using the spatial analysis of field trials with splines (SpATS)[20]. To specify the smooth component, a two-dimensional penalized tensor-product of marginal B-spline basis functions based on the P-spline ANOVA approach (PS-ANOVA) was defined with the default settings as a function of covariates plantation row and column, further denoted as *f(u, v)*. The following linear mixed model was fitted

$$\boldsymbol{y} = f(\boldsymbol{u}, \boldsymbol{v}) + \boldsymbol{Z}_g \boldsymbol{c}_g + \boldsymbol{Z}_r \boldsymbol{c}_r + \boldsymbol{Z}_c \boldsymbol{c}_c + \boldsymbol{\varepsilon} \quad (1)$$

with y being the vector of phenotypic values measured for each tree, u and v denoting the numeric positions, i.e., rows and columns, the vectors c_g, c_r and c_c being the random effect coefficients for the genotypes, rows and columns (as factors) associated with the design matrices Z_g, Z_r and Z_c, respectively, the ε denoting the random error vector. Spatial independence assumption of the error vector was visually assessed using a residuals' spatial plot for each model fit as described by Rodríguez-Álvarez et al.[20] To characterize the importance of model components, effective dimensions associated with each random factor (c_g, c_r and c_c), the PS-ANOVA spatial trend ($f_v(v)$, $f_u(u)$, $uh_v(v)$, $vh_u(u)$ and $f_{u,v}(u, v)$) and the total effective dimension ED_s (sum of partial effective dimensions associated with each component of the PS-ANOVA spatial trend) were assessed[20]. Predicted values of

genotypes adjusted for spatial heterogeneity within each environment were produced with random model terms of the smooth component and the random row and column effects excluded from the predictions. These values were used to visualize variability between traits and sites. Residuals for each tree (i.e., each replicate of a genotype) were extracted from the model fit and summed with the corresponding predicted values of genotypes adjusted for spatial heterogeneity within each environment to obtain adjusted phenotypic values of each tree.

Broad-sense heritability and phenotypic variation

To evaluate the efficiency of the spatial correction method when comparing raw data with data adjusted for spatial heterogeneity, individual-location clonal mean heritability H^2 was estimated for each of the traits and environments before and after the spatial correction. A random-effects model was fitted for each environment via restricted maximum likelihood (R package lmer[48])

$$y_{ik} = \mu + g_i + \varepsilon_{ik} \quad (2)$$

where y_{ik} was the k-th phenotypic value from genotype i (adjusted and non-adjusted phenotypic values of each tree), μ was the grand mean, g_i was the random effect of the i-th genotype and ε_{ik} was the error term. Individual-location clonal mean heritability was calculated from variance components of the model as total genotypic variance σ_g^2 over the phenotypic variance σ_p^2

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} \quad (3)$$

where the phenotypic variance was calculated from the genotypic variance, error variance σ_ε^2 and the mean number of replications $\bar{n}_r$.

$$\sigma_p^2 = \sigma_g^2 + \frac{\sigma_\varepsilon^2}{\bar{n}_r} \quad (4)$$

The individual-location clonal mean heritability was used to eliminate one trial with the heritability value below 0.1.

For the remaining environments, multi-location clonal mean heritability was estimated for each trait with a pooled analysis across environments using mixed-effects models fitted via restricted maximum likelihood (R package lmer[48])

$$y_{ijk} = \mu + g_i + l_j + gl_{ij} + \varepsilon_{ijk} \quad (5)$$

where y_{ijk} was the k-th adjusted phenotypic value of each tree from genotype i in environment j, μ was the grand mean, g_i was the random effect of the i-th genotype, l_j was the fixed effect of the j-th environment, gl_{ij} the interaction effect between the i-th genotype and j-th environment (random) and ε_{ijk} was the error term. Then, multi-location clonal mean heritability was calculated using the equation (3) with the phenotypic variance estimated as

$$\sigma_p^2 = \sigma_g^2 + \frac{\sigma_{gl}^2}{n_l} + \frac{\sigma_\varepsilon^2}{n_l \bar{n}_r} \quad (6)$$

where σ^2_{gl} was the genotype by environment interaction variance and n_l the number of environments.

From the model fit according to the equation 5, variance of each random effect was calculated. The fraction of phenotypic variation associated with the fixed effect was estimated as the variance of the vector of values predicted from the model fit when all random effects were set to zero. To assess the precision of the random effects, confidence intervals for the variance components were estimated using profiling likelihood method.

Phenotypic least square means

Phenotypic least square mean of each genotype across environments was estimated from the adjusted phenotypic values of each tree corrected for spatial heterogeneity within the environments. First, a multiple linear regression model was fitted for each trait

$$y_{ijk} = \mu + g_i + l_j + \varepsilon_{ijk} \quad (7)$$

where y_{ijk} was the k-th adjusted phenotypic value of each tree from genotype i in environment j, μ was the grand mean, g_i was the effect of the i-th genotype, l_j was the effect of the j-th environment and ε_{ijk} was the error term. Second, phenotypic least square means (or LS-means) of genotypes across environments were calculated with the R package doBy[49]. The phenotypic least square means of genotypes across environments were used later for genome-wide association analyses and genomic prediction.

2.2.8 Genome-wide association and prediction analysis

Genome-wide association study

The multi-locus mixed model (MLMM) method[28], a stepwise mixed-model regression for mapping complex traits under population structure, was applied to perform genome-wide association studies for both apple REFPOP groups together. As a response variable, the phenotypic least square means of genotypes across environments were used. MLMM was used as implemented in the R package GAPIT 3.0[50] with default settings and SNPs with low minor allele frequency (0.05) removed. Marker-trait associations were found as significant for p-values falling below a Bonferroni-corrected significance threshold $\alpha^* = \alpha/m$ with $\alpha = 0.05$ and m representing the number of tested markers. The proportion of phenotypic variance explained by each SNP marker significantly associated to the phenotypic least square means of genotypes across environments was estimated as a coefficient of determination (r^2). The r^2 was estimated from a simple linear regression model fitted using the numeric marker values of a single SNP as predictor and phenotypic least square means of genotypes across environments as response.

Genomic prediction

The RR-BLUP model[51] was used for genomic prediction of breeding values. The model was defined as

$$\boldsymbol{y} = \boldsymbol{WGu} + \boldsymbol{\varepsilon} \quad (8)$$

with the $\boldsymbol{y}$ being the vector of phenotypic least square means of genotypes across environments, $\boldsymbol{W}$ the design matrix relating genotypes to $\boldsymbol{y}$, $\boldsymbol{G}$ the SNP matrix, $\boldsymbol{u} \sim N(0, \boldsymbol{I}\sigma^2_u)$ the vector of SNP marker effects and $\boldsymbol{\varepsilon}$ the vector of errors. Five-fold cross-

validation was applied with the model and data of the whole apple REFPOP, each run masking 20% of the genotypes as validation set. Marker effects (BLUPs) were generated from the remaining 80% of the genotypes to make predictions for the validation set. Predictive ability was estimated as an average value of Pearson correlation coefficient calculated between observed phenotypic least square means of genotypes across environments from each validation set and predictions for the same genotypes. The cross-validation was repeated 100 times, the five folds being chosen randomly without replacement before each of the repetitions.

Comparison of GWAS and genomic prediction performance under various SNP densities

Additionally to the analyses with the full set of 303,239 SNPs produced with the 480K array, different subsets of SNPs were used to perform the GWAS and genomic prediction to investigate the effect of feature selection approaches on model performance. SNPs in subsets were chosen according to three main feature selection strategies, (i) the SNP set from the 20K array (7,060 SNPs) available in the full SNP marker dataset, (ii) the SNP set thinned according to linkage disequilibrium (12,374 SNPs, for details see population structure analysis described above) and (iii) an unsupervised SNP choice. For this latter strategy, data subsets were built for densities of 500, 1,000, 5,000, 10,000, 50,000, 100,000, 150,000, 200,000 and 250,000 markers. As suggested by Bermingham et al.[52], the markers were chosen evenly spaced from a random starting point. Since the 480K array was designed to cover not only the genic regions, but also to reach a uniform coverage of non-genic regions[18], the order of the markers on chromosomes was used as a proxy for their physical distance when sampling. For each of the nine SNP densities, the SNP choice was repeated with ten different seeds resulting in 90 subsets. For these subsets, GWAS and genomic prediction were performed as described above, with the five-fold cross-validation of the genomic prediction being repeated 10 times for each subset. The number of significant associations detected by GWAS was determined as the number of regions containing SNPs with p-values falling below a Bonferroni-corrected significance threshold. Mean number of the significant associations in GWAS and mean prediction ability of the genomic prediction as well as their 95% confidence interval were calculated for each number of chosen markers.

Analysis of prediction precision

Expectations of the precision of genomic estimated breeding values were approximated based on the equation 9, i.e., the equation 5 in the original article by Elsen[53]. The approximation of the precision ($\hat{r}^2$) of genomic estimated breeding values was determined for different values of parameters influenced by the experimental design (i.e., species, population size and composition, SNP density, environments or trait architecture) as

$$\hat{r}^2 \cong \frac{Nh^2}{Nh^2+M(1-h^2)} \quad (9)$$

with M the number of loci in linkage disequilibrium with genes underlying the trait, N the population size and h^2 the heritability. Parameter values were chosen to encompass and extrapolate beyond the apple REFPOP design: N taking values of 10, 269, 534 and 1,000, M between 1 and 10^6 and h^2 being equal to 0.5 (moderate) or 0.8 (high). To interpret the

output, trait architecture was classified using M into oligogenic ($M \leq 10$), complex ($10 < M \leq 100$) and very complex ($100 < M \leq 1000$). Precision of genomic estimated breeding values was considered very high when equal or larger than 0.8. All statistical analyses and data formatting in this article were performed with R[54] and visualized with the R package ggplot2[55], unless stated otherwise.

2.3 Results

2.3.1 Composition of the apple REFPOP

The apple REFPOP was ultimately built with (i) 269 diploid accessions representing a wide range of genetic diversity in apple, originating from various geographic regions around the globe and (ii) 265 diploid progeny from 27 parental combinations from several European breeding programs (Supplementary Table 1 and 2). Similar to a previous study[2], 194 European accessions were classified according to their origin into (i) Northern and Eastern Europe incl. Russia and Baltic countries (NEE, number of accessions n = 28), (ii) Western and Central Europe (WCE, n = 134), (iii) Southern Europe with accessions from Spain, Italy and Portugal (SE, n = 22) and (iv) Southeastern Europe containing accessions from Romania, Northern Macedonia, Moldova, Bulgaria and Turkey (SEE, n = 10). The 69 non-European accessions originated from (i) Australia and New Zealand (ANZ, n = 8), (ii) Canada (CAN, n = 16), (iii) Japan (JPN, n = 9), (iv) United States of America (USA, n = 34) and (v) South Africa (ZAF, n = 2). Six accessions previously analyzed with the 480K array but of unknown origin were also included in the apple REFPOP.

2.3.2 Marker imputation and validation of the imputation output

Missing marker values in the reference set obtained from the 480K array (see Material and methods, Table 1) were inferred through a minor imputation step to obtain a full dataset of 303,239 SNPs covering the 17 apple chromosomes. This minor imputation step was performed for a collection of 1,356 accessions (including 269 apple REFPOP accessions) and six progeny genotyped in separate studies[18,19]. Subsequently, a major marker imputation applied to the 259 of the 265 progeny that were genotyped with the 20K array (the remaining 6 progeny with 480K array data available were included in the reference set) increased the marker density from 7,060 to 303,239 SNPs (97.7% of marker values imputed). Additionally, imputation accuracy of 0.96 and 0.94 when calculating Pearson correlations between imputed and original values across individuals and across markers, respectively, was estimated for 86 genotypes from two parental combinations, ‘Fuji’ × ‘Pinova’ and ‘Golden Delicious’ × ‘Renetta Grigia di Torriana’.

2.3.3 Linkage disequilibrium and population structure

From the complete set of genome-wide SNP data for both apple REFPOP accessions and progeny, rapid decay of linkage disequilibrium was found (Figure 1 (a)). The loess smoother fitted to r^2 values of SNPs within a 5 kb distance dropped below a threshold of 0.2 at distance of 2.52 kb, the curve being very flat (Figure 1 (b)). Average r^2 calculated at 100 kb, 5 kb, 1 kb and 100 bp was 0.14, 0.21, 0.21 and 0.24, respectively. Separate analysis of the accession and progeny group showed a similar pattern of linkage disequilibrium in both groups (Supplementary Figure 1 and 2).

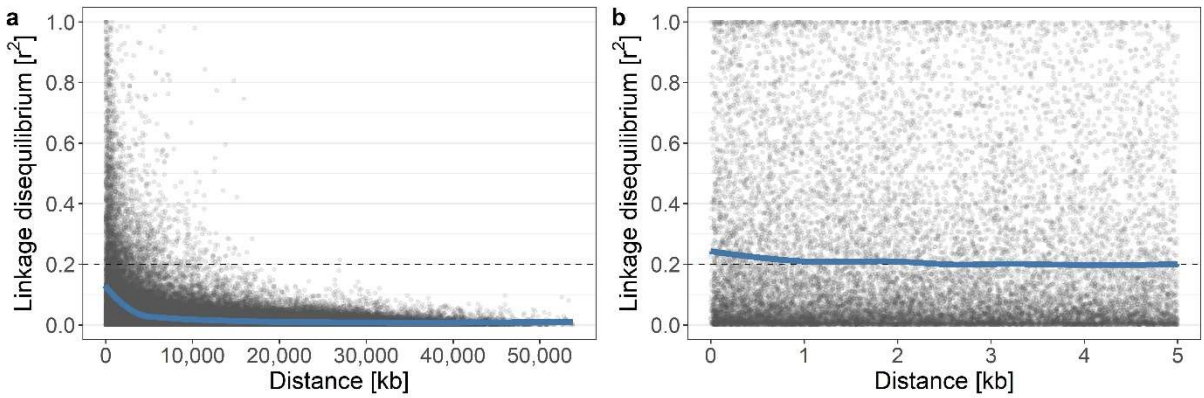


Figure 1: **a-b** Linkage disequilibrium with a loess smoother for (a) distances between SNPs across the span of chromosomes, (b) for SNPs within a 5 kb distance.

In an unrooted neighbor-joining tree of apple REFPOP accessions (Figure 2 (a)), the non-European accessions clustered at the upper side of the tree with a transition towards the European accessions clustering at the lower branches. In the principal component analysis (PCA) of the accession group (Figure 2 (b)), the first two principal components explained only 7.6% of the total variance in genetic markers. The first principal component showed a slight differentiation between European and non-European accessions with the majority of the non-European accessions positioned on the right side of the plot. The second principal component displayed a weak latitudinal cline in the European accessions with the southern and northern European accessions placed towards the opposite extremes of the second component. Progeny, added to the PCA as supplementary individuals after PCA loadings were estimated, were grouped tightly together among the accessions although they did not form any separate cluster (Figure 2 (b)). The ADMIXTURE analysis revealed that two local minima of the cross-validation error were reached at the number of clusters $K = 14$ and $K = 17$ (Supplementary Figure 3). For the first minimum ($K = 14$), genotypes sorted by cluster membership within groups defined by geographic region of origin appeared highly admixed (Figure 2 (c)). A PCA of the progeny group with 13 parents of the crosses included as supplementary individuals showed that the first two principal components described 10.4% of the total variance in the genomic data and the parents fell among the many small clusters formed by members of distinct parental combinations (Supplementary Figure 4). From all population structure analyses, we concluded that the apple REFPOP was composed of diverse germplasm with very weak population structure together with high levels of admixture.

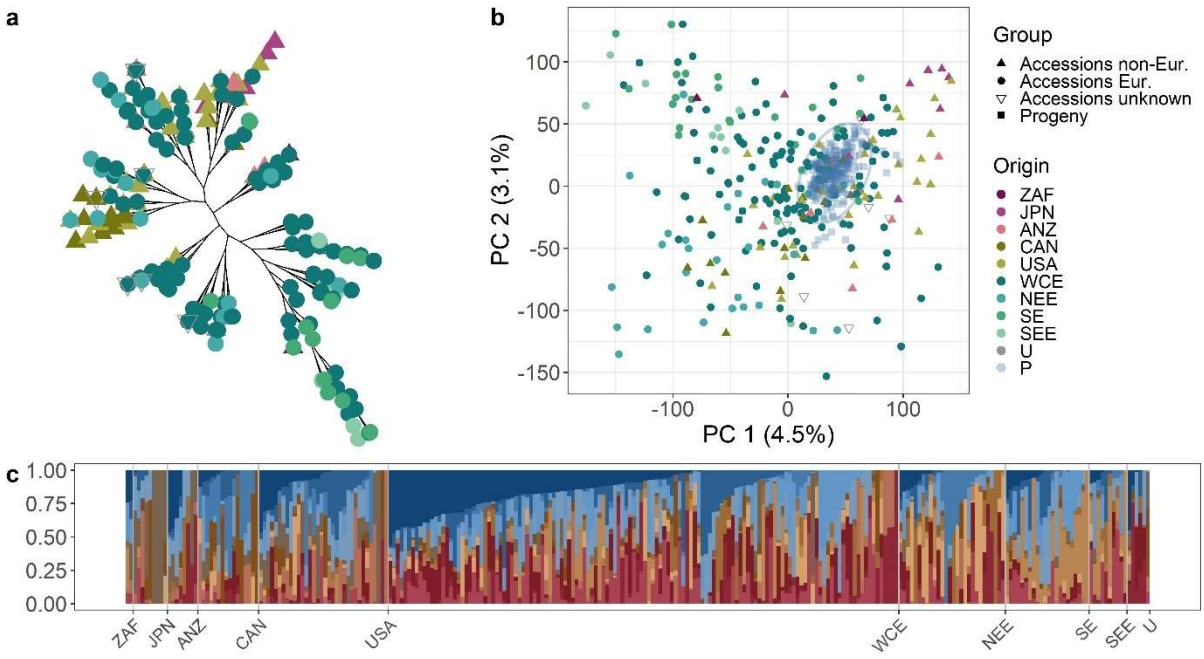


Figure 2: Structure of the apple reference population. **a** Unrooted neighbor-joining tree of the accession group, colors correspond to the legend in "b". **b** Principal component analysis of the accession group with progeny group as supplementary individuals encircled with a normal confidence ellipse (constructed using a multivariate normal distribution, level 0.95). Plot of the first two principal components with their respective proportion of the total variance shown within brackets. **c** ADMIXTURE bar plot of the accession group for $K = 14$. Labels in plots "a" to "c" refer to the geographic origin of genotypes: ZAF (South Africa), JPN (Japan), ANZ (Australia and New Zealand), CAN (Canada), USA (United States of America), WCE (Western and Central Europe), NEE (Northern and Eastern Europe), SE (Southern Europe), SEE (Southeastern Europe), U (accessions of unknown geographic origin) and P representing the progeny group in plot "b". In plot "c", each group of genotypes with a common geographic origin is labeled at its right side.

2.3.4 Phenotypic analyses

Using the spatial analysis of field trials with splines, spatial patterns were captured appropriately for all environments and traits as suggested by the homogeneous residuals (equation 1). Estimated effective dimensions of the spatial model, which are helpful for characterizing the importance of model components[20], showed a generally larger field variation due to spatial effects for floral emergence than for harvest date (Supplementary Table 3). Differences between environments for both floral emergence and harvest date were visualized using adjusted phenotypic values of each tree in every environment (Figure 3 (a, b)). Based on this adjusted phenotypic data from 2018, trees began flowering during a narrow period ranging between 16 days in Switzerland and 47 days in France. On average, the flowering began earlier at the southern sites (Spain and Italy). By contrast, harvest dates were distributed across a much longer time span of 99 to 150 days in Switzerland and Spain, respectively. As with the floral emergence, the mean harvest dates tended to be the earliest in the southern European sites. When comparing the adjusted data to the raw data, we found a small systematic increase in individual-location clonal mean heritability for all environments and both traits (Figure 3 (c)). Values of the individual-location clonal mean heritability after the adjustment were generally high for both traits and larger for harvest date than for floral emergence. Floral emergence data from Poland with individual-location clonal mean heritability below 0.1 were excluded. Using the adjusted phenotypic values of each tree from the remaining environments to

estimate the phenotypic least square means of genotypes across environments, the obtained values were distributed across a narrow period of 33 days for floral emergence whereas the values of harvest date were distributed over 97 days (Figure 3 (d)).

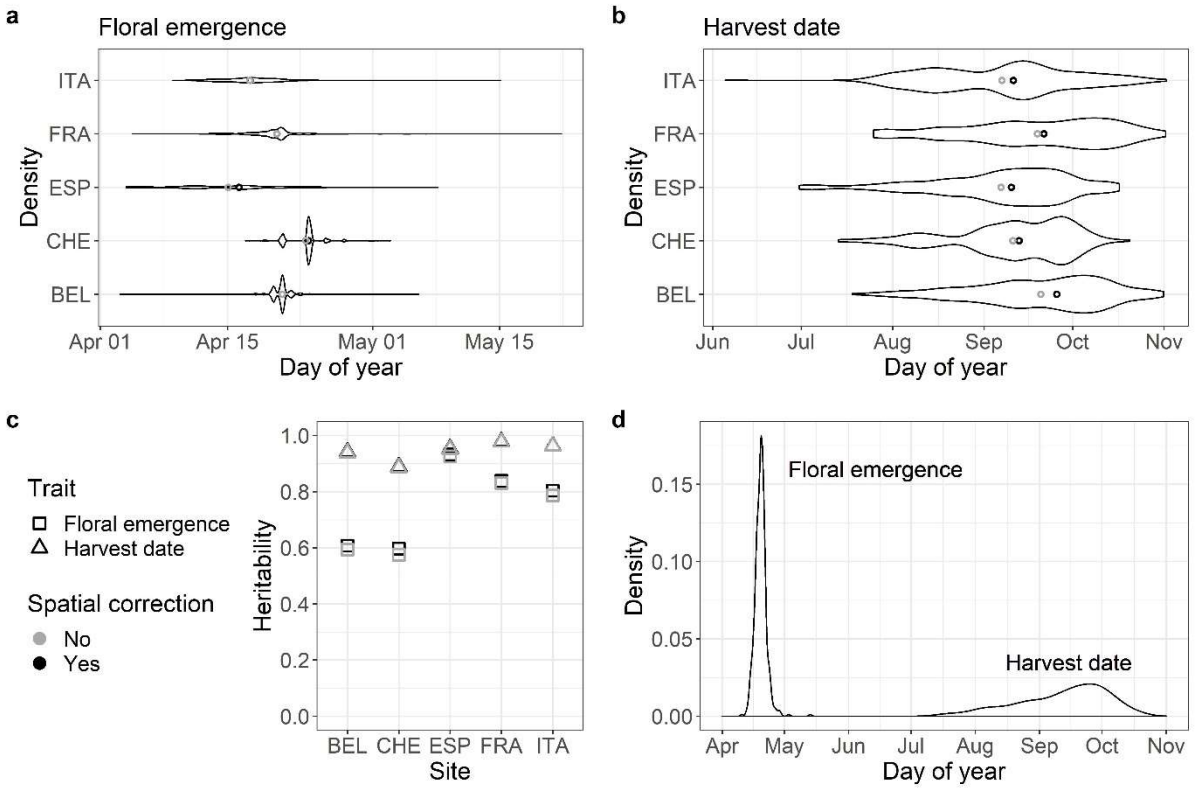


Figure 3: **a-b** Violin plots of (a) floral emergence and (b) harvest date for individual environments using the adjusted phenotypic values of each tree. Gray and black circles denote mean and median values, respectively. **c** Individual-location clonal mean heritability for two analyzed traits with values before and after the correction of spatial heterogeneity. **d** Density plot of phenotypic least square means of genotypes across environments with environmental effects removed, calculated from the adjusted phenotypic values of each tree corrected for spatial heterogeneity within environments. The environments were labeled with codes: Belgium (BEL), Switzerland (CHE), Spain (ESP), France (FRA) and Italy (ITA).

Predicted values of genotypes adjusted for spatial heterogeneity within each environment showed that individual genotypes appeared to respond differently to the various environments, with an apparent broader variation among genotypes for harvest date than for floral emergence (Figure 4 (a, b)). When modelling the effects of environment, genotype and their interaction (see Material and methods, equation 5), the proportion of variance in floral emergence explained by the environment, genotype and genotype by environment interaction was 43%, 22% and 18%, respectively (Figure 4 (c)). For harvest date, the environment explained only 5%, whilst genotype explained 74% and genotype by environment interaction explained 12% of the overall variance (Figure 4 (c)). Confidence intervals for the estimated variances of random effects of genotype, genotype by environment interaction and residuals were distinct from zero and therefore, the genotypes differed in the evaluated trait and interacted with the environment (Supplementary Figure 5 and 6).

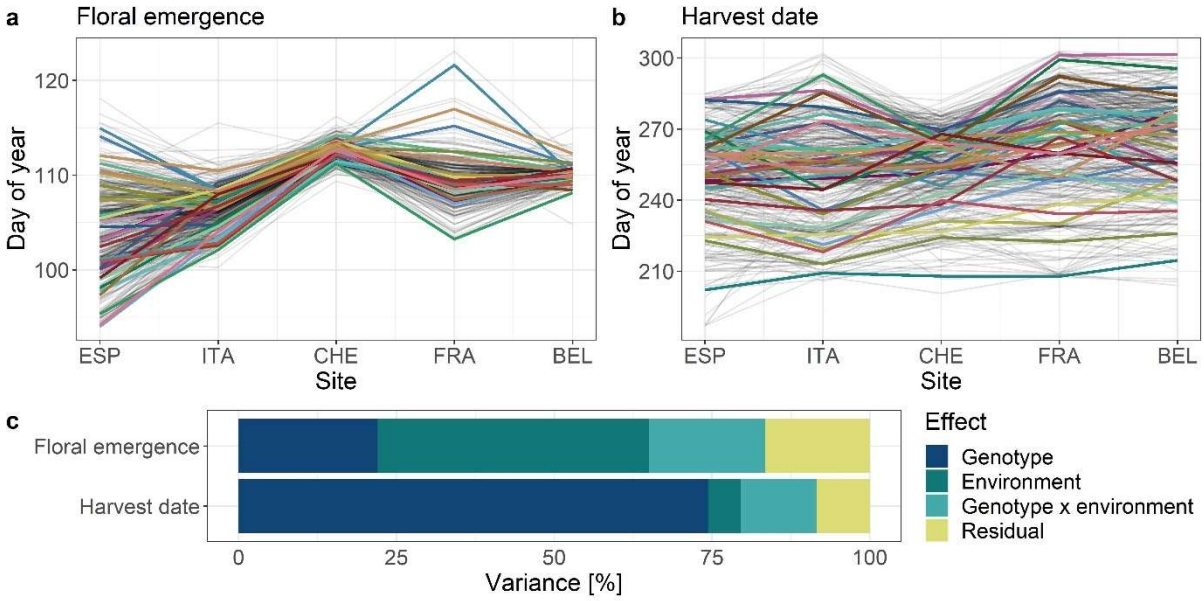


Figure 4: Visualization of variability between traits (floral emergence and harvest date), genotypes (both apple REFPOP groups) and environments. **a-b** Predicted values of genotypes adjusted for spatial heterogeneity within each environment, 30 randomly chosen genotypes were highlighted with colors. Order of the environments corresponds to their latitude. The environments were labeled with codes: Spain (ESP), Italy (ITA), Switzerland (CHE), France (FRA) and Belgium (BEL). **c** Stacked bar plots with the variance of the fixed effect of environment and the random effects of genotype, genotype by environment interaction and residuals; calculated from the model following the equation 5 (see Material and methods).

2.3.5 Genome-wide association and prediction analyses

The application of GWAS to apple REFPOP dataset identified three markers associated with floral emergence (Figure 5 (a), see Supplementary Table 4 for a list of p-values). Reported p-values were closely below the log-transformed Bonferroni-corrected significance threshold for GWAS performed with the full set of 480K array SNPs; all three associations were significant at density of 150,000 SNPs. Two SNPs were located at the top of chromosome 9 (proportion of explained phenotypic variance r^2 of 0.07 and 0.03) and a third one on chromosome 11 (r^2 = 0.10). Four SNPs were identified to be significantly associated with harvest date using the full set of 480K array SNPs (Figure 5 (b), see Supplementary Table 5 for a list of p-values). The strongest association was found on the chromosome 3 (r^2 = 0.39) with another significantly associated marker at a distance of 5.3 Mb ($r^2 \approx 0$). Two SNPs on chromosome 10 (r^2 = 0.15) and on chromosome 16 (r^2 = 0.11) were also significantly associated with harvest date. The QQ-plots for both traits indicated a good model fit (Supplementary Figure 7 and 8).

With the full set of 480K array SNPs, floral emergence was predicted using the RR-BLUP model with an average predictive ability of 0.57 whereas the multi-location clonal mean heritability was equal to 0.86 for this trait (Figure 5 (c)). Average predictive ability of harvest date was 0.75 with the multi-location clonal mean heritability of 0.97 for the trait.

Comparison of GWAS performance under various SNP densities showed that the number of significant associations was higher for harvest date than for floral emergence at all densities (Figure 5 (d)). For floral emergence, the number of significant associations increased up to density of 150,000 SNPs. Beyond this density, the number of significant associations decreased. The number remained at zero for the 12K marker set (12,374

SNPs) thinned for linkage disequilibrium. For harvest date, the number of significant associations was low (~1) for SNP densities of 0.5–1K. Across all densities larger than 1K, the number of significant associations with harvest date remained high (≥2). For the 12K marker set thinned for linkage disequilibrium, two significant associations on one chromosome were found. Similar to GWAS, predictive ability of genomic prediction was higher for harvest date than for floral emergence at all densities (Figure 5 (e)). For both traits, the plateau in predictive ability was reached at a density of 10K SNPs and the difference between predictive ability at 500 and 10,000 SNPs was ~0.1.

The precision analysis of genomic prediction (Figure 5 (f, g)) allowed the assessment of the trait architecture and heritability ranges that may lead to satisfying precision of genomic estimated breeding values. For a training population with a size of $N = 10$, very high precision values (precision ≥ 0.8) may be expected only for simple, oligogenic traits (number of markers underlying the trait $M \leq 10$). With a training population of the size of the apple REFPOP accession group ($N = 269$), more complex traits with $M = 100$ may be predicted with very high precision if the heritability was high ($h^2 = 0.8$), but with a decrease in precision for a moderate heritability ($h^2 = 0.5$). A training population size equal to that of the whole apple REFPOP ($N = 534$) may be sufficient to predict traits with $M = 100$ for both high and moderate heritability with very high precision. The precision analysis indicated that highly precise predictions for traits with very complex architecture ($100 < M \leq 1000$) may be possible if heritability was high and the training population had approximately twice the size of the apple REFPOP ($N = 1000$). However, the increase in estimated precision from $N = 534$ to $N = 1000$ was comparably smaller than the increase from $N = 269$ to $N = 534$.

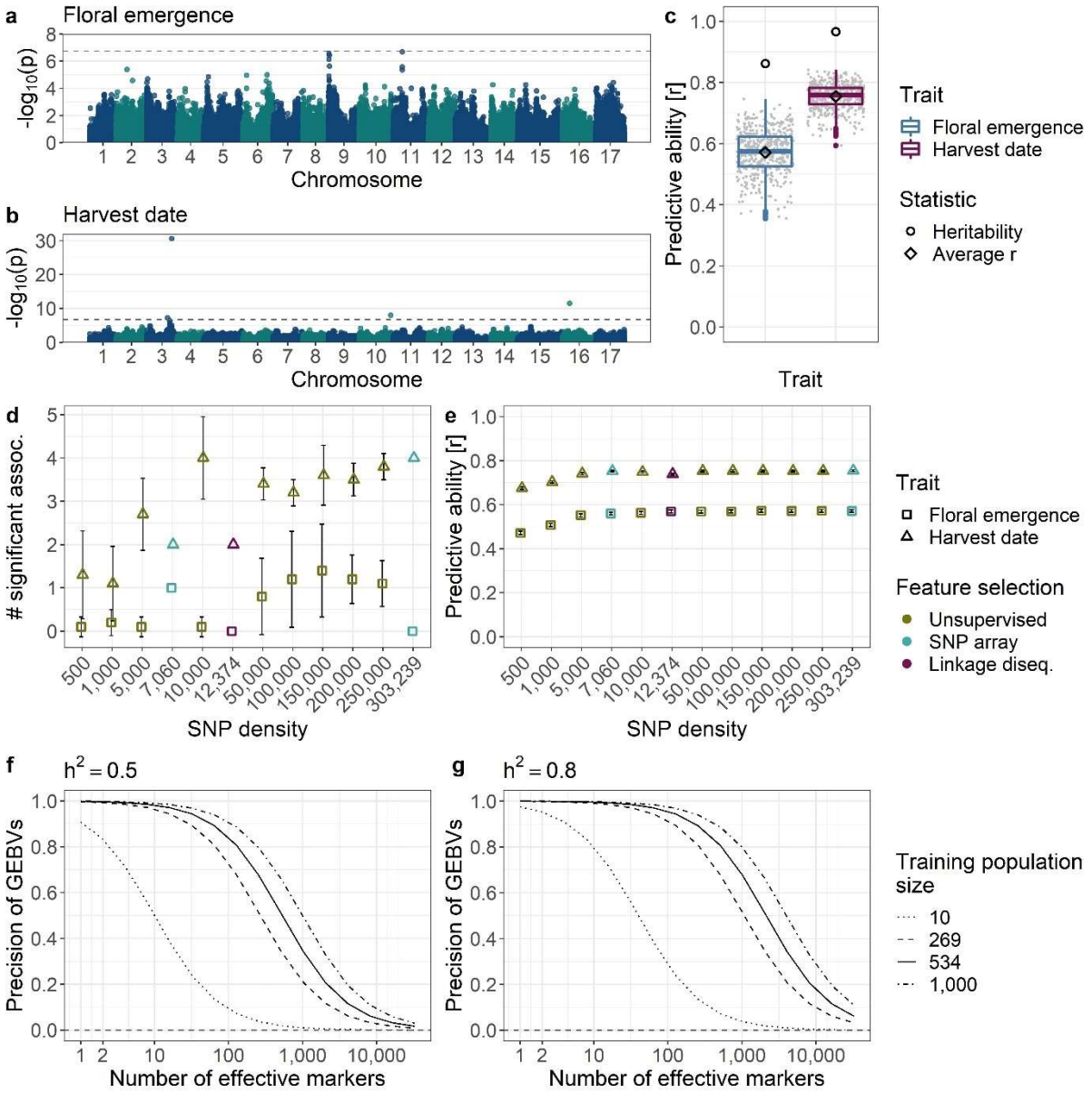


Figure 5: Results of the genome-wide association study (GWAS) and genomic prediction analysis. **a-b** Manhattan plots for (a) floral emergence and (b) harvest date with log-transformed p-values obtained with GWAS and Bonferroni-corrected significance threshold indicated with dashed line. **c** Genomic predictive ability measured with Pearson correlation coefficient, multi-location clonal mean heritability and average predictive ability. **d-e** Comparison of the number of significant associations in (d) GWAS and (e) genomic predictive ability measured with the respective mean values and their 95% confidence intervals under various SNP densities obtained through three feature selection strategies (see Methods). Plot "d" and "e" share a common legend placed in "e". **f-g** Expected precision of genomic estimated breeding values (GEBVs) with different training population sizes *N*, number of effective markers *M* in linkage disequilibrium with the genes underlying the trait and two heritability values h^2 of (f) 0.5 and (g) 0.8. The *N* values correspond to a minimum of 10 individuals, and to the sizes of the accession group (*N* = 269), the whole apple REFPOP (*N* = 534) and a population of an approx. double size of the apple REFPOP (*N* = 1000).

2.4 Discussion

2.4.1 Composition of the apple REFPOP

This study is the first report on the apple REFPOP, a reference population created to advance apple breeding with genomics-assisted methods. The population is divided into an accession and a progeny group. Genotypes of its accession group (all diploid) were chosen to represent a wide range of geographic origins and capitalize on the previously available high-density SNP array data. The progeny group stemmed from eight different European breeding programs to represent current tendencies in European apple breeding. For the progeny group, the choice of genotypes was mainly based on genetic distance sampling within progeny groups derived from a number of parental combinations. On the contrary, the choice of genotypes for the accession group did not correspond to a core collection design which generally aims at maximizing genetic diversity[21]. The incorporation of both cultivated accessions and modern breeding material is expected to promote the applicability of the apple REFPOP for genomics-assisted breeding beyond the scope of most current breeding programs, whose genetic bases are generally narrow[22]. Also, since relatedness between training and breeding populations is a crucial factor for successful genomic prediction[13], the European breeding programs can directly profit from the outcomes of the project because their breeding material has been included in the apple REFPOP. Furthermore, due to the high admixture of the accessions, which stem from across the globe, the levels of diversity in the apple REFPOP appear to be adequate to predict the performance of a broad spectrum of novel breeding material.

2.4.2 Expanding genomic information via imputation

Lately, resources for high-density marker genotyping in apple became available with the development of the 480K array[18], but using this array remained costly. Here, we combined the high-density marker information of the accessions with the recently inferred pedigrees for numerous cultivars[19] to perform a large-scale marker imputation of the progeny group genotyped with the 20K array[17]. Although marker imputation has been routinely applied at large scales in well-studied organisms such as cattle before[23], here we report for the first time a large scale, high density imputation in apple. The imputation has multiplied the amount of marker information by ~43x for the progeny group at relatively low cost. A high imputation accuracy was achieved, which is comparable to a similar study in poplar[24]. To evaluate the imputation accuracy, two parental combinations 'Fuji' × 'Pinova' and 'Golden Delicious' × 'Renetta Grigia di Torriana' consisting together of 86 progeny with known high density marker data were used. The parents 'Fuji' and 'Pinova' are present in the pedigree of five and three parental combinations of the imputation set, respectively. The parent 'Golden Delicious' can be found in all of the pedigrees from the imputation set, and therefore, relates to the whole progeny group of the apple REFPOP. Although the pedigree of 'Renetta Grigia di Torriana' is unknown, both parental combinations 'Fuji' × 'Pinova' and 'Golden Delicious' × 'Renetta Grigia di Torriana' are strongly related to the imputation set and therefore may provide a useful estimate of the imputation accuracy in other imputed parental combinations. An attempt to improve of the imputation accuracy might include expanding the reference set with all 19 parents of

the parental combinations as well as some of their ancestors for which high-density marker genotype data are not yet available.

2.4.3 Insights into apple population genomics

Earlier studies with lower density of genome-wide SNPs found that their number of markers was insufficient for GWAS because of a rapid decay in linkage disequilibrium in apple[10,25]. In concordance with previous findings, we found a strong linkage disequilibrium decay over short distances in the apple REFPOP dataset, the pattern of decay being very similar to that described by Urrestarazu et al.[3] The linkage disequilibrium decay was observed in the apple REFPOP from a distance as short as ~2.5 kb. Given the genome size of a doubled haploid derivative of 'Golden Delicious', estimated to 651 Mb[26], and the number of markers in this study, our marker resolution corresponded to one marker per ~2.1 kb on average. Based on this value, each group of loci in linkage disequilibrium should be on average represented by one marker. Thus, the marker coverage we report seems appropriate for genome-wide analyses, which is supported by our GWAS and prediction analyses. Due to the heterogeneity in the marker density over the genome, the statistical power could be reduced in specific regions with low SNP density.

Partial differentiation between European and non-European accessions in our material is consistent with significant differences between Old World and New World varieties found by Migicovsky et al.[10] The weak latitudinal cline in the European accessions may reflect the population structure reported in this germplasm by Urrestarazu et al.[2] The overall weak population structure in the apple REFPOP accession group, which was similar to a previously described weak genetic structure in apples of European origin, presumably reflected the generally highly admixed apple germplasm that is a result of the prominent gene flow characteristic of the cultivated apple gene pool[2,27].

2.4.4 An efficient design for genomics-assisted breeding

We applied GWAS using the multi-locus mixed-model (MLMM) method[28], which accounts for potential confounding effects of kinship and population structure. Thus, this method allows to combine highly related plant material, like in a pedigree-based-analysis, and more diversified, unrelated individuals to integrate different levels of linkage disequilibrium in the same analysis, hence maximizing the chances of finding regions associated to the target trait. By performing GWAS on the apple REFPOP dataset, SNPs associated with phenotypic variation could be identified and assigned to previously known genomic regions. A genomic region located on chromosome 9 was significantly associated with the trait floral emergence. Quantitative trait loci (QTL) at close locations were previously identified for floral bud break, floral emergence and flowering period[3,29,30]. The second SNP association with floral emergence was located on chromosome 11 approximately 1 Mb downstream from a SNP identified by Urrestarazu et al.[3] This association may be related to a minor QTL on chromosome 11 discovered using best linear unbiased predictors for genotype by year interaction effect of one season for the mapping of QTLs associated to bud break date (a trait highly correlated to floral emergence) in the study of Allard et al.[30] For harvest date, one significantly associated SNP on chromosome 3 was located 14,610 bp upstream from the transcription factor NAC18.1 listed as gene MD03G1222600 on the GDDH13 v1.1 genome[26]. Other studies

have identified associations between NAC18.1 and harvest date, and the gene is a known member of a family of conserved transcriptional regulators involved in ripening[3,10]. Another SNP on chromosome 3 associated with harvest date was located approximately 1 Mb upstream from a marker reported before by Urrestarazu et al.[3] The remaining two SNP associations were found on chromosomes 10 and 16 where QTL for harvest date have also been discovered before[3,31]. The identification of SNPs associated with both phenology traits in well-characterized genomic regions and for one possible minor QTL indicates the suitability of the apple REFPOP for discovering other alleles with large and small effects on trait variability in apple. Although novel marker-trait associations could not be identified and their number was lower than in Urrestarazu et al.[3] who also used genotypic data of the 480K array, the number of reported marker-trait associations in this work was higher than in GWAS studies using lower SNP densities obtained by genotyping by sequencing[10,32].

Applying a genomic prediction model with cross-validation, we were able to predict both phenology traits with moderate to high predictive ability when compared to the predictions for different apple traits reported previously[8–11,13], although cross-validation may have inflated predictive ability compared to a potential independent validation with a test set[33]. To our knowledge, floral emergence has not been predicted in apple before; in this work, an average predictive ability of 0.57 was reached for this trait. The predictions for harvest date had an average predictive ability of 0.75, which was higher than any previously reached accuracies of this, or equivalent, traits in apple[10,11]. The presented genomic prediction methodology may be directly applied for the breeding of floral emergence and harvest date. In particular, breeding programs using germplasm related to the apple REFPOP may capitalize on this work.

The high SNP density in this study allowed for powerful GWAS and genomic prediction analyses, with overall lower performance for floral emergence than harvest date. Floral emergence appeared under weak genetic control with the majority of phenotypic variance explained by the effects of environment and genotype by environment interaction. Markers significantly associated with the trait explained a low proportion of the phenotypic variance in our study, pointing to a complex genetic architecture of floral emergence with many influential genomic regions yet uncovered. With SNP density increasing up to 150,000 SNPs, increasing amount of the phenotypic variance of floral emergence can be explained with GWAS. Environment-specific GWAS or GWAS with phenotypes from across several seasons and locations may allow for improved GWAS performance in floral emergence. For harvest date, a trait under strong genetic control with a large proportion of the phenotypic variance explained by a few major genomic regions, our results suggest that SNP density as low as 10,000 markers may be required to discover genomic associations with the trait. The 10,000 marker subset was also sufficient to reach a plateau of the predictive ability in genomic prediction of both traits, with the difference in predictive ability between traits possibly attributable to the trait architecture. As previously shown in American cranberry[15], the SNP density at which the plateau of predictive ability is reached may be specific to the apple REFPOP independently of the traits, but may differ for other populations of the same species. These results can have a practical impact on apple breeding, where cost-effective genotyping of 10K SNPs may be sufficient for precise genomic prediction. In breeding research, the same genotyping coverage may be adequate to perform GWAS of oligogenic traits. For complex

traits under strong environmental control such as floral emergence, high-density SNP marker datasets remain desirable in GWAS. Future modelling including the effects of environment and genotype by environment interaction may contribute to a higher precision of genomic prediction in complex traits.

In our study, the size of the SNP marker dataset seemed to affect the genomic prediction ability stronger than the feature selection method. On the contrary, a higher power of GWAS to detect marker-trait associations can be obtained with a smaller SNP marker dataset than with a larger set of SNPs depending on the representation of genes underlying the traits in the SNP marker dataset. This way, more marker-trait associations could be revealed with the 7,060 SNPs of the 20K array than using SNPs pruned for linkage disequilibrium and such pruning should be avoided prior to GWAS.

In the light of prediction precision analysis, the lower predictive ability for floral emergence may be explained by a higher complexity of this trait, together with lower heritability. Importantly, the results were very likely impacted by the low phenotypic variability of floral emergence in the season of 2018. Additional phenotyping seasons will likely contribute to a better representation of the flowering variability in the population.

2.4.5 Prospect of the apple REFPOP for multi-environment, multi-management and multi-trait testing

The performance of breeding material in tested as well as untested but similar environments can be predicted accurately using genomic prediction models taking into account genotype by environment interactions[34]. We found a moderate but noticeable effect of the genotype by environment interaction on both phenology traits evaluated in 2018, which contrasts with the limited effect of genotype by environment interaction on a trait similar to harvest date reported in sweet cherry[34]. The replication of the apple REFPOP across six environments will enable the inclusion of these interactions into genomic prediction models. Furthermore, GWAS across separate environments can be performed in the future to identify environment-specific associations and evaluate the stability of associations across environments.

The apple REFPOP was also designed for comparing different management practices: one part of every orchard was grown under the conventional practice of each region, to evaluate the response of the germplasm to environmental effects; the second part can be managed in order to evaluate response to managements such as reduction of irrigation or pesticide application. However, the second management practice has not been applied so far to allow the trees to mature. Incorporating genotype by management and genotype by environment by management interactions in prediction models may help select new material adapted to drier climates or with stronger resistance to apple pests and diseases.

In addition to the multi-environment and multi-management design, protocols for phenotyping of various traits have been applied since 2018 to evaluate the apple REFPOP. More than ten different traits including yield, fruit quality and phenology are being simultaneously phenotyped with the same method at the six environments. Genotypes are replicated at least twice at each of the environments; each tree is evaluated for all traits separately to allow for variance decomposition up to the individual level (i.e., the tree). Using multi-trait genomic prediction models, prediction of traits with low

heritability or labor-intensive phenotyping can be supported by genetically correlated traits with higher heritability and available phenotypes[35,36].

2.5 Conclusion

This study benefits from a collaborative European approach dedicated to the improvement of apple breeding via genomics-assisted methods. A reference population, which sampled diverse apple germplasm and current European breeding material at larger scales than most current breeding programs, has been established. An extensive set of high-density SNP marker data has been assembled via cost-effective validated marker imputation while making use of the recently available SNP arrays and pedigrees. The imputation method of localized haplotype clustering together with the consolidated high-density SNP marker dataset can be implemented as a standard for cost-effective genomics-assisted breeding. Our diversity and quantitative genetics analyses showed that the reference population is representative of the current apple diversity and breeding material, and that the associated genotypic resources and experimental design allow for the development and application of genomics-assisted breeding methods in apple. This work emphasized the positive effects of high marker density on GWAS and the role of trait architecture in both GWAS and genomic prediction. The apple REFPOP with its unique multi-environment, multi-management and multi-trait design represents a rich source of data for future environment-specific GWAS and genomic predictions. Particularly, the predictions produced with models accounting for the interaction effects between genotype, environment and management as well as using multi-trait modelling can help untangle the effects underlying the traits and ultimately improve the efficiency and success of apple breeding. The apple REFPOP will be the cornerstone of many future projects including the application of genomic selection in apple and the work on apple in the EU-H2020-INVITE project (2019-2024).

Acknowledgements

The authors wish to thank Dr. Eric van de Weg (Wageningen University and Research, The Netherlands (WUR)) and Dr. Mario Di Guardo (WUR and Fondazione Edmund Mach, Italy; presently Dipartimento di Agricoltura, Alimentazione e Ambiente, University of Catania, Catania, Italy) for making raw genotypic data available. Additional genotyping was supported by an INRAE SelGen grant (project named GDivSelGen: "Efficient use of genetic diversity in genomic selection"). The authors thank the field technicians and staff at INRAE Experimental Unit (UE Horti), Angers, France and technical staff at other apple REFPOP sites for the maintenance of the orchards. We thank (i) INRAE Experimental Unit (UE Horti), Angers, France, (ii) Better3Fruit, Rillaar, Belgium and (iii) Consorzio Italiano Vivaisti, San Giuseppe, Italy for grafting of the plant material. This work was partially supported by the project RIS3CAT (COTPA-FRUIT3CAT) financed by the European Regional Development Fund through the FEDER frame of Catalonia 2014-2020 and by the CERCA Programme from Generalitat de Catalunya. We acknowledge financial support from the Spanish Ministry of Economy and Competitiveness through the “Severo Ochoa Programme for Centres of Excellence in R&D” 2016-2019 (SEV-20150533). This work has been partially funded by the EU seventh Framework Programme, the FruitBreedomics project N°265582: Integrated Approach for Increasing Breeding Efficiency in Fruit Tree Crops. The views expressed in this work are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission. Christian Dujak was supported by "DON CARLOS ANTONIO LOPEZ" Abroad Postgraduate Scholarship Program, BECAL-Paraguay.

References

1 Food and Agriculture Organization of the United Nations. *FAOSTAT* (2017).

2 Urrestarazu, J. *et al.* Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology* **16,** 130; 10.1186/s12870-016-0818-0 (2016).

3 Urrestarazu, J. *et al.* Genome-wide association mapping of flowering and ripening periods in apple. *Frontiers in Plant Science* **8,** 1923; 10.3389/fpls.2017.01923 (2017).

4 Kumar, S., Volz, R. K., Chagné, D. & Gardiner, S. in *Genomics of plant genetic resources: Volume 2. Crop productivity, food security and nutritional quality,* edited by R. Tuberosa, A. Graner & E. Frison (Springer Netherlands, Dordrecht, 2014), pp. 387-416.

5 Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157,** 1819 (2001).

6 Georges, M., Charlier, C. & Hayes, B. Harnessing genomic information for livestock improvement. *Nature Reviews Genetics* **20,** 135-156; 10.1038/s41576-018-0082-2 (2019).

7 Laurens, F. *et al.* An integrated approach for increasing breeding efficiency in apple and peach in Europe. *Horticulture Research* **5,** 11; 10.1038/s41438-018-0016-3 (2018).

8 Kumar, S. *et al.* Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). *PLOS ONE* **7,** e36674; 10.1371/journal.pone.0036674 (2012).

9 Muranty, H. *et al.* Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research* **2,** 15060; 10.1038/hortres.2015.60 (2015).

10 Migicovsky, Z. *et al.* Genome to phenome mapping in apple using historical data. *The Plant Genome* **9**; 10.3835/plantgenome2015.11.0113 (2016).

11 McClure, K. A. *et al.* A genome-wide association study of apple quality and scab resistance. *The Plant Genome* **11,** 170075; 10.3835/plantgenome2017.08.0075 (2018).

12 McClure, K. A. *et al.* Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Horticulture Research* **6,** 107; 10.1038/s41438-019-0190-y (2019).

13 Roth, M. *et al.* Prediction of fruit texture with training population optimization for efficient genomic selection in apple. *bioRxiv,* 862193; 10.1101/862193 (2019).

14 Kumar, S., Hilario, E., Deng, C. H. & Molloy, C. Turbocharging introgression breeding of perennial fruit crops: A case study on apple. *Horticulture Research* **7,** 47; 10.1038/s41438-020-0270-z (2020).

15 Covarrubias-Pazaran, G. *et al.* Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight in biparental populations of *Vaccinium macrocarpon* Ait. *Frontiers in Plant Science* **9,** 1310; 10.3389/fpls.2018.01310 (2018).

16 Peace, C. P. *et al.* Apple whole genome sequences: recent advances and new prospects. *Horticulture Research* **6,** 59; 10.1038/s41438-019-0141-7 (2019).

17 Bianco, L. *et al.* Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLOS ONE* **9,** e110377; 10.1371/journal.pone.0110377 (2014).

18 Bianco, L. *et al.* Development and validation of the Axiom®Apple480K SNP genotyping array. *The Plant Journal* **86,** 62-74; 10.1111/tpj.13145 (2016).

19 Muranty, H. *et al.* Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biology* **20,** 2; 10.1186/s12870-019-2171-6 (2020).
20 Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* **23,** 52-71; 10.1016/j.spasta.2017.10.003 (2018).
21 Lassois, L. *et al.* Genetic diversity, population structure, parentage analysis, and construction of core collections in the French apple germplasm based on SSR markers. *Plant Molecular Biology Reporter* **34,** 827-844; 10.1007/s11105-015-0966-7 (2016).
22 Kumar, S., Volz, R. K., Alspach, P. A. & Bus, V. G. M. Development of a recurrent apple breeding programme in New Zealand: A synthesis of results, and a proposed revised breeding strategy. *Euphytica* **173,** 207-222; 10.1007/s10681-009-0090-6 (2010).
23 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics* **48,** 1279-1283; 10.1038/ng.3643 (2016).
24 Pégard, M. *et al.* Sequence imputation from low density single nucleotide polymorphism panel in a black poplar breeding population. *BMC Genomics* **20,** 302; 10.1186/s12864-019-5660-y (2019).
25 Vanderzande, S., Micheletti, D., Troggio, M., Davey, M. W. & Keulemans, J. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. *Tree Genetics & Genomes* **13,** 125; 10.1007/s11295-017-1206-0 (2017).
26 Daccord, N. *et al.* High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49,** 1099-1106; 10.1038/ng.3886 (2017).
27 Cornille, A. *et al.* New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLOS Genetics* **8,** e1002703; 10.1371/journal.pgen.1002703 (2012).
28 Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44,** 825-830; 10.1038/ng.2314 (2012).
29 Celton, J.-M. *et al.* Deciphering the genetic determinism of bud phenology in apple progenies: A new insight into chilling and heat requirement effects on flowering dates and positional candidate. *New Phytologist* **192,** 378-392; 10.1111/j.1469-8137.2011.03823.x (2011).
30 Allard, A. *et al.* Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. *Journal of Experimental Botany* **67,** 2875-2888; 10.1093/jxb/erw130 (2016).
31 Chagné, D. *et al.* Genetic and environmental control of fruit maturation, dry matter and firmness in apple (*Malus* × *domestica* Borkh.). *Horticulture Research* **1,** 14046; 10.1038/hortres.2014.46 (2014).
32 Larsen, B. *et al.* Genome-wide association studies in apple reveal loci for aroma volatiles, sugar composition, and harvest date. *The Plant Genome* **12,** 180104; 10.3835/plantgenome2018.12.0104 (2019).
33 Gezan, S. A., Osorio, L. F., Verma, S. & Whitaker, V. M. An experimental validation of genomic selection in octoploid strawberry. *Horticulture Research* **4,** 16070; 10.1038/hortres.2016.70 (2017).

34 Hardner, C. M. *et al.* Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. *Horticulture Research* **6,** 6; 10.1038/s41438-018-0081-7 (2019).

35 Schulthess, A. W. *et al.* Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theoretical and Applied Genetics* **129,** 273-287; 10.1007/s00122-015-2626-6 (2016).

36 Lado, B. *et al.* Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theoretical and Applied Genetics* **131,** 2719-2731; 10.1007/s00122-018-3186-3 (2018).

37 Fernández-Fernández, F. SID 5 (Research Project Final Report): Fingerprinting the national apple & pear collections. Defra, 2010.

38 Bink, M. C. A. M. *et al.* Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. *Theoretical and Applied Genetics* **127,** 1073-1090; 10.1007/s00122-014-2281-3 (2014).

39 Jansen, J. & van Hintum, T. Genetic distance sampling: A novel sampling method for obtaining core collections using genetic distances with an application to cultivated lettuce. *Theoretical and Applied Genetics* **114,** 421-428; 10.1007/s00122-006-0433-9 (2007).

40 European Environment Agency. Biogeographical regions in Europe. Available at https://www.eea.europa.eu/data-and-maps/figures/biogeographical-regions-in-europe-2 (2017).

41 Vanderzande, S. *et al.* High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. *PLOS ONE* **14,** e0210928; 10.1371/journal.pone.0210928 (2019).

42 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81,** 1084-1097; 10.1086/521987 (2007).

43 Clayton, D. *snpStats: SnpMatrix and XSnpMatrix classes and methods* (2017).

44 Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289-290; 10.1093/bioinformatics/btg412 (2004).

45 Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software* **25,** 1-18 (2008).

46 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19,** 1655-1664 (2009).

47 International union for the protection of new varieties of plants (UPOV). *Guidelines for the conduct of tests for distinctness, uniformity and stability: TG/14/9 Apple (fruit varieties)* (Geneva, 2005).

48 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67** (2015).

49 Højsgaard, S. & Halekoh, U. *doBy: Groupwise statistics, LSmeans, linear contrasts, utilities* (2018).

50 Tang, Y. *et al.* GAPIT version 2: An enhanced integrated tool for genomic association and prediction. *The Plant Genome* **9**; 10.3835/plantgenome2015.11.0120 (2016).

51 Endelman, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome* **4,** 250-255; 10.3835/plantgenome2011.08.0024 (2011).

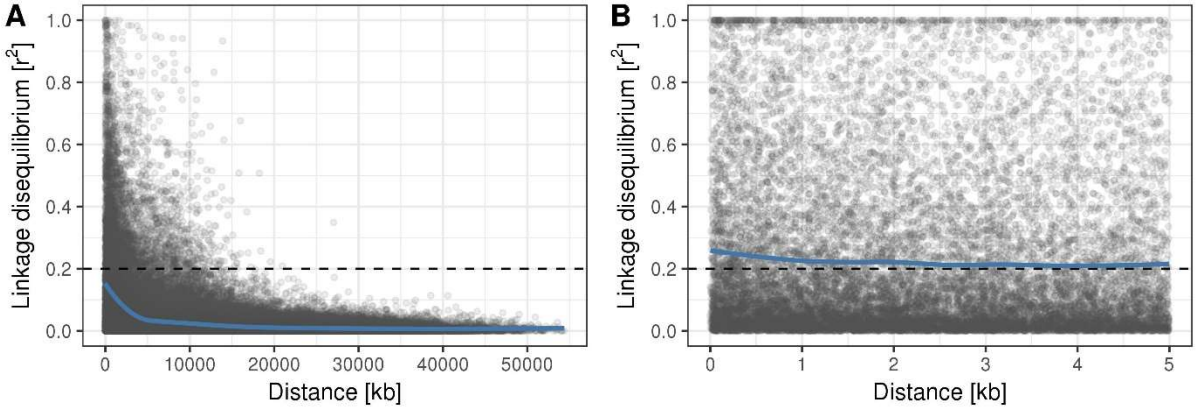
52 Bermingham, M. L. *et al.* Application of high-dimensional feature selection: Evaluation for genomic prediction in man. *Scientific Reports* **5,** 10312; 10.1038/srep10312 (2015).
53 Elsen, J.-M. An analytical framework to derive the expected precision of genomic selection. *Genetics Selection Evolution* **49,** 95; 10.1186/s12711-017-0366-6 (2017).
54 R Core Team. *R: A language and environment for statistical computing* (R Foundation for Statistical Computing, Vienna, Austria, 2014).
55 Wickham, H. *ggplot2: Elegant graphics for data analysis* (Springer, 2016).

Supplementary methods

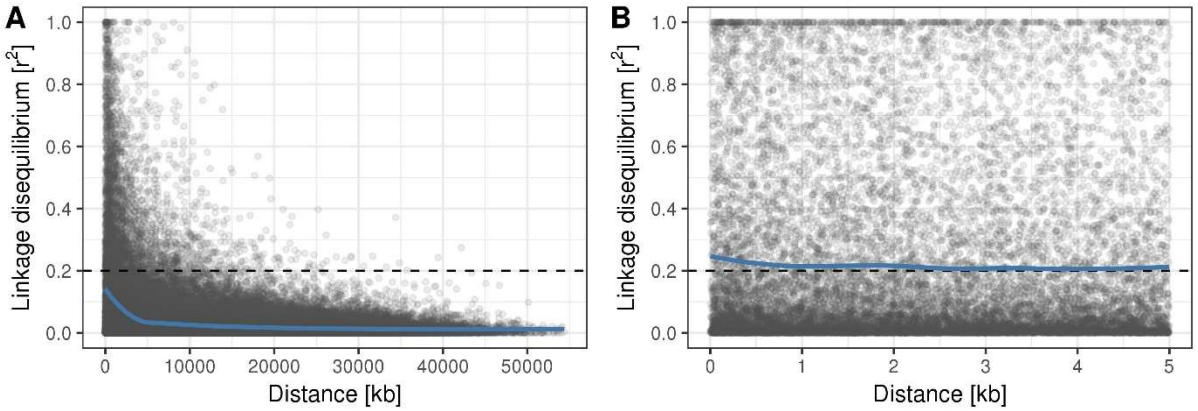
Supplementary Methods 1

The Axiom Analysis Suite software was used for processing raw hybridization intensity data of all genotyped samples, clustering and genotype calling. Samples with a dish quality control value < 0.82 and sample call rate < 0.97 were excluded. Within the Axiom Analysis Suite software, the SNPs were classified into categories (i) poly high resolution, (ii) mono high resolution, (iii) off target variant, (iv) call rate below threshold, (v) no minor homozygote and (vi) other. The data from the first and fifth category were exported and submitted to further filtering criteria. First, markers were excluded when they showed two or more Mendelian errors in (i) the pedigree identified using all genotyped diploid unique genotypes by Muranty et al. (2020) or (ii) 92 progeny of parental combinations 'Fuji' × 'Pinova' and 'Golden Delicious' × 'Renetta Grigia di Torriana' or (iii) between 'Golden Delicious' and its two doubled haploid offspring. Second, SNPs were removed if they showed (i) differences between duplicates in two or more groups of duplicates (based on 41 groups of duplicates, including 'Golden Delicious' genotyped in each plate), (ii) a heterozygous score in at least one of the two doubled haploid offspring of 'Golden Delicious', (iii) overall heterozygous scoring of five or less.

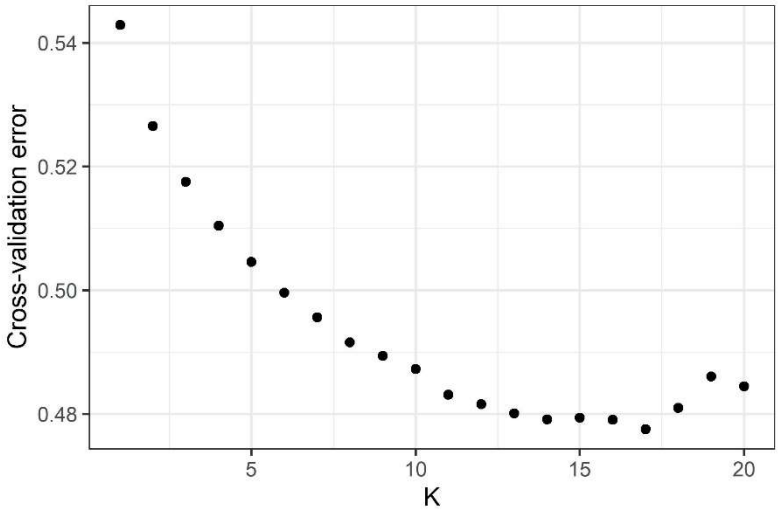
Supplementary figures



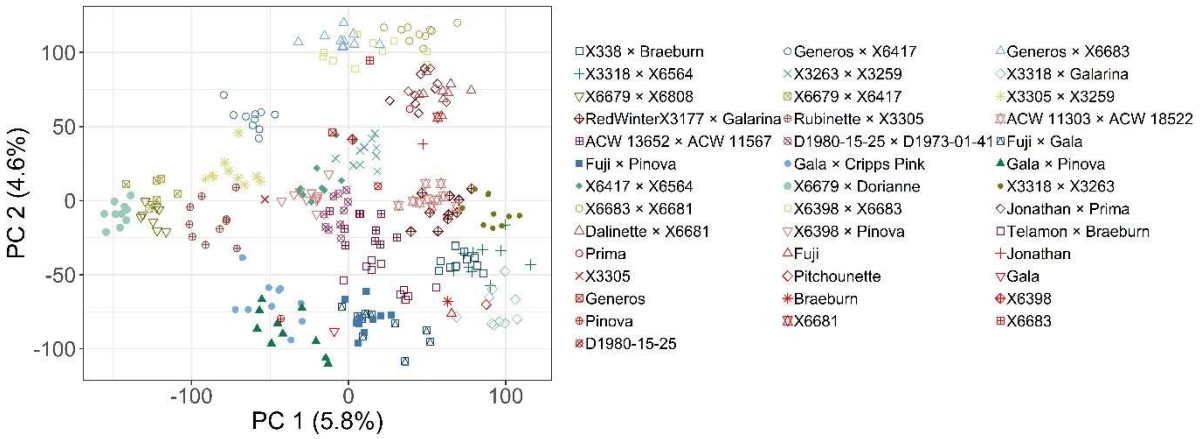
Supplementary Figure 1: Linkage disequilibrium of the accession section, with a loess smoother for (A) distances between SNPs across the span of chromosomes, (B) for SNPs within a 5 kb distance.



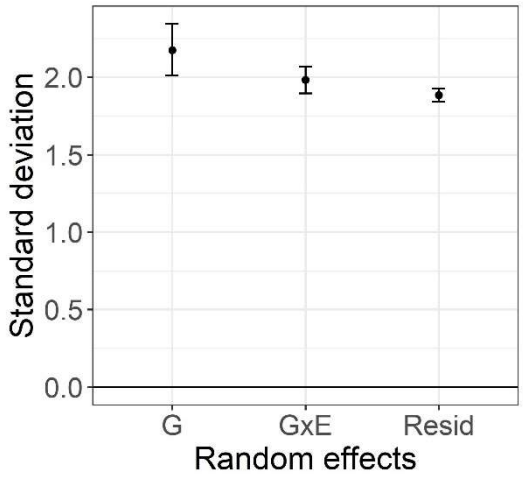
Supplementary Figure 2: Linkage disequilibrium of the progeny section, with a loess smoother for (A) distances between SNPs across the span of chromosomes, (B) for SNPs within a 5 kb distance.



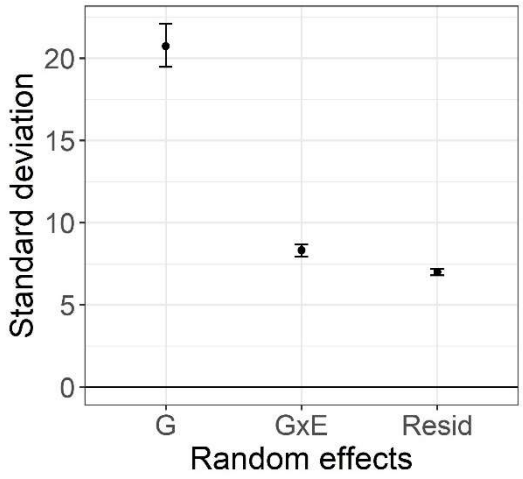
Supplementary Figure 3: ADMIXTURE cross-validation error



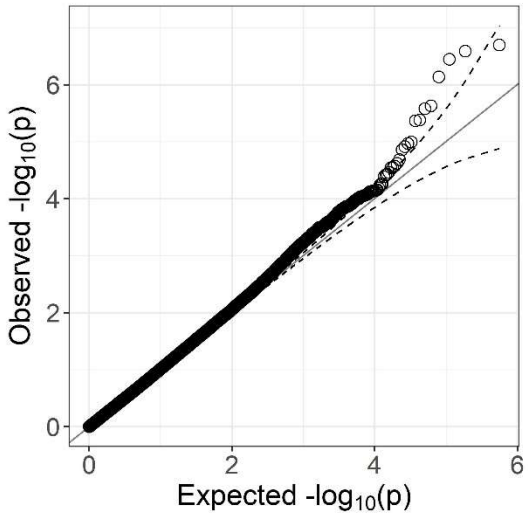
Supplementary Figure 4: Principal component analysis of the progeny section (27 parental combinations) with 13 parents as supplementary individuals shown in red color. Plot of the first two principal components with their respective proportion of the total variance shown within brackets.



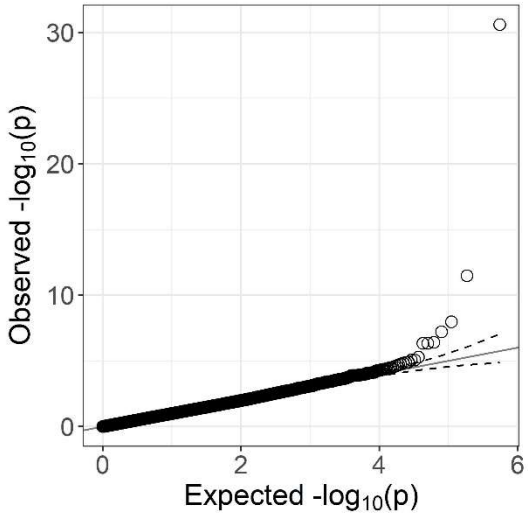
Supplementary Figure 5: Standard deviations and confidence intervals for the random effects of genotype (G), genotype by environment interaction (G×E) and residuals (Resid) explaining the floral emergence



Supplementary Figure 6: Standard deviations and confidence intervals for the random effects of genotype (G), genotype by environment interaction (G×E) and residuals (Resid) explaining the harvest date



Supplementary Figure 7: QQ-plot of the observed versus expected p-values for floral emergence



Supplementary Figure 8: QQ-plot of the observed versus expected p-values for harvest date

Supplementary tables

Supplementary table 1

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
1	1	1945-079	Brabant Bellefleur	WCE
101	101	1947-004	Rosa Mantovana	SE
1015	1015	1951-050	Schmidtberger Renette	WCE
1020	1020	1948-614	Chüsenrainer	WCE
1091	1091	1947-200	Ribonde	WCE
1113	1113	X4194	Red Winter	USA
112	112	1948-744	Democrat	ANZ
1123	1123	1953-014	Horei	JPN
1125	1125	1978-136	Lady Williams	ANZ
114	114	1947-318	Koestlicher	WCE
118	118	1948-236	Reinette Clochard	WCE
1218	1218	1953-002	Shinsei	JPN
1236	1236	X2039	O53T136	USA
130	130	1944-048	Groninger Kroon	WCE
131	131	1947-143	Reinette du Mans	WCE
133	133	1948-599	Spaanse Keiing	WCE
135	135	1947-138	Reinette de Saintonge	WCE
1377	1377	1948-400	Simonffy Piros	WCE
140	140	1946-034	Lange's Perfection	WCE
141	141	2000-070	Mrs. Phillimore	WCE
143	143	1954-026	Polly	WCE
1446	1446	1958-105	Hoti	SEE
1448	1448	1975-004	Osenee Polosatoe	NEE
1478	1478	Priscilla_NL	Priscilla-NL	WCE
1481	1481	1967-089	Indo	JPN
149	149	1977-204	Woodford	WCE
15	15	1957-175	Annie Elizabeth	WCE
152	152	1953-050	Histon Favourite	WCE
163	163	1907-002	Cox's Orange Pippin	WCE
167	167	1974-060	Melba	CAN
17	17	1923-058	Ontario	CAN
170	170	1973-114	Jincoa Zagarra	WCE
173	173	1951-170	Edelroter	SE
1870	1870	1905-004	Ard Cairn Russet	WCE
1872	1872	1916-006	Beeley Pippin	WCE
1887	1887	1923-078	Fillingham Pippin	WCE
1889	1889	1924-005	Scilly Pearl	WCE
1893	1893	1924-057	Bushey Grove	WCE
1900	1900	1929-036	New Bess Pool	WCE
1901	1901	1930-020	Summer Golden Pippin	WCE
1910	1910	1933-012	Sam Young	WCE
192	192	1948-233	Rose Double	WCE
1935	1935	1945-001	Storey's Seedling	WCE
1951	1951	1946-043	Brown's Seedling	WCE
1952	1952	1946-091	Holland Pippin	WCE
1962	1962	1947-480	Montfort	WCE
197	197	1973-051	Annurca	SE
1985	1985	1949-207	Barnhill Pippin	WCE
1988	1988	1949-221	Summer John	WCE
1989	1989	1949-222	Ballyfatten	WCE
1991	1991	1949-272	Lass o' Gowrie	WCE
1997	1997	1950-042	Stoke Allow	WCE
2	2	1958-156	Carrata	SE
200	200	1951-179	Gustavs Dauerapfel	WCE
2003	2003	1950-079	Eight Square	WCE
2010	2010	1951-224	Longstart	WCE
2013	2013	1952-055	Gilliflower of Gloucester	WCE

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
202	202	1947-086	Franc Roseau du Valais	WCE
2024	2024	1955-024	Pitmaston Pine Apple	WCE
2026	2026	1955-078	Cornish Honeypin	WCE
203	203	1924-009	Nanny	WCE
204	204	2006-011	Tydeman's Early Worcester	WCE
2046	2046	1960-044	Sergeant Peggy	WCE
2050	2050	1962-033	Lodgemore Nonpareil	WCE
2055	2055	1965-012	Lady Hollendale	WCE
206	206	1948-209	Pomme d'Ile	WCE
2061	2061	1974-496	Unnamed-1	U
2071	2071	1981-148	Old Rock Pippin	WCE
216	216	1973-116	Bordes	WCE
236	236	1951-017	Early Red Bird	CAN
238	238	1976-019	Sylvia	NEE
241	241	1957-084	Canvada	ANZ
2410	2410	X4355	P7 R4A4	WCE
2422	2422	X6398	TN R42A60	WCE
2443	2443	2002-056	Mason's Pippin	U
249	249	1974-065	Orleans	USA
25	25	1954-054	White Transparent	NEE
26	26	1949-202	Directeur Lesage	WCE
261	261	1979-176	Merton Worcester	WCE
262	262	1944-005	Thomas Jeffrey	WCE
2625	2625	1981-117	Orin	JPN
2651	2651	1976-107	Siugisdesert	NEE
2652	2652	1976-105	Piltsamasskoe Zimnee	NEE
2653	2653	1976-104	Noris	NEE
267	267	1972-019	Prima	USA
2670	2670	1974-323	Rosanne	ANZ
2671	2671	1974-322	Stowell Cox	ANZ
270	270	1947-158	Belle Flavoise	WCE
2720	2720	1967-068	Edgar (Canada)	CAN
2722	2722	1967-066	Diana	CAN
2732	2732	1966-044	Balder	WCE
2762	2762	1961-087	Tropical Beauty	ZAF
2779	2779	1958-100	Cretesc de Breaza	SEE
2784	2784	1958-043	Giambun	SE
2785	2785	1958-037	Gabiola	SE
2794	2794	1957-205	Stonehenge	CAN
2796	2796	1957-202	Milfor	CAN
2802	2802	1957-193	Emilia	CAN
2809	2809	1957-079	Titovka	NEE
2810	2810	1957-078	Kitchovka	SEE
2813	2813	1957-073	Aivaniya	SEE
2817	2817	1956-035	Caroline Hopkins	ZAF
2820	2820	1956-006	Toyo	JPN
2823	2823	1955-015	President Boudewijn	WCE
2827	2827	1955-011	Primus	WCE
2830	2830	1955-004	Prinses Marijke	WCE
2831	2831	1955-003	Prinses Beatrix	WCE
2836	2836	1954-069	Unnamed-5	U
2847	2847	1953-010	Golden Melon	JPN
2848	2848	1953-006	Shinko	JPN
2849	2849	1953-005	Shin Indo	JPN
2850	2850	1953-004	Kyokko	JPN
2855	2855	1952-178	Gold Reinette	WCE
2856	2856	1952-130	Jersey Black	USA
286	286	1968-035	Katja	NEE
2862	2862	1952-120	Golden Russet of Western New York	USA
2865	2865	1952-113	Jefferis	USA
2872	2872	1952-031	Pero Dourado	SE

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
29	29	1957-190	Borowitsky	NEE
290	290	1957-213	Chelmsford Wonder	WCE
2902	2902	1951-061	Fireside	USA
2908	2908	1951-021	Henry Clay	USA
2936	2936	1950-037	Primate	USA
2971	2971	1948-724	Petit Pippin	U
2974	2974	1948-688	Mere de Baia	SEE
2975	2975	1948-671	Karolka	SEE
2979	2979	1948-644	Domnicele	SEE
2980	2980	1948-642	Meri de Saminta	SEE
2990	2990	1948-403	Szaszpap alma	WCE
2992	2992	1948-401	Szechenyi Renet	WCE
2997	2997	1948-381	Keresi Muskotaly	WCE
30	30	1999-072	Alexander	NEE
301	301	1953-051	Spätblühender Taffetapfel	WCE
3010	3010	1948-354	Banffy Pal	WCE
3013	3013	1948-344	Angyal Dezso	WCE
3026	3026	1948-038	Unnamed-6	U
3042	3042	1947-324	Josephine Kreuter	WCE
3062	3062	1947-062	Frau Margarete von Stosch	WCE
308	308	1927-013	Åkerö	NEE
3148	3148	1930-027	Baxter	CAN
3160	3160	1927-004	Bodil Neergaard	NEE
317	317	1947-051	Marie-Joseph d'Othée	WCE
3170	3170	1925-019	Thurso	CAN
3172	3172	1925-016	Joyce	CAN
3173	3173	1925-015	Ascot	CAN
3190	3190	1921-089	Grimes Golden	USA
32	32	1965-025	Ingrid Marie	NEE
321	321	1947-462	Imperiale	WCE
323	323	1948-774	Bec d'Oie	WCE
324	324	1946-088	Winter Marigold	WCE
327	327	1958-139	Abbondanza	SE
33	33	1952-108	Gros Api	WCE
330	330	1958-128	Sicilia Piccola	SE
333	333	1915-102	William Crump	SE
334	334	1943-007	Rome Beauty	USA
338	338	1950-155	Cabusse	WCE
34	34	1927-007	Signe Tillisch	NEE
345	345	1927-005	Skovfoged	NEE
353	353	1948-778	Rouget de Born	WCE
366	366	1971-048	Vista Bella	USA
37	37	1999-004	King of the Pippins	WCE
378	378	1947-297	Friandise	WCE
40	40	1941-022	Benoni	USA
405	405	1947-223	Reinette d'Anthezieux	WCE
409	409	1958-140	Campanino	SE
410	410	1958-141	Durello	SE
424	424	1947-161	Barraude	WCE
428	428	1950-178	Rose Rouge	WCE
43	43	1948-246	Colapuis	NEE
435	435	1958-030	Renetta Grigia di Torriana	SE
440	440	1947-383	Amasya	SEE
445	445	1958-164	Mela Violetta	SE
447	447	1974-052	Winesap	USA
45	45	1957-181	Gascoyne's Scarlet	WCE
457	457	1947-216	Gazerau	WCE
46	46	2000-082	Reinette Rouge Etoilee	WCE
462	462	1976-149	Wagener	USA
465	465	1951-203	Tiroler Spitzlederer	SE
468	468	1979-186	Stark Earliest	USA

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
472	472	1948-278	Rouget	WCE
473	473	1929-029	Broad-Eyed Pippin (of Bultitude)	WCE
483	483	1955-067	Snövit	NEE
488	488	1975-317	Mother	USA
489	489	1944-004	Venus Pippin	WCE
49	49	1947-084	Freiherr von Berlepsch	WCE
491	491	1947-263	Reinette Sanguine du Rhin	WCE
5	5	1975-339	Antonovka Polotora Funtovaja	NEE
50	50	1921-094	Winter Banana	USA
508	508	2006-014	McIntosh	CAN
51	51	2000-075	Peasgood's Nonsuch	WCE
515	515	1950-150	Bancroft	CAN
52	52	1957-253	Zigeunerin	WCE
522	522	1950-033	Esopus Spitzenburg	USA
526	526	1982-200	Melrose	USA
533	533	1975-006	Pepin Shafrannyj	NEE
543	543	1955-066	Mio	NEE
546	546	1953-133	Rall's Janet	USA
548	548	1976-145	Granny Smith	ANZ
549	549	1984-001	Hocking's Green	WCE
55	55	1951-197	Rosmarina bianca	SE
56	56	1949-082	Zuccalmaglios Renette	WCE
560	560	1927-028	Opalescent	USA
562	562	1947-136	Collins	USA
564	564	1946-108	Caroline	WCE
565	565	1973-251	Kansas Queen	U
567	567	1958-041	Gambefine Fine Lunghe	SE
57	57	1979-164	Jonathan	USA
570	570	1947-165	Pomme Fer	WCE
575	575	1950-127	Atlas	CAN
577	577	1950-044	Norman's Pippin	WCE
58	58	1975-351	Calville des Prairies	WCE
582	582	2000-101	Wyken Pippin	WCE
584	584	1971-046	Kidd's Orange Red	ANZ
587	587	1974-264	York Imperial	USA
592	592	1952-194	Ivöäpple	USA
6	6	1950-123	Wealthy	USA
601	601	1953-081	Sops-in-Wine	WCE
603	603	1963-106	Mantet	CAN
61	61	1929-026	Reinette de Landsberg	WCE
629	629	1953-047	Dutch Mignonne	WCE
639	639	1947-469	Fiessers Erstling	WCE
65	65	1971-054	Golden Delicious	USA
652	652	1975-342	Bellefleur Krasny	NEE
653	653	1947-128	Grosse de Saint-Clément	WCE
66	66	1957-241	Elise Rathke	WCE
662	662	1958-029	Rus Cavallotta	SE
67	67	1948-310	Belle Fille de l'Indre	WCE
670	670	1947-059	Sauergrauschapfel	WCE
673	673	1958-028	Gian d'Andre'	SE
676	676	1976-103	Bessemyanka Michurina	NEE
678	678	1974-203	Dukat	WCE
68	68	1947-325	Maglemer	NEE
681	681	1948-210	Vernade	WCE
691	691	1953-150	Giant Geniton	USA
697	697	1958-040	Grenoble	SE
71	71	1943-006	King David	USA
72	72	2000-085	Grahams Jubileum	WCE
722	722	1948-656	Danziger Kantapfel	WCE
737	737	1976-108	Zhigulevskoe	NEE
739	739	1976-144	Gala	ANZ

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
74	74	1921-015	Edward VII	WCE
740	740	1948-363	Daru Sovari	WCE
743	743	1974-205	Růžena Bláhová	WCE
748	748	1950-175	Reale d'Entraygues	WCE
75	75	1973-189	Discovery	WCE
750	750	1950-120	Tönnes	NEE
757	757	1947-056	La Nationale	WCE
764	764	1956-014	Schweizer Orangenapfel	WCE
77	77	1953-140	Yellow Bellflower	USA
779	779	1951-058	Schöner aus Herrnhut	WCE
786	786	1945-155	Devonshire Buckland	WCE
787	787	1963-025	Newtown Pippin	USA
8	8	1979-159	Egremont Russet	WCE
80	80	1968-039	Aroma	NEE
809	809	1944-046	Zoete Ermgaard	WCE
81	81	1947-050	Prinzen Apfel	WCE
810	810	1957-082	Verde Doncella	SE
82	82	1999-084	Red Astrachan	NEE
89	89	1947-189	Reinette de Geer	WCE
9	9	1949-182	Lombarts Calville	WCE
918	918	1921-010	Charles Eyre	WCE
929	929	1958-018	Barry	USA
93	93	1974-347	Grenadier	WCE
94	94	1927-023	Filippa	NEE
95	95	2002-037	Calville Blanc d'Hiver	WCE
950	950	1957-191	Ecklinville	WCE
953	953	1947-133	Pero Mingan	SE
97	97	1951-048	Neue Goldparmane	WCE
997	997	1948-643	Nemtesc cu Miezul Rosu	SEE
12_B006		12_B006		P
12_B020		12_B020		P
12_B023		12_B023		P
12_B057		12_B057		P
12_B061		12_B061		P
12_B064		12_B064		P
12_B074		12_B074		P
12_B081		12_B081		P
12_B093		12_B093		P
12_B095		12_B095		P
12_E003		12_E003		P
12_E014		12_E014		P
12_E015		12_E015		P
12_E017		12_E017		P
12_E024		12_E024		P
12_E025		12_E025		P
12_E034		12_E034		P
12_E035		12_E035		P
12_E039		12_E039		P
12_E043		12_E043		P
12_F002		12_F002		P
12_F007		12_F007		P
12_F014		12_F014		P
12_F019		12_F019		P
12_F023		12_F023		P
12_F040		12_F040		P
12_F045		12_F045		P
12_F046		12_F046		P
12_F052		12_F052		P
12_F054		12_F054		P
12_I002		12_I002		P
12_I003		12_I003		P

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
12_I005		12_I005		P
12_I010		12_I010		P
12_I037		12_I037		P
12_I049		12_I049		P
12_I056		12_I056		P
12_I059		12_I059		P
12_I062		12_I062		P
12_I064		12_I064		P
12_J001		12_J001		P
12_J004		12_J004		P
12_J011		12_J011		P
12_J013		12_J013		P
12_J015		12_J015		P
12_J019		12_J019		P
12_J027		12_J027		P
12_J028		12_J028		P
12_J029		12_J029		P
12_J031		12_J031		P
12_K001		12_K001		P
12_K010		12_K010		P
12_K016		12_K016		P
12_K018		12_K018		P
12_K023		12_K023		P
12_K024		12_K024		P
12_K044		12_K044		P
12_K045		12_K045		P
12_K049		12_K049		P
12_L001		12_L001		P
12_L005		12_L005		P
12_L012		12_L012		P
12_L013		12_L013		P
12_L018		12_L018		P
12_L021		12_L021		P
12_N021		12_N021		P
12_N034		12_N034		P
12_N039		12_N039		P
12_N041		12_N041		P
12_N046		12_N046		P
12_N052		12_N052		P
12_N054		12_N054		P
12_N057		12_N057		P
12_N062		12_N062		P
12_N063		12_N063		P
12_O010		12_O010		P
12_O011		12_O011		P
12_O014		12_O014		P
12_O017		12_O017		P
12_O026		12_O026		P
12_O056		12_O056		P
12_O061		12_O061		P
12_O063		12_O063		P
12_O072		12_O072		P
12_O074		12_O074		P
12_P001		12_P001		P
12_P002		12_P002		P
12_P003		12_P003		P
12_P018		12_P018		P
12_P025		12_P025		P
12_P027		12_P027		P
12_P029		12_P029		P
12_P034		12_P034		P

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
12_P055		12_P055		P
12_P056		12_P056		P
12_196		12-196	B3Fx12-196	P
12_197		12-197	B3Fx12-197	P
12_204		12-204	B3Fx12-204	P
12_205		12-205	B3Fx12-205	P
13_10		13-10	B3Fx13-10	P
13_109		13-109	B3Fx13-109	P
13_139		13-139	B3Fx13-139	P
13_32		13-32	B3Fx13-32	P
13_35		13-35	B3Fx13-35	P
13_38		13-38	B3Fx13-38	P
ACW_25358		ACW 25358		P
ACW_25363		ACW 25363		P
ACW_25365		ACW 25365		P
ACW_25366		ACW 25366		P
ACW_25368		ACW 25368		P
ACW_25371		ACW 25371		P
ACW_25372		ACW 25372		P
ACW_25375		ACW 25375		P
ACW_25454		ACW 25454		P
ACW_25469		ACW 25469		P
ACW_25580		ACW 25580		P
ACW_25581		ACW 25581		P
ACW_25584		ACW 25584		P
ACW_25586		ACW 25586		P
ACW_25591		ACW 25591		P
ACW_25592		ACW 25592		P
ACW_25607		ACW 25607		P
ACW_25619		ACW 25619		P
ACW_25622		ACW 25622		P
ACW_25628		ACW 25628		P
DLO_12_066		DLO-12_ 066		P
DLO_12_084		DLO-12_ 084		P
DLO_12_120		DLO-12_ 120		P
DLO_12_140		DLO-12_ 140		P
DLO_12_141		DLO-12_ 141		P
DLO_12_162		DLO-12_ 162		P
DLO_12_221		DLO-12_ 221		P
DLO_12_233		DLO-12_ 233		P
DLO_12_259		DLO-12_ 259		P
DLO_12_289		DLO-12_ 289		P
FuGa_015		FuGa 015		P
FuGa_016		FuGa 016		P
FuGa_140		FuGa 140		P
FuGa_187		FuGa 187		P
FuGa_215		FuGa 215		P
FuGa_217		FuGa 217		P
FuGa_243		FuGa 243		P
FuGa_270		FuGa 270		P
FuGa_275		FuGa 275		P
FuGa_352		FuGa 352		P
FuPi_005		FuPi 005		P
FuPi_013		FuPi 013		P
FuPi_023		FuPi 023		P
FuPi_052		FuPi 052		P
FuPi_054		FuPi 054		P
FuPi_087		FuPi 087		P
FuPi_097		FuPi 097		P
FuPi_106		FuPi 106		P
FuPi_107		FuPi 107		P

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
FuPi_114		FuPi 114		P
GaCr_002		GaCr 002		P
GaCr_010		GaCr 010		P
GaCr_017		GaCr 017		P
GaCr_041		GaCr 041		P
GaCr_043		GaCr 043		P
GaCr_052		GaCr 052		P
GaCr_055		GaCr 055		P
GaCr_056		GaCr 056		P
GaCr_057		GaCr 057		P
GaCr_061		GaCr 061		P
GaPi_007		GaPi 007		P
GaPi_014		GaPi 014		P
GaPi_017		GaPi 017		P
GaPi_023		GaPi 023		P
GaPi_028		GaPi 028		P
GaPi_033		GaPi 033		P
GaPi_035		GaPi 035		P
GaPi_038		GaPi 038		P
GaPi_051		GaPi 051		P
GaPi_056		GaPi 056		P
I_BB002		I_BB002		P
I_BB012		I_BB012		P
I_BB015		I_BB015		P
I_BB022		I_BB022		P
I_BB028		I_BB028		P
I_BB031		I_BB031		P
I_BB039		I_BB039		P
I_BB049		I_BB049		P
I_BB056		I_BB056		P
I_BB070		I_BB070		P
I_CC003		I_CC003		P
I_CC044		I_CC044		P
I_CC048		I_CC048		P
I_CC052		I_CC052		P
I_CC055		I_CC055		P
I_CC057		I_CC057		P
I_CC059		I_CC059		P
I_CC067		I_CC067		P
I_CC070		I_CC070		P
I_CC075		I_CC075		P
I_J012		I_J012		P
I_J030		I_J030		P
I_J033		I_J033		P
I_J036		I_J036		P
I_J046		I_J046		P
I_J054		I_J054		P
I_J056		I_J056		P
I_J063		I_J063		P
I_J064		I_J064		P
I_J069		I_J069		P
I_M007		I_M007		P
I_M013		I_M013		P
I_M014		I_M014		P
I_M024		I_M024		P
I_M026		I_M026		P
I_M029		I_M029		P
I_M047		I_M047		P
I_M048		I_M048		P
I_M052		I_M052		P
I_M056		I_M056		P

Genotype Code	MUNQ	Original Code	Preferred Name	Origin
I_W002		I_W002		P
I_W010		I_W010		P
I_W034		I_W034		P
I_W036		I_W036		P
I_W037		I_W037		P
I_W044		I_W044		P
I_W045		I_W045		P
I_W046		I_W046		P
I_W049		I_W049		P
I_W050		I_W050		P
JoPr_1990_102_001		JoPr 1990-102-001		P
JoPr_1990_102_054		JoPr 1990-102-054		P
JoPr_1990_102_097		JoPr 1990-102-097		P
JoPr_1990_102_108		JoPr 1990-102-108		P
JoPr_1990_102_169		JoPr 1990-102-169		P
JoPr_1990_102_179		JoPr 1990-102-179		P
JoPr_1990_102_202		JoPr 1990-102-202		P
JoPr_1990_102_232		JoPr 1990-102-232		P
JoPr_1990_102_242		JoPr 1990-102-242		P
JoPr_1990_102_246		JoPr 1990-102-246		P
JoPr_1990_102_257		JoPr 1990-102-257		P
NOVADI_0010		NOVADI / 0010		P
NOVADI_0023		NOVADI / 0023		P
NOVADI_0042		NOVADI / 0042		P
NOVADI_0114		NOVADI / 0114		P
NOVADI_0256		NOVADI / 0256		P
NOVADI_0268		NOVADI / 0268		P
NOVADI_0286		NOVADI / 0286		P
NOVADI_0296		NOVADI / 0296		P
NOVADI_0329		NOVADI / 0329		P
NOVADI_0330		NOVADI / 0330		P
NOVADI_0347		NOVADI / 0347		P
NOVADI_0356		NOVADI / 0356		P
NOVADI_0371		NOVADI / 0371		P
NOVADI_0382		NOVADI / 0382		P
NOVADI_0622		NOVADI / 0622		P
NOVADI_0731		NOVADI / 0731		P
NOVADI_0766		NOVADI / 0766		P
NOVADI_0980		NOVADI / 0980		P
NOVADI_1015		NOVADI / 1015		P
TeBr_012		TeBr_012		P
TeBr_021		TeBr_021		P
TeBr_022		TeBr_022		P
TeBr_027		TeBr_027		P
TeBr_043		TeBr_043		P
TeBr_071		TeBr_071		P
TeBr_084		TeBr_084		P
TeBr_180		TeBr_180		P
TeBr_248		TeBr_248		P
TeBr_263		TeBr_263		P

Supplementary table 2

Family	Size	Parent 1	Parent 2	Institute	Original Project	Original Code
12- or 13-	10	X338	Braeburn	BETTER3FRUIT	This study	
12_B	10	Generos	X6417	INRA	Bianco et al. (2014)	
12_E	10	Generos	X6683	INRA	Bianco et al. (2014)	
12_F	10	X3318	X6564	INRA	Bianco et al. (2014)	
12_I	10	X3263	X3259	INRA	Bianco et al. (2014)	
12_J	10	X3318	Galarina	INRA	Bianco et al. (2014)	
12_K	9	X6679	X6808	INRA	Bianco et al. (2014)	
12_L	6	X6679	X6417	INRA	PBA, FruitBreedomics (Bink et al. 2014; Laurens et al. 2018)	
12_N	10	X3305	X3259	INRA	Bianco et al. (2014)	
12_O	10	RedWinterX3177	Galarina	INRA	PBA, FruitBreedomics (Bink et al. 2014; Laurens et al. 2018)	
12_P	10	Rubinette	X3305	INRA	Bianco et al. (2014)	
ACW 1	10	ACW 11303	ACW 18522	AGROSCOPE	This study	ACW 253XX, ACW 254XX
ACW 2	10	ACW 13652	ACW 11567	AGROSCOPE	This study	ACW 255XX, ACW 256XX
DLO	10	D1980-15-25	D1973-01-41	WUR (Wageningen University & Research)	Bianco et al. (2014)	
FuGa	10	Fuji	Gala	LAIMBURG (duplicate of Bologna progeny)	Bianco et al. (2014)	
FuPi	10	Fuji	Pinova	LAIMBURG	Bianco et al. (2014, 2016)	
GaCr	10	Gala	Cripps Pink	LAIMBURG	PBA, FruitBreedomics (Bink et al. 2014; Laurens et al. 2018)	
GaPi	10	Gala	Pinova	LAIMBURG	Bianco et al. (2014)	
I_BB	10	X6417	X6564	INRA	Bianco et al. (2014)	
I_CC	10	X6679	Dorianne	INRA	Bianco et al. (2014)	
I_J	10	X3318	X3263	INRA	Bianco et al. (2014)	
I_M	10	X6683	X6681	INRA	Bianco et al. (2014)	
I_W	10	X6398	X6681	INRA	Bianco et al. (2014)	
JoPr	11	Jonathan	Prima	SLU (WUR population)	Bianco et al. (2014)	
NOVADI 1	9	Dalinette	X6681	NOVADI	This study	NOVADI / 02XX, NOVADI / 03XX
NOVADI 2	10	X6398	Pinova	NOVADI	This study	NOVADI / 00XX, NOVADI / 01XX, NOVADI / 06XX, NOVADI / 07XX, NOVADI / 09XX, NOVADI / 10XX
TeBr	10	Telamon	Braeburn	BETTER3FRUIT/KUL	Bianco et al. (2014)	

Supplementary table 3

Trait	Environment	c_g	c_c	c_r	$f_v(v)$	$f_u(u)$	$uh_v(v)$	$vh_u(u)$	$f_{u,v}(u, v)$	ED_s
Floral emergence	BEL	330.81	0.47	0.02	0	0	0	0	0.01	0.01
Floral emergence	FRA	451.32	1.73	9.56	0	0	1.73	0.61	0.07	2.41
Floral emergence	ITA	372.18	5.98	28.85	0	0	0	0.5	0.01	0.51
Floral emergence	ESP	492.07	3.45	0.03	2.95	2.44	0.87	0	5.16	11.42
Floral emergence	CHE	289.35	9.42	17.68	0	0	0.4	0.33	0	0.73
Harvest date	BEL	454.64	0	4.64	0	0.59	1.64	0	0	2.23
Harvest date	FRA	517.55	4.74	4.77	0.85	0	0.01	0	2.78	3.64
Harvest date	ITA	459.9	0	0.1	0.7	0.76	0	2.53	0	3.99
Harvest date	ESP	479.17	4.68	0.01	0	2.66	0	0	0	2.66
Harvest date	CHE	436.42	0	1.29	2.02	0	3.82	1.09	0	6.93

Supplementary table 4

SNP	Chromosome	Position	P.value
AX-115476635	11	11277966	2.02E-07
AX-115409139	9	681280	2.58E-07
AX-115198592	9	1078953	3.57E-07
AX-115410038	9	514635	7.30E-07
AX-115649181	9	1365711	2.34E-06
AX-115253961	11	11322737	2.65E-06
AX-115271245	2	13180576	4.15E-06
AX-115320293	11	11333668	4.30E-06
AX-115294802	6	30019604	1.01E-05
AX-115489653	6	4850252	1.07E-05
AX-115630047	9	530386	1.23E-05
AX-115424668	5	5989381	1.39E-05
AX-115200723	3	898531	2.06E-05
AX-115448780	9	713562	2.41E-05
AX-115405800	2	19950789	2.66E-05
AX-115293162	9	809693	2.83E-05
AX-115574727	6	29060380	2.85E-05
AX-115606670	5	4250031	3.44E-05
AX-115271653	17	21254231	3.75E-05
AX-115480906	3	152501	3.81E-05
AX-115413851	6	32333057	4.14E-05
AX-115375001	15	9950335	5.47E-05
AX-115358144	14	2051861	5.50E-05
AX-115375491	10	7852768	5.84E-05
AX-115413247	17	17157259	6.46E-05
AX-115659250	9	834337	6.97E-05
AX-115462290	10	7701768	7.01E-05
AX-115424792	9	897049	7.22E-05
AX-115509946	12	7456260	7.29E-05
AX-115395942	5	4298159	7.35E-05
AX-115325950	10	31221112	7.50E-05
AX-115438747	9	1079392	7.54E-05
AX-115641129	6	32325150	7.57E-05
AX-115537093	9	700032	7.86E-05
AX-115280605	9	6787582	8.36E-05
AX-115442530	9	740328	8.45E-05
AX-115409127	9	686455	8.52E-05
AX-115225691	17	19613564	8.69E-05
AX-115364676	9	733336	8.77E-05
AX-115263051	5	4539872	8.85E-05
AX-115248798	10	7694294	8.90E-05
AX-115626528	3	873611	9.14E-05
AX-115398042	17	19529047	9.16E-05
AX-115659049	9	1311208	9.39E-05
AX-115186296	12	7123996	9.40E-05
AX-115277096	3	117096	9.63E-05
AX-115624034	11	22368266	0.000100061
AX-115651254	17	1106421	0.00010048
AX-115489648	6	4848782	0.000100702
AX-115364881	17	207036	0.000102727
AX-115342473	4	9680012	0.000105213
AX-115652380	11	8157514	0.000106655
AX-115206028	10	7921602	0.000108087
AX-115455521	3	901698	0.000108991
AX-115315275	11	35922407	0.000115041
AX-115509805	6	30513600	0.000115183
AX-115252388	16	15218544	0.000118232
AX-115315768	3	465830	0.000118784
AX-115255461	9	922375	0.000123356

SNP	Chromosome	Position	P.value
AX-115425719	9	917784	0.000123717
AX-115631393	9	1680658	0.000127179
AX-115409123	9	690145	0.000130129
AX-105200711	12	22271000	0.000131734
AX-115644628	15	9949778	0.00013359
AX-115231767	17	19556765	0.000134472
AX-115559472	17	19567272	0.000134472
AX-115476992	9	7213924	0.000136205
AX-115635011	5	4850709	0.000137172
AX-115423441	17	22719604	0.000137334
AX-115573675	17	19457214	0.000138007
AX-115267901	6	33402022	0.00013804
AX-115390388	6	30148716	0.000138734
AX-115257133	9	1314934	0.000143546
AX-115189446	15	9374210	0.000143653
AX-115499603	14	26468205	0.00014508
AX-115506650	8	797503	0.000146988
AX-115639090	11	35922143	0.000150042
AX-115524062	15	23288953	0.000152795
AX-115195302	9	6702999	0.000154855
AX-115383113	17	37796	0.000155357
AX-115308249	5	42624991	0.000155724
AX-115480902	3	150621	0.000157893
AX-115425721	9	918327	0.000158403
AX-115264997	13	23172600	0.000158666
AX-115340860	9	363423	0.000164922
AX-115435194	3	919699	0.000165532
AX-115305148	3	27424583	0.000165667
AX-115364674	9	732953	0.000166918
AX-115393402	3	457780	0.000172928
AX-115514951	17	12969651	0.000173161
AX-115390389	6	30148756	0.000173634
AX-115342474	4	9679939	0.000179559
AX-115655386	9	425367	0.000183772
AX-115634292	6	24957968	0.000187343
AX-115480578	2	1513518	0.000188725
AX-115251645	9	284467	0.000191663
AX-115246876	7	35449965	0.000192354
AX-115634977	5	41827741	0.000195583
AX-115442234	3	272381	0.000197724
AX-115435196	3	919991	0.000197949

Supplementary table 5

SNP	Chromosome	Position	P.value
AX-115366114	3	30681581	2.47E-31
AX-115350952	16	8932940	3.31E-12
AX-115296686	10	38808504	1.07E-08
AX-115193921	3	25342871	6.25E-08
AX-115610924	3	27890504	3.86E-07
AX-115267785	3	28196675	4.61E-07
AX-115194965	3	28196867	4.61E-07
AX-115408872	3	28211584	5.20E-06
AX-105215757	3	28710671	8.61E-06
AX-115332166	3	30522111	8.79E-06
AX-115267783	3	28197544	1.21E-05
AX-115323519	3	3974709	1.35E-05
AX-115323522	3	3976063	1.45E-05
AX-115267794	3	28191292	1.83E-05
AX-115350519	3	28225257	2.00E-05
AX-115382132	15	16282258	2.36E-05
AX-115494676	15	16543549	2.67E-05
AX-115440163	2	31757614	3.04E-05
AX-115408866	3	28209841	3.10E-05
AX-115451439	13	8734863	3.31E-05
AX-115271556	3	28713955	3.80E-05
AX-115589269	3	28356736	3.94E-05
AX-115394105	3	30566958	4.18E-05
AX-115325035	15	15974089	4.20E-05
AX-115249524	1	23157595	4.62E-05
AX-115571663	14	19433330	4.65E-05
AX-115382135	15	16275614	5.14E-05
AX-115475033	3	27662815	5.26E-05
AX-115517369	15	15898261	5.40E-05
AX-115349967	10	32302414	5.60E-05
AX-115306487	11	32417830	5.72E-05
AX-115194969	3	28194438	6.12E-05
AX-115495480	10	1714411	7.41E-05
AX-115271564	3	28716603	7.42E-05
AX-115382124	15	16290612	7.69E-05
AX-115589264	3	28353807	7.83E-05
AX-115228351	12	24553230	8.26E-05
AX-115585030	3	28327507	8.28E-05
AX-115325049	15	15965285	8.62E-05
AX-115585026	3	28324807	8.84E-05
AX-115214826	2	32732031	8.86E-05
AX-115382114	15	16308333	8.95E-05
AX-115382109	15	16309660	8.95E-05
AX-115265665	15	16589887	9.95E-05
AX-115290203	15	16344145	0.000101575
AX-115390916	11	32563374	0.000104877
AX-115382099	15	16315365	0.000109633
AX-115335312	15	16251437	0.000110102
AX-115312018	15	16257792	0.000110102
AX-115382129	15	16286615	0.000110102
AX-115424522	2	13090152	0.000111453
AX-115373926	3	28120363	0.000115418
AX-115341567	15	15879594	0.000117125
AX-115517377	15	15890668	0.000117125
AX-115305419	15	16025305	0.000117125
AX-115475016	3	27644712	0.000122172
AX-115441737	15	16330184	0.000122353
AX-115469686	13	9019850	0.000123996
AX-115382142	15	16272123	0.000125483

SNP	Chromosome	Position	P.value
AX-115382141	15	16272282	0.000125483
AX-115325043	15	15960256	0.000125844
AX-115408870	3	28210596	0.000126056
AX-115219544	15	16068979	0.000126135
AX-115453753	14	19418814	0.000126638
AX-115290206	15	16343126	0.000126872
AX-115649770	15	15712727	0.000128947
AX-115267786	3	28196171	0.000129672
AX-115496550	3	27639600	0.000131182
AX-115511599	3	4905406	0.000133303
AX-115290818	7	27464315	0.000136126
AX-115636644	14	2927472	0.000138108
AX-115283115	15	16460881	0.000145929
AX-115409564	3	4933387	0.000149175
AX-115335173	15	16414293	0.000157476
AX-115589256	3	28348021	0.000157631
AX-115319081	12	16790040	0.000160079
AX-115305173	12	16984299	0.00017416
AX-115439319	12	17277180	0.000174634
AX-115480177	14	2894345	0.000175288
AX-115312612	17	34466225	0.000180303
AX-115511582	3	4893424	0.000188649
AX-115382648	15	16755728	0.000191962
AX-115556899	2	13275475	0.00019294
AX-115289711	3	29592255	0.000193645
AX-115315391	3	25120784	0.000196641
AX-115589257	3	28348396	0.00019807
AX-115208348	13	8869377	0.000199053
AX-115348786	11	32499704	0.000207056
AX-115350530	3	28260114	0.000225459
AX-115350524	3	28257357	0.000225915
AX-115441728	15	16327426	0.000227823
AX-115208331	13	8889817	0.000228517
AX-115197115	11	32063918	0.000231565
AX-115315392	3	25120334	0.000233251
AX-115375083	2	32935499	0.000236573
AX-115373932	3	28116535	0.000238279
AX-115275372	11	11308254	0.000244765
AX-115256894	15	16776153	0.000245191
AX-115438655	2	13371090	0.000248806
AX-115438653	2	13373422	0.000248806

Supplementary table 6

Genotype code	Floral emergence	Harvest date
1	113.97	249.42
101	102.02	270.02
1015	108.44	251.34
1020	109.53	260.69
1091	116.73	236.41
1113	110.51	257.01
112	108.00	273.35
1123	112.57	282.41
1125	105.02	287.35
114	106.59	250.34
118	106.68	272.06
12_196	104.78	270.85
12_197	107.25	264.50
12_204	107.98	262.96
12_205	108.77	268.23
12_B006	107.04	246.74
12_B020	107.90	261.66
12_B023	109.97	255.64
12_B057	107.37	262.79
12_B061	107.68	258.35
12_B064	105.84	259.19
12_B074	109.00	261.52
12_B081	108.47	270.66
12_B093	107.17	255.82
12_B095	110.41	249.04
12_E003	103.78	238.13
12_E014	106.45	267.05
12_E015	110.13	267.31
12_E017	107.25	259.77
12_E024	110.47	281.27
12_E025	111.02	291.27
12_E034	109.27	272.41
12_E035	108.58	277.03
12_E039	109.65	291.59
12_E043	104.54	241.77
12_F002	110.34	250.95
12_F007	110.44	271.60
12_F014	112.35	258.49
12_F019	108.82	269.93
12_F023	110.72	272.13
12_F040	111.24	252.37
12_F045	111.20	250.13
12_F046	109.89	259.91
12_F052	110.25	263.26
12_F054	108.79	263.74
12_I002	107.03	264.18
12_I003	110.37	260.57
12_I005	108.76	239.48
12_I010	110.22	234.83
12_I037	108.60	228.61
12_I049	108.98	268.69
12_I056	105.73	267.38
12_I059	103.93	239.14
12_I062	108.73	261.96
12_I064	110.38	261.81
12_J001	111.02	258.58
12_J004	110.37	267.43
12_J011	108.86	260.80
12_J013	107.35	261.15

Genotype code	Floral emergence	Harvest date
12_J015	112.11	250.33
12_J019	110.46	271.11
12_J027	109.31	275.62
12_J028	109.83	276.82
12_J029	110.96	272.72
12_J031	109.14	264.50
12_K001	109.81	240.80
12_K010	109.03	245.87
12_K016	108.12	261.81
12_K018	110.74	267.81
12_K023	109.45	282.62
12_K024	106.94	270.32
12_K044	106.26	277.76
12_K045	109.35	275.29
12_K049	111.99	264.87
12_L001	109.46	258.78
12_L005	107.13	265.69
12_L012	109.24	262.79
12_L013	107.13	269.93
12_L018	109.36	267.06
12_L021	106.35	275.94
12_N021	108.61	282.28
12_N034	109.62	277.97
12_N039	109.42	289.40
12_N041	106.10	242.67
12_N046	109.55	260.33
12_N052	106.65	255.80
12_N054	106.93	253.84
12_N057	108.52	258.02
12_N062	106.30	239.72
12_N063	107.84	229.53
12_O010	111.79	279.10
12_O011	108.75	269.78
12_O014	110.97	279.20
12_O017	112.19	283.31
12_O026	111.50	288.58
12_O056	108.40	261.89
12_O061	109.35	259.62
12_O063	110.25	271.99
12_O072	108.62	250.65
12_O074	110.88	281.67
12_P001	110.18	280.68
12_P002	109.72	281.59
12_P003	109.58	283.88
12_P018	113.09	278.74
12_P025	106.27	271.59
12_P027	107.13	291.43
12_P029	110.35	275.59
12_P034	109.32	259.20
12_P055	108.92	292.67
12_P056	110.40	273.13
1218	106.80	236.61
1236	107.64	247.99
13_10	105.95	270.90
13_109	105.07	221.34
13_139	101.64	258.02
13_32	105.47	251.01
13_35	105.55	238.51
13_38	106.52	277.00
130	108.15	239.51
131	114.63	279.91

Genotype code	Floral emergence	Harvest date
133	110.06	280.89
135	108.04	283.72
1377	103.92	231.75
140	109.15	257.78
141	110.06	251.88
143	111.72	223.04
1446	112.86	278.31
1448	106.77	222.29
1478	107.62	211.93
1481	103.29	284.39
149	110.35	262.51
15	111.73	255.55
152	108.04	249.87
163	109.35	244.06
167	105.95	197.33
17	109.16	275.96
170	115.24	275.71
173	110.88	259.85
1870	110.56	236.37
1872	104.37	219.70
1887	112.96	212.61
1889	112.06	214.89
1893	106.77	237.91
1900	114.88	251.95
1901	105.67	215.60
1910	102.40	251.97
192	113.13	265.75
1935	108.95	277.02
1951	108.90	271.21
1952	109.74	261.77
1962	109.18	279.69
197	108.46	268.14
1985	107.98	212.15
1988	106.97	227.09
1989	110.67	261.43
1991	108.30	205.23
1997	108.96	225.96
2	105.23	276.66
200	106.74	268.71
2003	117.49	215.36
2010	107.63	221.49
2013	108.96	245.98
202	113.22	271.38
2024	109.12	251.59
2026	107.34	213.66
203	105.53	236.49
204	107.89	216.60
2046	108.64	249.70
2050	108.86	270.80
2055	103.40	199.02
206	104.55	273.00
2061	107.70	264.31
2071	112.13	271.81
216	109.48	280.63
236	103.37	195.67
238	105.95	205.22
241	105.55	223.86
2410	105.69	261.20
2422	109.13	253.58
2443	108.61	228.97
249	108.81	257.53

Genotype code	Floral emergence	Harvest date
25	108.68	204.60
26	106.37	212.32
261	106.54	222.34
262	107.65	220.44
2625	105.72	282.60
2651	104.95	199.38
2652	106.63	236.69
2653	109.28	248.90
267	104.74	227.77
2670	104.43	222.98
2671	108.14	258.06
270	106.31	209.59
2720	108.92	247.68
2722	109.77	218.39
2732	108.86	231.15
2762	107.58	280.34
2779	118.29	272.52
2784	110.42	276.39
2785	112.46	280.18
2794	106.77	232.43
2796	106.29	233.66
2802	109.36	270.74
2809	105.42	266.59
2810	105.83	278.38
2813	106.59	273.04
2817	99.58	250.44
2820	107.05	276.17
2823	111.78	244.99
2827	110.27	238.59
2830	109.21	250.83
2831	109.05	234.86
2836	112.68	221.65
2847	108.52	261.58
2848	110.42	236.13
2849	105.51	287.82
2850	110.00	256.70
2855	106.58	230.33
2856	108.48	260.22
286	106.34	213.42
2862	103.45	274.15
2865	106.87	221.44
2872	110.52	270.69
29	104.88	212.00
290	108.74	238.68
2902	108.78	251.32
2908	102.79	197.02
2936	105.31	216.21
2971	109.27	272.05
2974	110.25	235.72
2975	108.45	267.81
2979	108.91	264.52
2980	113.27	247.46
2990	110.38	266.40
2992	105.47	242.37
2997	106.24	282.71
30	107.39	215.09
301	132.43	233.88
3010	104.64	261.45
3013	109.88	269.93
3026	109.26	238.58
3042	102.77	256.19

Genotype code	Floral emergence	Harvest date
3062	105.59	210.11
308	106.85	220.10
3148	108.95	236.37
3160	107.00	250.06
317	117.63	270.31
3170	113.16	222.81
3172	102.84	211.83
3173	109.91	239.44
3190	107.98	261.57
32	109.29	230.68
321	108.51	276.00
323	113.05	278.06
324	112.27	237.59
327	109.62	274.31
33	108.52	281.56
330	107.11	273.98
333	109.05	245.81
334	113.50	268.86
338	110.65	269.74
34	110.02	234.88
345	106.27	212.73
353	110.72	270.59
366	105.46	196.77
37	110.17	243.36
378	106.65	256.49
40	108.04	213.85
405	109.76	257.21
409	108.93	288.99
410	106.57	280.98
424	110.64	274.74
428	109.90	272.60
43	111.49	253.27
435	110.12	274.32
440	103.92	268.70
445	112.35	287.79
447	110.87	267.23
45	112.03	247.42
457	112.06	267.47
46	115.94	237.24
462	107.71	268.26
465	110.84	285.92
468	104.62	196.65
472	107.96	241.46
473	107.64	237.34
483	106.76	205.11
488	112.46	234.92
489	110.81	212.19
49	108.05	245.13
491	110.38	265.99
5	107.40	216.35
50	106.80	266.22
508	105.68	249.00
51	109.46	232.36
515	103.09	267.10
52	106.31	200.28
522	109.11	266.85
526	110.43	254.82
533	106.64	221.58
543	110.05	219.45
546	111.92	284.58
548	107.67	290.14

Genotype code	Floral emergence	Harvest date
549	110.02	232.08
55	107.04	276.78
56	108.81	258.17
560	108.89	247.28
562	106.68	277.55
564	109.73	241.97
565	108.70	212.67
567	112.81	269.76
57	110.38	254.04
570	112.00	244.44
575	106.81	229.30
577	107.76	256.14
58	112.37	226.06
582	109.68	255.28
584	108.83	242.07
587	109.11	250.02
592	110.04	229.98
6	108.43	232.95
601	108.67	215.87
603	106.59	202.93
61	107.63	254.01
629	103.90	262.77
639	103.99	240.20
65	108.64	264.41
652	105.93	224.48
653	107.52	231.69
66	111.57	254.44
662	106.82	290.19
67	122.11	266.84
670	108.82	235.48
673	113.18	272.19
676	103.91	211.38
678	106.37	254.36
68	107.54	241.33
681	116.02	282.25
691	108.03	275.35
697	105.78	271.11
71	109.70	253.32
72	113.82	231.19
722	107.62	233.53
737	104.36	222.41
739	107.95	232.15
74	113.27	266.26
740	111.67	260.39
743	106.92	237.42
748	110.53	290.15
75	108.64	203.25
750	106.68	263.31
757	113.66	272.19
764	109.98	258.76
77	106.03	261.26
779	105.22	240.30
786	110.33	262.74
787	108.20	287.24
8	104.47	250.48
80	107.25	221.25
809	113.14	260.78
81	108.26	213.93
810	106.13	265.99
82	103.54	198.58
89	108.72	244.97

Genotype code	Floral emergence	Harvest date
9	110.58	264.97
918	108.54	255.11
929	107.53	229.95
93	110.57	221.92
94	108.89	221.83
95	110.94	266.47
950	108.64	219.11
953	110.51	268.82
97	107.51	258.15
997	108.23	249.30
ACW_25358	108.57	255.88
ACW_25363	108.82	274.84
ACW_25365	110.45	260.58
ACW_25366	109.56	263.50
ACW_25368	111.09	267.07
ACW_25371	108.64	264.84
ACW_25372	109.30	272.07
ACW_25375	107.57	273.73
ACW_25454	107.26	270.71
ACW_25469	107.92	262.40
ACW_25580	106.86	274.37
ACW_25581	108.44	262.55
ACW_25584	107.68	252.36
ACW_25586	108.97	243.98
ACW_25591	105.36	256.23
ACW_25592	106.06	259.26
ACW_25607	109.47	265.88
ACW_25619	105.63	257.23
ACW_25622	108.12	256.97
ACW_25628	107.71	255.10
DLO_12_066	108.47	234.55
DLO_12_084	110.11	232.69
DLO_12_120	108.60	265.50
DLO_12_140	110.31	244.14
DLO_12_141	109.24	247.12
DLO_12_162	107.17	235.16
DLO_12_221	111.96	270.63
DLO_12_233	108.64	244.32
DLO_12_259	108.28	265.46
DLO_12_289	108.64	267.97
FuGa_015	109.84	270.22
FuGa_016	108.89	218.68
FuGa_140	109.87	241.55
FuGa_187	108.49	279.53
FuGa_215	107.98	264.56
FuGa_217	110.00	264.17
FuGa_243	109.24	271.73
FuGa_270	108.35	268.51
FuGa_275	106.93	273.10
FuGa_352	109.04	255.54
FuPi_005	110.62	267.64
FuPi_013	106.75	282.19
FuPi_023	106.25	245.13
FuPi_052	108.03	275.21
FuPi_054	110.69	282.59
FuPi_087	110.02	256.52
FuPi_097	106.75	270.21
FuPi_106	108.75	273.11
FuPi_107	109.62	276.19
FuPi_114	109.20	256.81
GaCr_002	107.64	277.67

Genotype code	Floral emergence	Harvest date
GaCr_010	106.44	251.64
GaCr_017	107.59	280.60
GaCr_041	106.71	290.73
GaCr_043	108.04	262.70
GaCr_052	105.88	260.78
GaCr_055	108.75	273.00
GaCr_056	109.68	259.42
GaCr_057	108.31	269.73
GaCr_061	107.03	255.51
GaPi_007	109.61	255.91
GaPi_014	108.93	227.36
GaPi_017	109.45	254.42
GaPi_023	109.11	239.85
GaPi_028	109.64	277.35
GaPi_033	108.78	223.32
GaPi_035	109.10	219.20
GaPi_038	109.07	232.74
GaPi_051	108.61	253.47
GaPi_056	110.25	244.77
I_BB002	110.74	271.51
I_BB012	108.63	256.92
I_BB015	108.46	259.86
I_BB022	110.94	279.82
I_BB028	110.43	273.74
I_BB031	109.35	249.59
I_BB039	107.55	261.01
I_BB049	109.78	261.40
I_BB056	106.82	241.06
I_BB070	107.63	272.63
I_CC003	108.87	268.22
I_CC044	109.47	281.94
I_CC048	108.29	275.54
I_CC052	107.64	253.69
I_CC055	106.88	272.79
I_CC057	109.46	251.27
I_CC059	109.94	250.24
I_CC067	110.53	279.85
I_CC070	108.88	264.48
I_CC075	110.35	282.05
I_J012	110.28	244.45
I_J030	108.72	263.57
I_J033	108.87	213.34
I_J036	110.61	269.85
I_J046	108.74	266.13
I_J054	107.53	250.52
I_J056	107.12	248.61
I_J063	108.80	243.65
I_J064	109.30	242.39
I_J069	106.98	252.44
I_M007	103.68	275.96
I_M013	106.99	252.17
I_M014	103.78	272.56
I_M024	104.77	253.16
I_M026	109.15	269.68
I_M029	107.06	241.48
I_M047	108.16	242.62
I_M048	104.19	263.87
I_M052	107.77	251.73
I_M056	105.69	237.42
I_W002	103.76	260.70
I_W010	108.30	255.81

Genotype code	Floral emergence	Harvest date
I_W034	105.53	260.86
I_W036	109.14	264.82
I_W037	109.67	273.07
I_W044	106.91	255.53
I_W045	110.49	257.09
I_W046	108.00	259.09
I_W049	106.41	269.23
I_W050	110.55	274.45
JoPr_1990_102_001	107.87	234.10
JoPr_1990_102_054	105.75	226.45
JoPr_1990_102_097	104.75	219.40
JoPr_1990_102_108	106.47	243.99
JoPr_1990_102_169	107.15	230.61
JoPr_1990_102_179	107.92	228.55
JoPr_1990_102_202	106.71	230.01
JoPr_1990_102_232	107.41	224.76
JoPr_1990_102_242	103.92	239.83
JoPr_1990_102_246	110.95	238.70
JoPr_1990_102_257	105.48	236.37
NOVADI_0010	109.42	250.50
NOVADI_0023	109.06	271.36
NOVADI_0042	107.46	268.98
NOVADI_0114	107.29	280.48
NOVADI_0256	113.64	275.64
NOVADI_0268	115.23	273.57
NOVADI_0286	110.21	241.96
NOVADI_0296	107.81	257.76
NOVADI_0329	110.50	252.62
NOVADI_0330	109.94	271.70
NOVADI_0347	109.78	263.41
NOVADI_0356	109.98	266.73
NOVADI_0371	110.28	266.68
NOVADI_0382	108.90	265.75
NOVADI_0622	106.91	262.49
NOVADI_0731	108.59	284.10
NOVADI_0766	105.36	249.10
NOVADI_0980	107.97	270.11
NOVADI_1015	105.50	233.87
TeBr_012	105.85	261.05
TeBr_021	106.97	268.59
TeBr_022	106.00	272.09
TeBr_027	104.54	263.08
TeBr_043	104.49	248.10
TeBr_071	107.44	274.42
TeBr_084	103.92	247.59
TeBr_180	107.55	259.61
TeBr_248	107.58	276.12
TeBr_263	108.82	262.42

3 Genetic architecture and genomic prediction accuracy of apple quantitative traits across environments

Jung, Michaela[1,2]; Keller, Beat[1,2]; Roth, Morgane[1,3]; Aranzana, Maria José[4,5]; Auwerkerken, Annemarie[6]; Guerra, Walter[7]; Al-Rifaï, Mehdi[8]; Lewandowski, Mariusz[9]; Sanin, Nadia[7]; Rymenants, Marijn[6,10]; Didelot, Frédérique[11]; Dujak, Christian[5]; Font i Forcada, Carolina[4]; Knauf, Andrea[1,2]; Laurens, François[8]; Studer, Bruno[2]; Muranty, Hélène[8]; Patocchi, Andrea[1]

[1]Breeding Research group, Agroscope, 8820 Wädenswil, Switzerland
[2]Molecular Plant Breeding, Institute of Agricultural Sciences, ETH Zurich, 8092 Zurich, Switzerland
[3]GAFL, INRAE, 84140 Montfavet, France
[4]IRTA (Institut de Recerca i Tecnologia Agroalimentàries), 08140 Caldes de Montbui, Barcelona, Spain
[5]Centre for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Campus UAB, 08193 Bellaterra, Barcelona, Spain
[6]Better3fruit N.V., 3202 Rillaar, Belgium
[7]Research Centre Laimburg, 39040 Auer, Italy
[8]Univ Angers, Institut Agro, INRAE, IRHS, SFR QuaSaV, F-49000 Angers, France
[9]The National Institute of Horticultural Research, Konstytucji 3 Maja 1/3, 96-100 Skierniewice, Poland
[10]Laboratory for Plant Genetics and Crop Improvement, KU Leuven, B-3001, Leuven, Belgium
[11]Unité expérimentale Horticole, INRAE, F-49000 Angers, France

Material from: 'JUNG et al., GENETIC ARCHITECTURE AND GENOMIC PREDICTION ACCURACY OF APPLE QUANTITATIVE TRAITS ACROSS ENVIRONMENTS, bioRxiv, published 2021, doi: https://doi.org/10.1101/2021.11.29.470309'

Abstract

Implementation of genomic tools is desirable to increase the efficiency of apple breeding. The apple reference population (apple REFPOP) proved useful for rediscovering loci, estimating genomic prediction accuracy, and studying genotype by environment interactions (G×E). Here we show contrasting genetic architecture and genomic prediction accuracies for 30 quantitative traits across up to six European locations using the apple REFPOP. A total of 59 stable and 277 location-specific associations were found using GWAS, 69.2% of which are novel when compared with 41 reviewed publications. Average genomic prediction accuracies of 0.18–0.88 were estimated using single-environment univariate, single-environment multivariate, multi-environment univariate, and multi-environment multivariate models. The G×E accounted for up to 24% of the phenotypic variability. This most comprehensive genomic study in apple in terms of trait-environment combinations provided knowledge of trait biology and prediction models that can be readily applied for marker-assisted or genomic selection, thus facilitating increased breeding efficiency.

3.1 Introduction

Apple (*Malus domestica* Borkh.) is the third most produced fruit crop worldwide[1]. Since its domestication in the Tian Shan mountains of Central Asia, the cultivated apple developed into a separated near-panmictic species[2]. Over the centuries, thousands of apple cultivars have been raised and conserved thanks to grafting[3]. Extensive relatedness among cultivars with a strong influence of a few founders through the history of apple breeding has been reported despite their high genetic diversity[4-6]. Only a fraction of the existing cultivars are grown commercialy[3] and they require an intensive use of pesticides for crop protection. To diversify apple production, it is desirable to produce new cultivars for sustainable intensive agriculture and adapted to future climate, while remaining attractive to consumers.

Apple breeding is labor- and time-intensive, but selection efficiency can be improved by integrating DNA-informed techniques into the breeding process[7]. Marker-assisted selection allows breeders to predict the value of a target trait based on its association with a genetic marker. The method leads to removal of inferior seedlings without phenotyping, thus reducing the labor costs when decreasing the number of individuals passing to the next selection step[7]. Quantitative trait locus (QTL) mapping has been traditionally used to investigate the genetic basis of variation in traits such as pathogen resistance, phenology, and some fruit quality traits[8-11]. To bridge the gap between the discovery of marker-trait associations and their application in breeding, protocols that transfer the knowledge obtained by QTL analyses into DNA tests were established[12,13]. However, marker-assisted selection in apple remains restricted to a limited number of traits associated with single genes or a handful of large-effect QTL, such as pathogen resistance and fruit firmness, acidity, or color[14]. DNA-informed selection is rarely deployed in apple when breeding for quantitative traits with complex genetic architecture, though this task became feasible with the recent technological developments in apple genomics.

In the genomics era, advancements in genotyping and sequencing technologies led to a broad range of new tools for genetic analyses. In the case of apple, several reference genomes have been produced[15-19], single nucleotide polymorphism (SNP) genotyping arrays of different densities such as 20K or 480K SNPs have been developed[20,21], and genotyping-by-sequencing methods have been adopted[22,23]. Genome-wide association study (GWAS) emerged as a method for exploring the genetic basis of quantitative traits[24]. GWAS in apple have been used to identify associations between markers and various traits such as fruit quality and phenology traits[22,23,25-29]. The associations found with GWAS can be translated into DNA tests for marker-assisted selection. Besides GWAS, genomic selection was developed to exploit the effects of genome-wide variation at loci of both large and low effect on quantitative traits using a single model[30] and is sometimes called marker-assisted selection on a genome-wide scale[31]. For genomic selection, prediction models are first trained with phenotypic and genomic data of a training population. In a second step, the models predict the performance of breeding material based on the genomic data alone. These genomic estimated breeding values are then used to make selections among the breeding material, thus increasing the breeding efficiency and genetic gain. Several studies have assessed genomic prediction accuracy for apple quantitative traits related to fruit quality and phenology[22,23,29,32-36]. Genomic selection can double genetic gain, as demonstrated by yield traits in dairy cattle[37], but the accuracy of

genomic prediction for yield traits in apple has not been studied. Analyses of genomic datasets beyond 100K SNPs have been limited to flowering and harvest time (GWAS and genomic prediction)[26,36], fruit firmness and skin color (GWAS)[28,38]. Marker density, trait architecture, and heritability have been shown to differentially affect prediction performance in simulated data and for apple[34,36,39] and their impact on genomic analyses should therefore be further empirically tested. Moreover, GWAS for the same traits measured at different locations, the effect of genotype by environment interaction (G×E) on genomic prediction accuracy, and predictions with multivariate genomic prediction models have not been evaluated yet in apple.

Plants are known for their strong phenotypic response to environmental factors, a phenomenon regularly tested in plant breeding using multi-environment trials. In general, when statistical models are applied to measurements from multi-environment trials, the effect of environment on individuals remains constant at single locations, but the G×E leads to changes in the ranking of genotypes across locations. With an increasing proportion of G×E effect relative to genotypic effect, both heritability and response to selection decrease[40]. A noticeable effect of contrasting European environments and G×E on two apple phenology traits – floral emergence and harvest date – has been reported, which demands testing the multi-environment modelling approaches in apple[36]. A location-specific GWAS may be used to identify loci with stable effects across environments and loci specific to individual locations[41]. Multi-environment prediction models can account for G×E by explicitly modeling interactions between all available markers and environments[42]. Borrowing information from other genotypes across environments through markers, the G×E method can outperform more simple modelling approaches that ignore G×E[42-44]. Additionally, taking advantage of information that traits provide about one another, a multivariate (also called multi-trait) genomic prediction can be applied. This method may be useful in case the assessment of one trait remains costly, but another correlated trait with less expensive measurement is available or can be assessed more easily[45]. The multivariate prediction can also be extended to a multi-environment approach when treating measurements from different environments as distinct traits[46].

A population of 269 diverse apple accessions from across the globe and 265 progeny from 27 parental combinations originating in recent European breeding programs constitutes our apple reference population (apple REFPOP)[36]. The apple REFPOP has a high-density genomic dataset of 303K SNPs and was deemed suitable for the application of genomics-assisted breeding[36]. Combined with extensive phenotypic information, the apple REFPOP provides the groundwork for marker-assisted and genomic selection across contrasting European environments. Hence, 30 traits related to productivity, tree vigor, phenology, and fruit quality were measured in the apple REFPOP during up to three years and at up to six locations with various climatic conditions of Europe (Belgium, France, Italy, Poland, Spain, and Switzerland). First, GWAS was performed to dissect the genetic architecture of the studied traits, identify associated loci stable across locations and location-specific loci, and to observe signs of selection on loci of large effect. Second, this study aimed to measure prediction accuracy for these traits using single-environment univariate, single-environment multivariate, multi-environment univariate, and multi-environment multivariate genomic prediction models. Finally, a critical analysis of our results provided

recommendations for future implementation of genomic prediction tools in apple breeding.

3.2 Materials and methods

3.2.1 Plant material

The apple REFPOP was designed and established by the collaborators of the FruitBreedomics project[47] as described by Jung et al.[36]. The apple REFPOP consists of 534 genotypes from two groups of diploid germplasm. The accession group consists of 269 accessions of European and non-European origin representing the diversity in cultivated apple. The progeny group of 265 genotypes stemmed from 27 parental combinations produced in the current European breeding programs. In 2016, the apple REFPOP was planted in six locations representing several biogeographical regions in Europe, in (i) Rillaar, Belgium, (ii) Angers, France, (iii) Laimburg, Italy, (iv) Skierniewice, Poland, (v) Lleida, Spain and (vi) Wädenswil, Switzerland (one location per country). Every genotype was replicated at least twice per location. All plants included in this study were treated with agricultural practice common to each location. Calcium spraying was avoided due to its influence on bitter pit. Flowers were not thinned, but the fruits were hand-thinned after the June fruit drop and up to two apples per fruit cluster were retained.

3.2.2 Genotyping

A high-density genome-wide SNP marker dataset was produced as reported by Jung et al.[36]. Briefly, SNPs from two overlapping SNP arrays of different resolution, (i) the Illumina Infinium® 20K SNP genotyping array[20] and (ii) the Affymetrix Axiom® Apple 480K SNP genotyping array[21], were curated and then joined applying imputation with Beagle 4.0[48] using the recently inferred pedigrees[4]. Non-polymorphic markers were removed to obtain a set of 303,148 biallelic SNPs. Positions of SNPs were based on the apple reference genome obtained from the doubled haploid GDDH13 (v1.1)[16].

3.2.3 Phenotyping

Thirty phenotypic traits related to phenology, productivity, fruit size, outer fruit, inner fruit, and vigor were evaluated at up to six locations of the apple REFPOP during up to three seasons (2018–2020). Trunk diameter was measured in 2017 in some locations, enabling for a trunk increment calculation for 2018. The traits were recorded as described in the Supplementary Methods. Two phenology traits measured in 2018, i.e., floral emergence and harvest date, were previously analyzed by Jung et al.[36].

3.2.4 Phenotypic data analyses

Spatial heterogeneity was modeled separately for each trait and environment (nested factor of location and year) using the spatial analysis of field trials with splines (SpATS) to account for the replicate effects and differences due to soil characteristics[49]. Phenotypic values of traits adjusted for spatial heterogeneity within each environment were estimated at the level of trees (adjusted phenotypic values of each tree) and genotypes (adjusted phenotypic values of each genotype)[36].

The general statistical model for the following phenotypic data analyses fitted via restricted maximum likelihood (R package lme4[50]) was:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zb} + \boldsymbol{\varepsilon} \text{ (Equation 1)}$$

where $\boldsymbol{y}$ was a vector of trait phenotypes, $\boldsymbol{X}$ the design matrix for the fixed effects, $\boldsymbol{\beta}$ the vector of fixed effects, $\boldsymbol{Z}$ the design matrix for the random effects, $\boldsymbol{b}$ the vector of random effects and $\boldsymbol{\varepsilon}$ the vector of random errors. The $\boldsymbol{b}$ was a $q \times 1$ vector assuming $\boldsymbol{b} \sim N(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ was a variance-covariance matrix of the random effects. The assumptions for the $N \times 1$ vector of random errors were $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$ with $N \times N$ identity matrix $\boldsymbol{I}$ and the variance σ_ε^2, the N being the number of trees.

To assess the reliability of environment-specific data, a random-effects model was first fitted separately for each trait and environment to estimate an environment-specific clonal mean heritability. Applying the Equation 1, the response $\boldsymbol{y}$ was a vector of the raw (non-adjusted) phenotypic values of each tree. On the place of $\boldsymbol{X}$, a vector of ones was used to model the intercept β. The vector of genotypes acted as a random effect in $\boldsymbol{Z}$. The environment-specific clonal mean heritability was calculated from the variance components of the random-effects model as:

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2} \text{ (Equation 2)}$$

where the phenotypic variance $\sigma_p^2 = \sigma_g^2 + \sigma_\varepsilon^2/\bar{n}_r$ was obtained from the genotypic variance σ_g^2, error variance σ_ε^2 and the mean number of genotype replications $\bar{n}_r$. The environment-specific clonal mean heritability was used to eliminate location-year-trait combinations with a heritability value below 0.1.

For the remaining location-year combinations, a mixed-effects model following the Equation 1 was fitted to the vector of the adjusted phenotypic values of each tree as response ($\boldsymbol{y}$). The effects of environments, i.e., combination of location and years, were used as fixed effects and the effects of genotypes and genotype by environment interactions as random effects. Estimated variances of the model components were used to evaluate the across-environment clonal mean heritability calculated using the Equation 2 with the phenotypic variance estimated as:

$$\sigma_p^2 = \sigma_g^2 + \frac{\sigma_{ge}^2}{n_e} + \frac{\sigma_\varepsilon^2}{n_e \bar{n}_r} \text{ (Equation 3)}$$

where σ_{ge}^2 was the genotype by environment interaction variance and n_e represented the number of environments.

An additional mixed-effects model following the Equation 1 was fitted to the adjusted phenotypic values of each tree ($\boldsymbol{y}$) using the effects of location, year and their interaction as fixed effects and the effects of genotypes as random effects. Due to the skewness of their distributions, $\boldsymbol{y}$-values of the traits weight of fruits, number of fruits and trunk diameter were log-transformed. BLUPs ($\widehat{\boldsymbol{b}}$) extracted from the model were further denoted as across-location BLUPs. To estimate the location-specific BLUPs, a model according to the Equation 1 was fitted with a subset of the adjusted phenotypic values of each tree from single locations ($\boldsymbol{y}$) using the effects of years as fixed effects and the effects of genotypes as random effects. The across-location BLUPs and the adjusted phenotypic values of each genotype were used to assess phenotypic correlation as the Pearson correlation between pairs of traits and between pairs of environments within traits, respectively. The across-location BLUPs with the addition of location-specific BLUPs for traits measured at a single location were further denoted as the main BLUPs. In the main BLUPs, the missing values

were replaced with the mean of the BLUPs of the same trait and the data were scaled and centered to finally estimate a principal component analysis biplot[51], where multivariate normal distribution was assumed for the ellipses.

3.2.5 Genome-wide association studies

The Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK)[52] implemented in the R package GAPIT 3.0[53] was applied using the genomic matrix $\boldsymbol{M}$, an $n \times m$ matrix for a population of size $n = 534$ genotypes (i.e., accessions and progeny) with $m = 303{,}148$ markers, with across-location BLUPs (across-location GWAS) or location-specific BLUPs (location-specific GWAS) as the response. BLINK was used with two principal components and the minor allele frequency threshold was set to 0.05. Marker-trait associations were identified as significant for p-values falling below a Bonferroni-corrected significance threshold $\alpha^* = \alpha/m$ with $\alpha = 0.05$ $(-log_{10}(p) > 6.74)$. The proportion of phenotypic variance explained by each significantly associated SNP was assessed with a coefficient of determination (R^2). The R^2 was estimated from a linear regression model, which was fitted with a vector of SNP marker values (coded as 1, 2, 3) as predictor and either the across-location BLUPs or location-specific BLUPs as response. GWAS based on the across-location BLUPs with the addition of location-specific BLUPs, in cases where traits were measured at a single location only, was further denoted as the global GWAS. The position of the last SNP on a chromosome was used to estimate chromosome length, which was used to divide each chromosome into three equal segments, i.e., top, center and bottom. The marker-trait associations were assigned to these chromosome segments based on their positions.

Previous reports on QTL mapping and GWAS in apple were reviewed to perform an extensive comparison with our GWAS results (Supplementary Table 4). Published results for traits measured similarly to the traits studied in the present work were considered, with the traits being assembled into trait groups: harvest time (harvest date and similar), flowering time (floral emergence, full flowering, end of flowering and similar), productivity (flowering intensity, weight of fruits, number of fruits and similar), fruit size (single fruit weight, fruit diameter, fruit length, maximum fruit size, fruit volume and similar), ground color (ground color, yellow color and similar), over color (red over color, green color and similar), bitter pit (bitter pit frequency, bitter pit grade and similar), russet (russet cover, russet frequency overall, at stalk, on cheek and in the eye and similar), acidity (titratable acidity and similar), sugar (soluble solids content and similar), firmness (fruit firmness and similar), water core (water core frequency, water core grade and similar) and trunk (trunk diameter, trunk increment and similar). The positions of published associations within respective chromosomes were visually assigned to the three chromosome segments, i.e., top, center and bottom. The total number of markers used was recorded (Supplementary Table 4). Where the number of overlapping markers between the maternal and paternal linkage maps was not provided in a publication, the marker numbers for both maps were summed.

In the global GWAS results, the allele frequency was studied over generations. The ancestors of genotypes were identified making use of the apple pedigrees of Muranty et al.[4]. For all significant marker-trait associations from the global GWAS, frequency of the allele associated with increased phenotypic value was estimated for the progeny group and for its five ancestor generations. To represent the ancestors, the allele frequency was

estimated for the 30 accessions of them included in the apple REFPOP. For major significant marker-trait associations with $R^2 > 0.1$ reported in the global GWAS, linkage disequilibrium was estimated as squared Pearson's correlations in a window of 3,000 markers surrounding each of the associations. A smaller window size was used for associations located towards the end of a chromosome.

A mixed-effects model following the Equation 1 was fitted to the vector of the adjusted phenotypic values of each tree as response ($\boldsymbol{y}$) using the effects of environments as fixed effects and the effects of genotypes, genotype by environment interactions, and additional effects for each SNP significantly associated with the trait (a factor of the respective SNP values in $\boldsymbol{M}$) as random effects. In cases where traits with no marker-trait associations were found in the global GWAS, the additional random effects of significantly associated SNPs were omitted from the model. The mixed-effects model for every trait was used to estimate proportions of phenotypic variance explained by the model components as described in Jung et al.[36]. The proportions of phenotypic variance explained by the random effects of genotypes and significantly associated SNPs were summed to obtain the proportion of variance explained by a genotypic effect. The proportions of phenotypic variance explained by genotypic, environmental, genotype by environment interaction, and residual effects were scaled and centered to be finally used for discovering similarities between the traits. For this purpose, a hierarchical clustering following Ward[54] was applied to the distance matrix of the set of effects. The number of clusters was estimated from a dendrogram, which was cut where the distance between splits was the largest.

3.2.6 Genomic prediction

The general statistical model for genomic prediction was

$$\boldsymbol{y} = 1\boldsymbol{\mu} + \boldsymbol{u} + \boldsymbol{\varepsilon} \text{ (Equation 4)}$$

where $\boldsymbol{y}$ was a vector of trait phenotypes, $\boldsymbol{\mu}$ was an intercept, $\boldsymbol{u}$ represented a vector of random effects and $\boldsymbol{\varepsilon}$ was a vector of residuals. Different vectors of $\boldsymbol{y}$ and assumptions for $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ were used across eight single- and multi-environment genomic prediction models.

Single-environment genomic prediction

The single-environment genomic prediction models were fitted after the environmental effects were accounted for during the phenotypic data analysis, a process also called two-step genomic prediction. Therefore, the across-location BLUPs and location-specific BLUPs acted here as phenotypes from a single environment. Four univariate prediction models and one multivariate model were implemented. First, regression with random forest (RF) was performed[55]. In this and the following three univariate models, the response $\boldsymbol{y}$ was defined as a $n \times 1$ vector of the main BLUPs. The centered and scaled additive genomic matrix $\boldsymbol{M}$, an $n \times m$ matrix for a population of size $n = 534$ with $m = 303{,}148$ markers, was used as further input. The number of trees to grow in the RF was 500 and the number of variables randomly sampled as candidates at each split was (rounded down) $mtry = m/3$. Second, BayesCπ was applied[56], where the random marker effects $\boldsymbol{u} = \sum_{k=1}^{m} z_k a_k$ with z_k an $n \times 1$ vector of the number of copies of one allele at the marker k and a_k being the additive effect of the marker k. The prior for a_k depended on the variance $\sigma^2_{a_k}$ and the prior probability π that a marker k had zero effect, the priors

of all marker effects having a common variance $\sigma^2_{a_k} = \sigma^2_a$. The π parameter was treated as an unknown with uniform(0,1) prior. The random vector of residual effects followed a normal distribution $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_\varepsilon)$ with $n \times n$ identity matrix $\boldsymbol{I}$ and the variance σ^2_ε. Third, the Bayesian reproducing kernel Hilbert spaces regression (RKHS) was implemented using a multi-kernel approach[57]. The multi-kernel model was fitted with $L = 3$ random marker effects $\boldsymbol{u} = \sum_{l=1}^{L} \boldsymbol{u_l}$ following a distribution $\boldsymbol{u} \sim N(0, \boldsymbol{K_l}\sigma^2_{ul})$, with $\boldsymbol{K_l}$ being the reproducing kernel evaluated at the lth value of the bandwidth parameter $h = \{h_1, \dots, h_L\} = \{0.1, 0.5, 2.5\}$ and the variance σ^2_{ul}. For each random effect, the kernel matrix $\boldsymbol{K} = \{K(x_i, x_{i'})\}$ was an $n \times n$ matrix $K(x_i, x_{i'}) = \exp\{-h \times D_{ii'}\}$, where $\boldsymbol{D} = \left\{D_{ii'} = \frac{\sum_{k=1}^{m}(x_{ik} - x_{i'k})^2}{m}\right\}$ was the average squared-Euclidean distance matrix between genotypes, and x_{ik} the element on line i (genotype i) and column k (kth marker) of the centered and scaled additive genomic matrix $\boldsymbol{M}$. The residual effect assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_\varepsilon)$. Fourth, from the centered and scaled additive genomic matrix $\boldsymbol{M}$, the genomic relationship matrix $\boldsymbol{G}$ was computed as $\boldsymbol{G} = \boldsymbol{M}\boldsymbol{M}'/m$ and used to fit the genomic-BLUP (G-BLUP) model applying a semi-parametric RKHS algorithm, with the random marker effects following $\boldsymbol{u} \sim N(0, \boldsymbol{G}\sigma^2_u)$ with variance σ^2_u and the model residuals assuming $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma^2_\varepsilon)$[58]. Fifth, a multivariate model with an unstructured covariance matrix of the random marker effect (here abbreviated as MTM.UN) was fitted for chosen pairs of traits using the Bayesian multivariate Gaussian model environment MTM (http://quantgen.github.io/MTM/vignette.html). The main BLUPs acted as the response $\boldsymbol{y}$, which was a vector of length $n \cdot t$ with $t = 2$ being the number of traits used in the model. The vector of the random marker effects followed $\boldsymbol{u} \sim N(0, \boldsymbol{U} \otimes \boldsymbol{G})$ where $\boldsymbol{U}$ was an unstructured covariance matrix of the random marker effect with dimension $t \times t$. Model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{R} \otimes \boldsymbol{I})$ with $\boldsymbol{R}$ being an unstructured covariance matrix of the residual effect. To choose the pairs of traits for MTM.UN, a G-BLUP model was applied using all genotypes to estimate genomic BLUPs, which were then used to obtain pairwise genomic correlations between traits. The pairs with the genomic correlations larger than 0.3 were retained for the MTM.UN analysis. In case a trait was included in more than one pair of traits, the result for the pair with the highest average predictive ability for this trait was reported.

BayesCπ, RKHS, G-BLUP and MTM.UN were applied with 12,000 iterations of the Gibbs sampler, a thinning of 5, and a burn-in of 2,000 discarded samples. With all models, a five-fold cross-validation repeated five times was performed, generating 25 estimates of prediction accuracy. The folds were chosen randomly without replacement to mask phenotypes of 20% of the genotypes in each run. Prediction accuracy was estimated as a Pearson correlation coefficient between phenotypes of the masked genotypes and the predicted values for the same genotypes. The RF model was implemented in the R package ranger[59], the models BayesCπ, RKHS and G-BLUP in the R package BGLR[60] and the MTM.UN model in the R package MTM (http://quantgen.github.io/MTM/vignette.html).

Multi-environment genomic prediction

Two univariate multi-environment genomic prediction algorithms and one multivariate multi-environment algorithm were implemented, the response $\boldsymbol{y}$ being a vector of the adjusted phenotypic values of each genotype of length $n \times r$ with r equal to the number of environments (nested factor of location and year). The two univariate multi-

environment models reported by Lopez-Cruz et al.[42] and implemented in the R package BGLR[60] were applied to explore the effects of genotypes, environments and their interaction in genomic prediction. Of the two models, the across-environment G-BLUP model (G-BLUP.E) assumed that marker effects were constant across environments. The random marker effects followed $\boldsymbol{u} \sim N(0, \boldsymbol{G_0}\sigma_u^2)$ where $\boldsymbol{G_0} = \boldsymbol{J} \otimes \boldsymbol{G}$, the $\boldsymbol{J}$ being an $r \times r$ matrix of ones. The model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$. Additionally to the constant effects of markers across environments as assumed in the previous model, the marker by environment interaction G-BLUP model (G-BLUP.E.G×E) allowed the marker effects to change across environments, i.e., to borrow information across environments. The random marker effects were defined as $\boldsymbol{u} = \boldsymbol{u_0} + \boldsymbol{u_1}$ where $\boldsymbol{u_0} \sim N(0, \boldsymbol{G_0}\sigma_{u0}^2)$ and $\boldsymbol{u_1} \sim N(0, \boldsymbol{G_1})$ with

$$\boldsymbol{G_1} = \begin{bmatrix} \sigma_{u1}^2\boldsymbol{G} & 0 & 0 \\ 0 & \sigma_{u2}^2\boldsymbol{G} & 0 \\ 0 & 0 & \sigma_{u3}^2\boldsymbol{G} \end{bmatrix}$$

assuming $r = 3$ here for easier notation. The model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{I}\sigma_\varepsilon^2)$. Finally, a multivariate multi-environment factor-analytic model (here abbreviated as MTM.FA) using the Bayesian multivariate Gaussian model environment implemented in the R package MTM (http://quantgen.github.io/MTM/vignette.html) was fitted to the data. As in the previous two models, phenotypes of the same trait from multiple environments acted as response, although this model was originally designed to analyze multiple traits. The traits measured at only one location during two seasons (full flowering, end of flowering, fruit volume, water core frequency and water core grade) were not modeled using MTM.FA because the analysis required at least three environments. The vector of the random marker effects assumed $\boldsymbol{u} \sim N(0, \boldsymbol{C} \otimes \boldsymbol{G})$ where $\boldsymbol{C}$ was an $r \times r$ covariance matrix. For the factor analysis, the $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{B}' + \boldsymbol{\Psi}$ where $\boldsymbol{B}$ was a matrix of loadings (regressions of the original random effects into common factors) and $\boldsymbol{\Psi}$ was a diagonal matrix whose entries gave the variances of environment-specific factors. The loadings were estimated for all environments and the variance of the Gaussian prior assigned to the unknown loadings was set to 100. The model residuals assumed $\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{R} \otimes \boldsymbol{I})$ with $\boldsymbol{R}$ being an unstructured covariance matrix of the residual effect.

All three multi-environment genomic prediction models were applied with 12,000 iterations of the Gibbs sampler, a thinning of 5 and a burn-in of 2,000 discarded samples. The folds of a five-fold cross-validation were chosen randomly without replacement. The cross-validation was repeated under two scenarios. In the first cross-validation scenario (CV1), the phenotypes of 20% of the genotypes were masked across all environments. For the second cross-validation scenario (CV2), the phenotypes of 20% of the genotypes were masked across all environments except for three Swiss environments, i.e., phenotypes of all genotypes from the environments “CHE.2018”, “CHE.2019” and “CHE.2020” were used for model training. Ten traits were measured in only one location and therefore excluded from CV2 (i.e., full flowering, end of flowering, fruit diameter, fruit length, maximum fruit size, fruit volume, yellow color, green color, water core frequency and water core grade). Prediction accuracy was estimated as a Pearson correlation coefficient between the phenotypes of the masked genotypes and the predicted values for these genotypes. The correlations were estimated for each predicted environment separately.

3.2.7 Genomic heritability

The BayesCπ model was applied for each trait as described before but trained with a full set of the main BLUPs as response. The genomic heritability $h^2 = V_g/(V_g + V_e)$ was estimated as the proportion of phenotypic variance explained by the markers, where V_g and V_e represented the amount of phenotypic variance explained and unexplained by the markers, respectively[61,62]. The genomic heritability was calculated from the marker effects saved in each iteration and averaged over iterations to obtain the mean genomic heritability per trait.

3.3 Results

3.3.1 Phenotypic data analysis

The accession and progeny groups of the apple REFPOP were evaluated for 30 quantitative traits at up to six locations. The measurements for ten traits were collected at one location, while the remaining 20 traits were available from at least two locations (three traits were measured in two locations, three traits in four locations, eleven traits in five locations and three traits in six locations, Supplementary Table 1). Most traits (25) were assessed during three seasons while five traits were measured during two seasons (Supplementary Table 1). Accounting for environmental effects in the phenotypic data, BLUPs of traits (best linear unbiased prediction of random effects of genotypes, see Equation 1) were produced across all locations and separately for each location. The traits showed unimodal as well as multimodal distributions (Supplementary Figure 1). Differences of various extent between the accession and progeny groups were observed (Supplementary Figure 2). As expected, high phenotypic and genotypic correlations (>0.7) between traits were observed within trait categories, namely the phenology, productivity, fruit size, outer fruit, inner fruit, and vigor category (Figure 1a). A few moderate positive phenotypic correlations (0.3–0.7) were found between trait categories such as harvest date and fruit firmness (0.51), yellow color and russet cover (0.55), soluble solids content and russet cover (0.36), or between yield (weight and number of fruits) and vigor trait category (0.36–0.51, Figure 1a). High average correlations were observed between the environments (combinations of location and year) for harvest date (0.82 [0.73, 0.95]) or red over color (0.80 [0.62, 0.92]) whereas low average correlations (<0.3) were present between environments for flowering intensity (0.18 [-0.49, 0.68]) and trunk increment (0.16 [-0.31, 0.55], Supplementary Table 2, Supplementary Figure 3). A shift of the progeny group compared to the accession group towards smaller, more numerous and less russeted fruits was observed (Figure 1b).

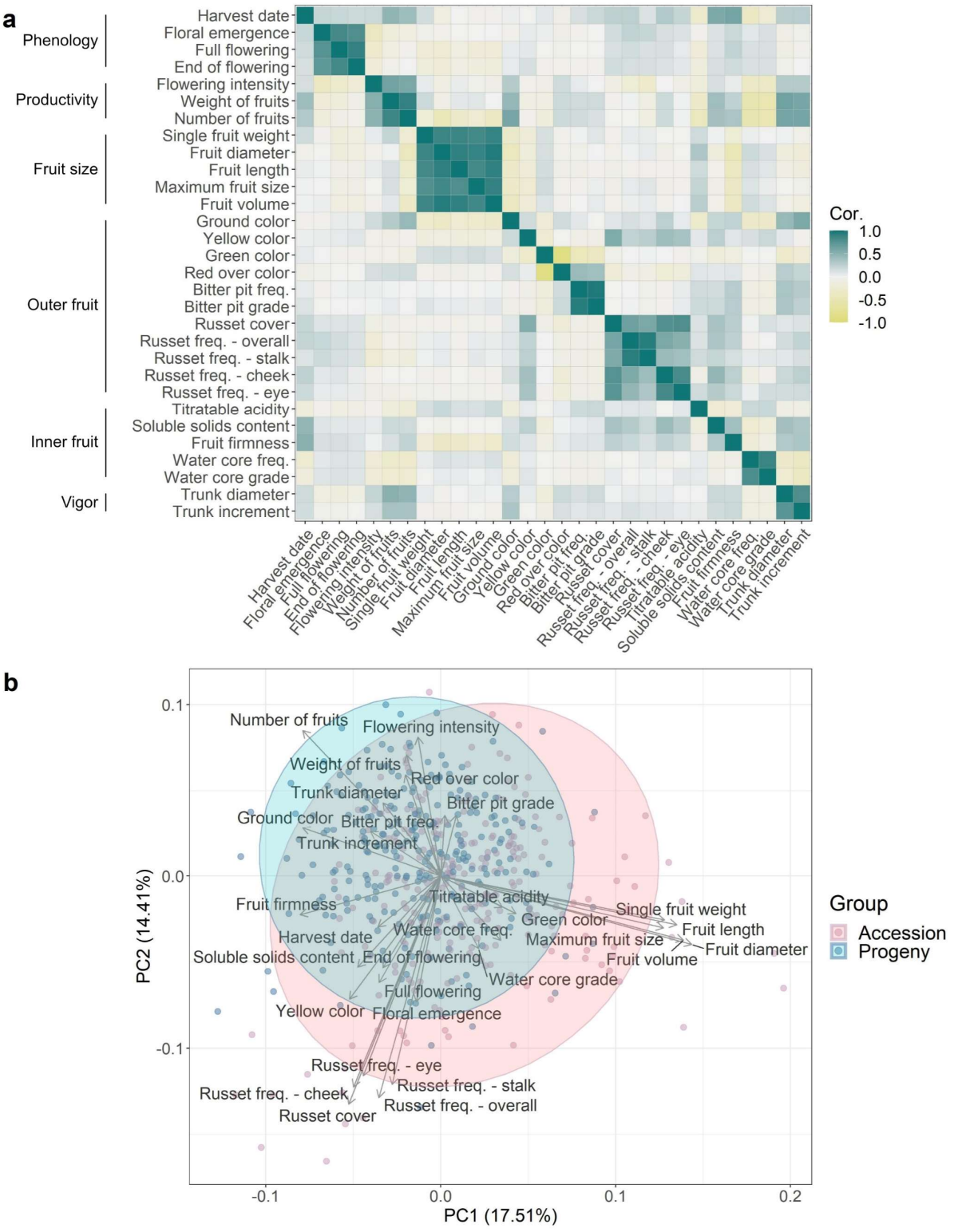


Figure 1: Exploratory phenotypic data analysis of the studied quantitative apple traits. **a** Pairwise correlations between traits with the phenotypic and genomic correlations in the lower and upper triangular part, respectively. Phenotypic correlation was assessed as Pearson correlation between pairs of across-location BLUPs, the genomic correlation as Pearson correlation between pairs of genomic BLUPs estimated from a G-BLUP model. Trait categories are outlined along the vertical axis. **b** Principal component analysis biplot based on across-location BLUPs of apple traits with the addition of location-specific BLUPs for traits measured at a single location.

3.3.2 Genome-wide association studies

Across-location GWAS for 20 traits measured at more than one location (Supplementary Table 1) and location-specific GWAS for all 30 traits were used to explore the genetic basis of the assessed traits. The quantile-quantile plots showed that the observed and expected distributions of p-values corresponded well and no apparent inflation of p-values was found (Supplementary Figure 4 and 5). Across-location GWAS revealed 59 significant ($-log_{10}(p) > 6.74$) marker-trait associations in 18 traits (Figure 2a, Supplementary Table 3). No significant associations were observed for trunk diameter and russet cover in the across-location GWAS. In the location-specific GWAS, 309 significant marker-trait associations for all 30 traits were discovered (Figure 2b, Supplementary Table 3). Of these 309 marker-trait associations, 32 associations for twelve traits were shared between the location-specific GWAS and the across-location GWAS (Supplementary Table 3). The coefficient of determination (R^2) of significant associations was the largest for red over color (0.71), green color (0.55) and harvest date (0.42, Figure 2c, Supplementary Table 3).

Significant associations with different traits co-localized at identical positions or occurred very close in some genomic regions (distance between marker positions below 100 kb, Figure 2c, Supplementary Table 3). In the across-location GWAS, a marker significantly associated with harvest date on chromosome 3 (position 30,681,581 bp) was located next to two markers associated with fruit firmness (positions 30,587,378 and 30,590,166 bp). The same marker on the position 30,681,581 bp was also associated with harvest date, ground color, overall russet frequency and soluble solids content measured at several different locations (location-specific GWAS). Similarly, the association with harvest date on chromosome 16 (position 9,023,861 bp) was closely located to a marker associated with fruit firmness (position 8,985,888 bp) in the across-location GWAS. The traits related to bitter pit analyzed in the across-location GWAS, i.e., bitter pit frequency and grade, showed significant associations on chromosome 16, position 7,681,416 bp. Several associations with traits measuring fruit skin russet in the across-location GWAS co-localized on chromosome 12 (position 23,013,281 bp, russet frequency on cheek and in the eye) and 17 (position 27,249,890 bp, overall russet frequency and russet frequency at stalk). A marker at position 18,679,105 bp on chromosome 1 was associated with both single fruit weight from the across-location GWAS and fruit diameter from Switzerland (found with the location-specific GWAS). The association with marker at position 2,005,502 bp on chromosome 8 was shared between fruit diameter and fruit volume from Switzerland and single fruit weight from Belgium. On chromosome 11, fruit diameter, fruit volume and single fruit weight from Switzerland, as well as single fruit weight from Belgium, shared the association at position 18,521,895 bp. Additionally, position 3,622,193 bp on chromosome 11 was shared between the associations of fruit length and single fruit weight from Switzerland. For red over color and green color, the association with a marker on chromosome 9 (position 33,799,120 bp) occurred in across-location and four location-specific GWAS, while a close marker (position 33,801,013 bp, less than 2kb away) was associated in the two other location-specific GWAS. Additional significant marker-trait associations occurred in the same genomic regions among the location-specific GWAS and between the across-location and location-specific GWAS (Supplementary Table 3).

Previous reports on QTL mapping and GWAS in apple were extensively reviewed and 41 publications reporting on traits measured similarly to our own were found and taken for comparison (Supplementary Table 4). The QTL positions from literature and the marker-trait associations found in this study were assigned to chromosome segments (top, center, and bottom of a chromosome). Unique segment-trait combinations were discovered in the literature (166), in the across-location GWAS (52) and in the location-specific GWAS (172,

Figure 3a). Out of all segment-trait combinations across our GWAS, 30.8% overlapped with the previously published results of QTL mapping or GWAS and the rest (69.2%) were novel. All previously published segment-trait combinations for the trait groups bitter pit and trunk were also detected in our study, whereas no overlap between the former and present associations was found for ground color and sugar trait groups (Figure 3b, Supplementary Figure 6).

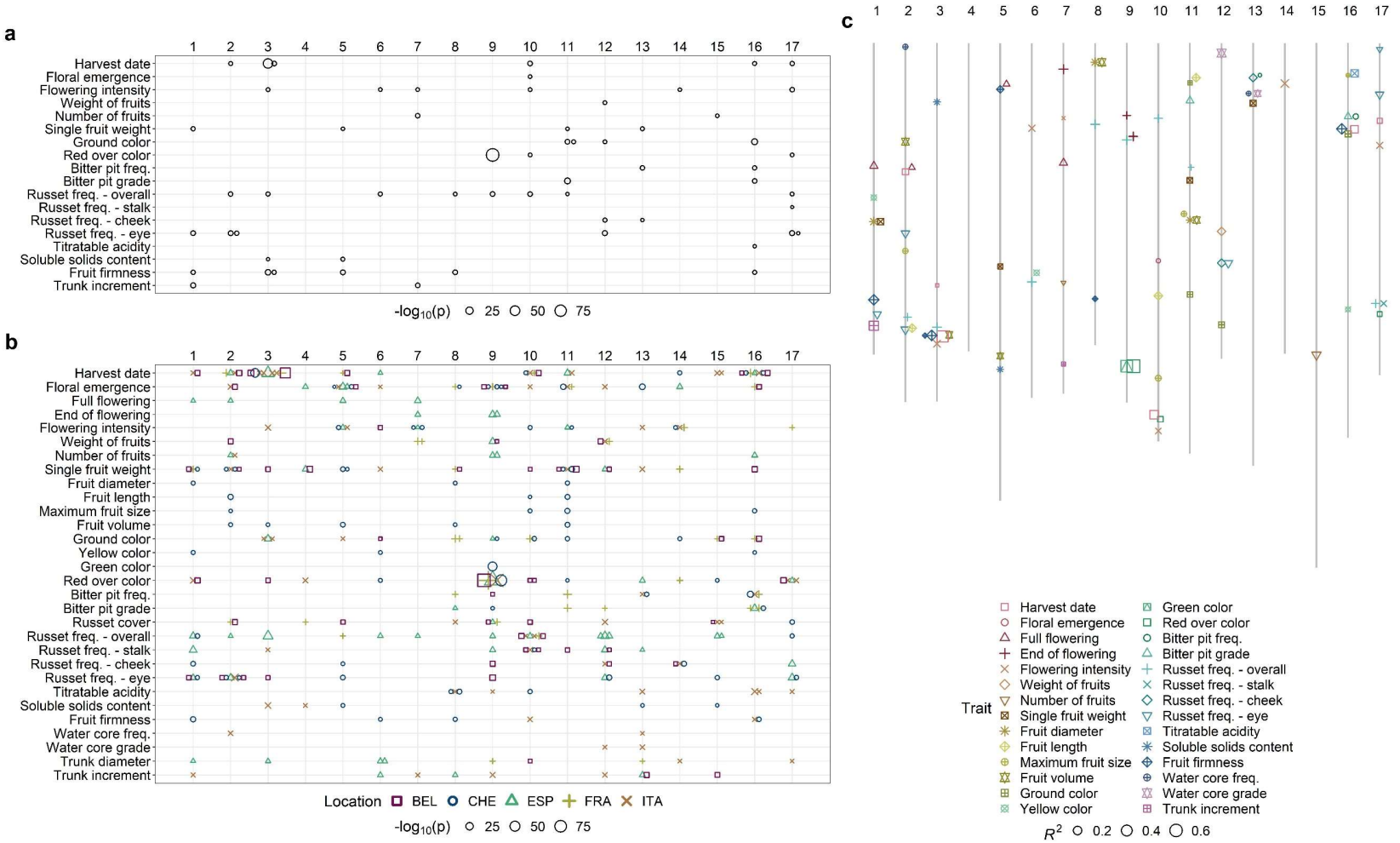


Figure 2: Significant marker-trait associations found by GWAS. **a** Distribution of the significant associations and corresponding p-values from across-location GWAS over the 17 apple chromosomes. **b** Distribution of the significant associations and corresponding p-values from location-specific GWAS over the 17 apple chromosomes. Locations are labeled as BEL (Belgium), CHE (Switzerland), ESP (Spain), FRA (France) and ITA (Italy). **a-b** Size of the symbols indicate the $-log_{10}(p)$. The x-axis shows chromosome numbers. **c** Physical positions (in bp) of the significant associations on chromosomes with their respective coefficients of determination (R^2) from the across-location GWAS complemented with the location-specific GWAS for traits measured at a single location. Size of the symbols indicate the R^2. The x-axis shows chromosome numbers.

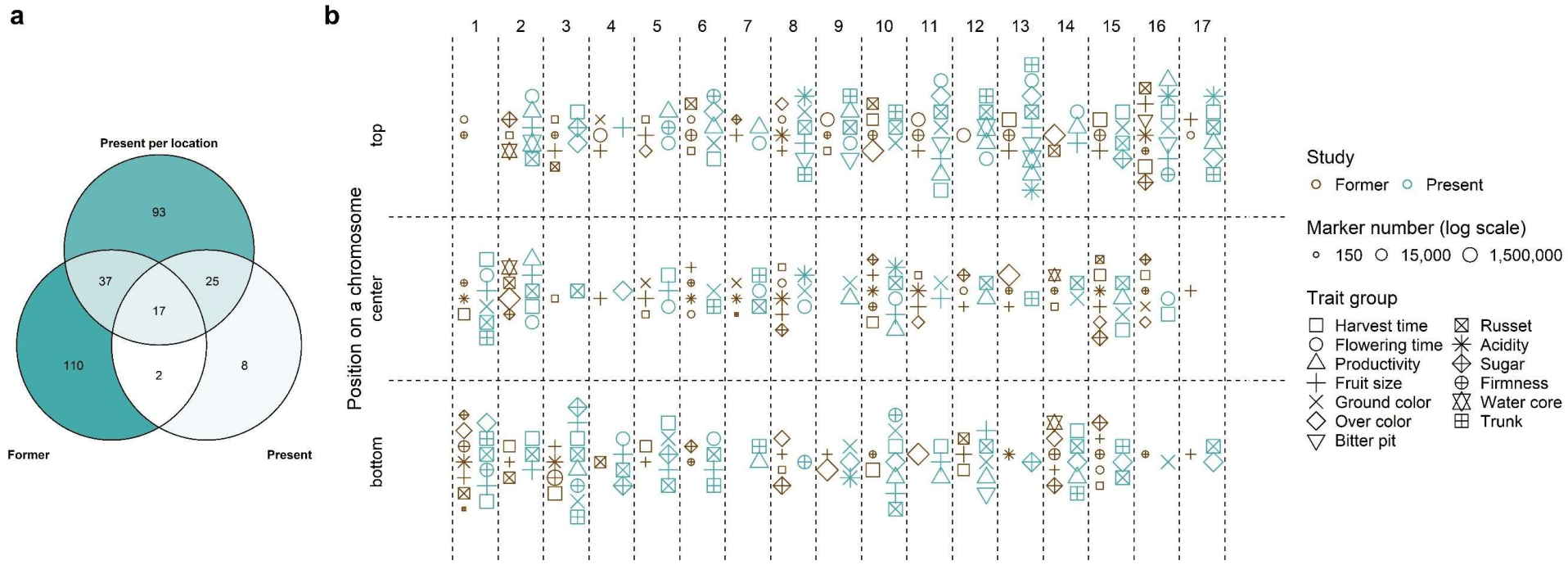


Figure 3: Comparison of the significant marker-trait associations with previously published associations. **a** Venn diagram comparing the unique associations, which were either previously published (former), reported in the across-location GWAS (present) or the location-specific GWAS (present per location). Color intensity and the values reflect the number of associations per diagram area. **b** Scatterplot of unique associations comparing published associations (former) with the merged across-location and location-specific GWAS (present). The traits were assembled into trait groups based on their similarity. Symbol size reflects the number of markers used in the studies. In case more than one publication reported an association in the same chromosome segment, only the report with the largest number of markers is shown (see Supplementary Table 4 for the complete list of previously published associations). **a-b** Positions of associations were assigned to three chromosome segments: top, center and bottom. Only the unique combinations of trait groups with segments and type of study (former or present) are shown

3.3.3 Allele frequency dynamics over generations

Eleven major significant marker-trait associations (R^2>0.1) were identified in the global GWAS results (across-location GWAS with the addition of location-specific GWAS for traits measured at a single location only, Figure 4). Among these major associations, changes in the frequency of alleles with an increasing effect on trait phenotypes were quantified in 30 ancestral accessions (five ancestor generations of the progeny group, Supplementary Table 5) and all 265 progenies included in the apple REFPOP (Figure 4a). Compared to the ancestral accessions, the frequency of the allele with an increasing effect on phenotype (Figure 4c) was higher in the progeny for the alleles associated with later harvest date and increased flowering intensity, titratable acidity, fruit firmness and trunk increment (Figure 4a). For the marker associated with green color and red over color, the allele frequencies were equivalent for ancestors and progeny, which reflected the minor allele frequency of nearly 0.5 for both traits (Figure 4b,d). Noticeably, at the markers closely associated with harvest date and fruit firmness on chromosome 3, the allele associated with later harvest date and firmer fruits was fixed in all progeny, while the allele with a decreasing effect on the phenotype was present with a frequency below 0.1 in the whole apple REFPOP (Figure 4a-d). The allele associated with larger trunk increment on chromosome 1 was found in progeny known to segregate for *Rvi6*, and it was present in only two accessions ('Prima' and X6398) that are also known to carry the apple scab resistance gene *Rvi6*, which is located about 1.8 Mb from the SNP associated with trunk increment (Figure 4b-c). The remaining associations (R^2≤0.1) reported by the global GWAS showed various trends in allele frequencies across generations such as increased frequency of alleles associated with increased weight of fruits in the progeny (Supplementary Figure 7). The individual parental combinations of the progeny group were often fixed for single alleles (Figure 4b, Supplementary Figure 8). Boxplots of the across-location BLUPs against the dosage of the reference allele (0, 1, 2) for the eleven major significant marker-trait associations showed additive effects of the alleles on phenotypes (Supplementary Figure 9). Squared Pearson's correlations in a window of ~3,000 markers surrounding each of the major significant marker-trait associations showed that markers in linkage disequilibrium extended over larger distances around some marker-trait associations (Supplementary Figure 10). When visually compared with other loci, the associations with harvest date and fruit firmness on chromosome 3 as well as red over color and green color on chromosome 9 were found in genomic regions of the highest linkage disequilibrium between markers (Supplementary Figure 10). The markers associated with trunk increment and *Rvi6* also showed signs of linkage disequilibrium (Supplementary Figure 10).

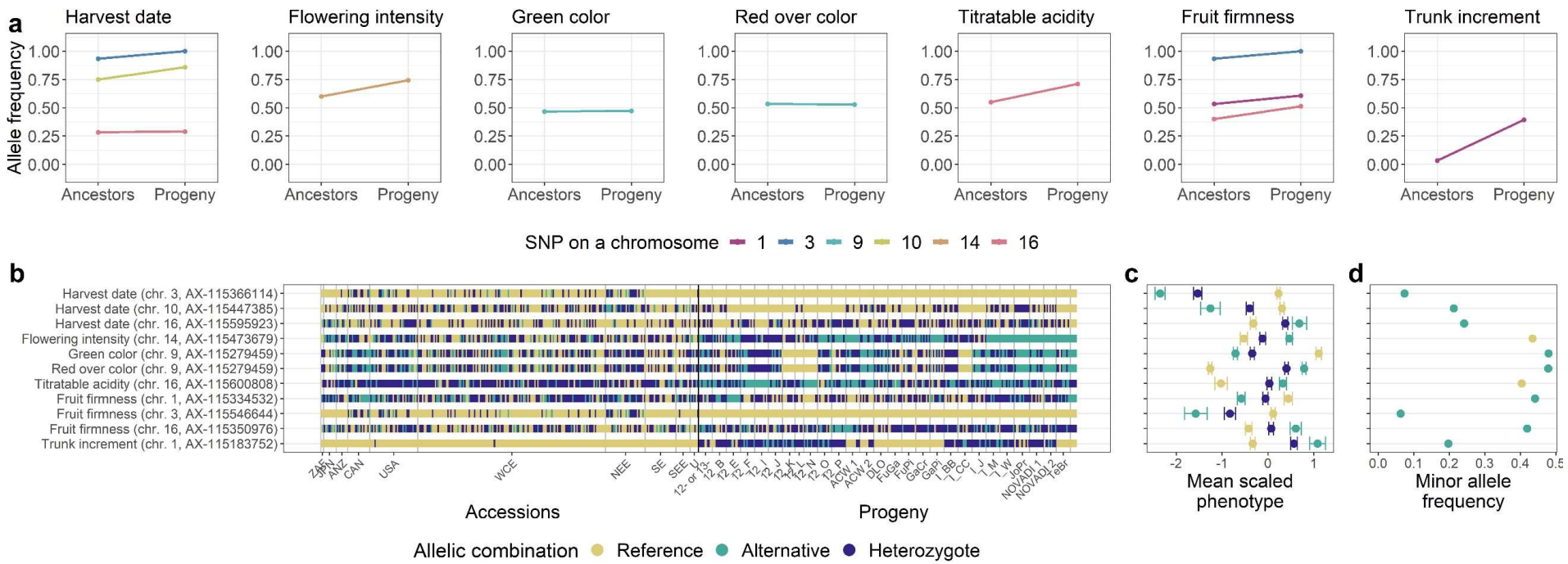


Figure 4: Allele frequency dynamics of the major significant marker-trait associations. **a-d** The associations were chosen based on the coefficient of determination (R^2>0.1) from the global GWAS. **a** For each association, frequency of the allele with increasing effect on trait phenotypes in the apple REFPOP is shown. For the progeny group (progeny) and its five ancestor generations (ancestors), the allele frequencies are shown as points connected with a line. Out of all known ancestors, the allele frequency was estimated for 30 accessions included in the apple REFPOP. Colors of the points and lines correspond to chromosome locations of the associated SNPs. **b** Allelic combinations carried by the apple REFPOP genotypes, sorted according to geographic origin of accessions and affiliation of progeny to parental combinations (the x-axis was labeled according to Supplementary Table 1 and 2 in Jung et al.[36]). **c** Phenotypic BLUPs of traits and their standard error for each allelic combination, centered to mean 0 and scaled to standard deviation of 1. **d** Frequency of the minor allele in the whole apple REFPOP. **b-d** The legend and y-axis are shared between plots. In d, the color of an allele corresponds to the color of the homozygous allelic combination of the same allele in b and c.

3.3.4 Genomic prediction

Four single-environment univariate prediction models – random forest (RF), BayesCπ, Bayesian reproducing kernel Hilbert spaces regression (RKHS) and genomic-BLUP (G-BLUP) – and a single-environment multivariate model with an unstructured covariance matrix of the random marker effect (MTM.UN) were compared using across-location BLUPs and location-specific BLUPs as phenotypes from a single environment. Among these models, the average prediction accuracies per trait ($\bar{r}_t$) ranged between 0.18 for russet cover and 0.88 for red over color, both extreme values observed with RF (Supplementary Table 6). The prediction accuracies estimated for G-BLUP were further used as reference for model comparisons due to its widespread use in genomic prediction. When the prediction accuracy of the G-BLUP model was averaged over all traits ($\bar{r}$), the obtained $\bar{r}$ was equal to 0.50. The RF showed an $\bar{r}_t$ higher than G-BLUP for 9 out of 30 traits and an $\bar{r}$ of 0.49. BayesCπ, RKHS and MTM.UN showed an $\bar{r}$ of 0.50, 0.51 and 0.50 and exceeded $\bar{r}_t$ of G-BLUP in one, twelve and ten traits, respectively. Generally, a similar performance of all five models was observed (Figure 5a).

When compared with the baseline model G-BLUP, the single-environment multivariate model MTM.UN showed an improved prediction accuracy for several traits when they were modelled in combination with a correlated trait (genomic correlation larger than 0.3, Figure 5a, Supplementary Table 6). The inclusion of floral emergence as correlated trait improved $\bar{r}_t$ of full flowering and end of flowering. A combination with weight of fruits improved $\bar{r}_t$ of flowering intensity. Fitting the model using fruit length showed an increased $\bar{r}_t$ of single fruit weight and using single fruit weight led to an increase in $\bar{r}_t$ for fruit diameter, fruit length, maximum fruit size and fruit volume. Using soluble solids content resulted in an increase of $\bar{r}_t$ for russet cover, while using russet frequency at cheek led to an improved $\bar{r}_t$ of russet frequency at stalk. Prediction accuracies for all possible combinations of correlated traits can be found in Supplementary Table 7.

Two multi-environment univariate models – across-environment G-BLUP (G-BLUP.E) and marker by environment interaction G-BLUP (G-BLUP.E.G×E) – and the multi-environment multivariate factor-analytic model (MTM.FA) were compared using two cross-validation scenarios corresponding to different experimental scenarios. In the first cross-validation scenario (CV1), traits were predicted for 20% of genotypes in each environment (i.e., their phenotypes were masked in all environments for model training). In the second cross-validation scenario (CV2), traits were predicted for 20% of genotypes in all but the Swiss environments (i.e., for these genotypes the environments "CHE.2018", "CHE.2019" and "CHE.2020" were retained for model training). For the models applied with CV1, the $\bar{r}_t$ ranged between 0.13 (for russet frequency in the eye obtained with MTM.FA) and 0.70 (for harvest date estimated with G-BLUP.E.G×E, Supplementary Table 6). With CV2, the lowest $\bar{r}_t$ of 0.29 was measured for trunk increment with G-BLUP.E.G×E and the maximum $\bar{r}_t$ of 0.86 was found for harvest date with both G-BLUP.E and G-BLUP.E.G×E models (Supplementary Table 6). The prediction performance of G-BLUP.E, G-BLUP.E.G×E and MTM.FA was generally lower under CV1 than under CV2 (Figure 5b, Supplementary Table 6). For all traits, the G-BLUP.E.CV1, G-BLUP.E.G×E.CV1 and MTM.FA.CV1 showed lower $\bar{r}_t$ than the single-environment G-BLUP, the $\bar{r}$ being equal to 0.40, 0.40 and 0.36, respectively. The G-BLUP.E.G×E.CV1 performed better than G-BLUP.E.CV1 for 14 out of 30 traits. The G-BLUP.E.CV2 and G-BLUP.E.G×E.CV2 outperformed G-BLUP for 13 out of

20 traits. The G-BLUP.E.CV2 and G-BLUP.E.G×E.CV2 both showed $\bar{r}$ equal to 0.57. The increase in $\bar{r}_t$ from G-BLUP to G-BLUP.E.CV2 (0.35) as well as from G-BLUP to G-BLUP.E.G×E.CV2 (0.36) was the most pronounced for russet cover. The performance of G-BLUP.E.CV2 and G-BLUP.E.G×E.CV2 remained below the level of G-BLUP predictions for productivity traits (flowering intensity, weight and number of fruits), ground color, soluble solids content, fruit firmness and trunk increment. The G-BLUP.E.G×E.CV2 performed better than G-BLUP.E.CV2 for 8 out of 20 traits. The $\bar{r}$ of MTM.FA.CV2 was equal to 0.52 and therefore similar to G-BLUP, however, the model outperformed G-BLUP for nine out of 20 predicted traits (Supplementary Table 6). The MTM.FA showed higher prediction accuracy than both G-BLUP.E and G-BLUP.E.G×E for two traits under CV1 and five traits under CV2 (Supplementary Table 6).

Across all model groups, the best prediction performance was found for harvest date, green color and red over color (Figure 5, Supplementary Table 6). The lowest prediction accuracy was found for traits related to bitter pit and russet as well as yellow color. Additionally, the prediction accuracy for flowering intensity and trunk increment with the multi-environment models remained strongly below the $\bar{r}_t$ of the corresponding single-environment models.

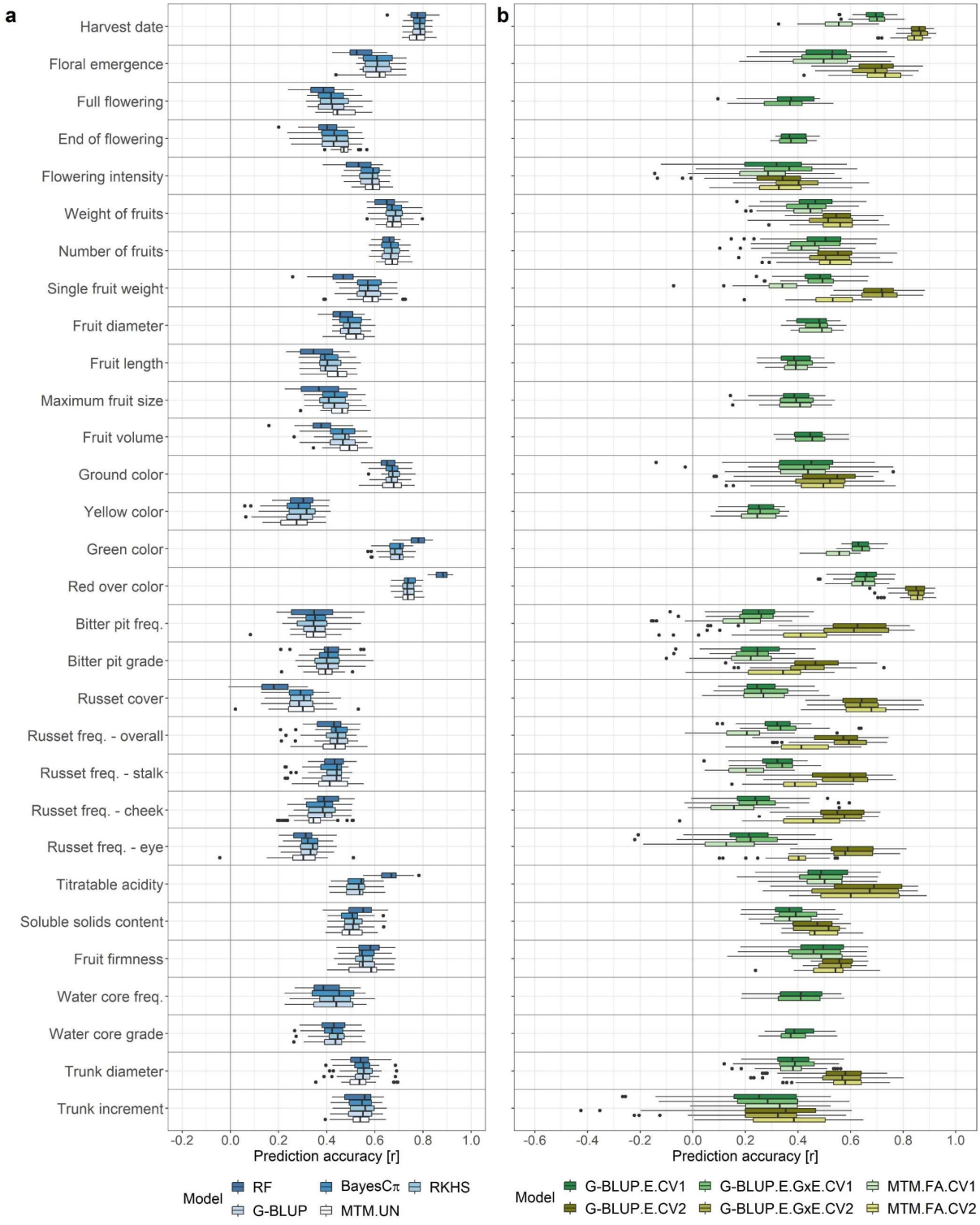


Figure 5: Genomic prediction accuracy in apple quantitative traits using eight genomic prediction models and two cross-validation scenarios. **a** Prediction accuracy of four single-environment univariate models, i.e., random forest (RF), BayesCπ, Bayesian reproducing kernel Hilbert spaces regression (RKHS) and genomic-BLUP (G-BLUP), and one single-environment multivariate model with an unstructured covariance matrix of the random marker effect (MTM.UN). The models were applied with a five-fold cross-validation where 20% of the genotypes were masked in each of the five runs. The MTM.UN was used in case a trait showed genomic correlation larger than 0.3 with at least one other trait. **b** Prediction accuracy of two multi-environment univariate models, i.e., across-environment G-BLUP (G-BLUP.E) and marker by environment interaction G-BLUP (G-BLUP.E.G×E), and the multi-environment multivariate factor-analytic model (MTM.FA). The models were applied under two five-fold cross-validation scenarios CV1 and CV2. The CV1 was applied for

all traits using G-BLUP.E and G-BLUP.E.G×E and for traits measured in at least three environments using MTM.FA. The CV2 was applied for traits measured in Switzerland and in at least a one other location. **a-b** Prediction accuracy was estimated as a Pearson correlation coefficient between the observed and the predicted values of genotypes whose phenotypes were masked in a five-fold cross-validation. For the multi-environment models, the correlation coefficients were estimated for each environment separately. In the box plot, the bottom and top line of the boxes indicate the 25th percentile and 75th percentile quartiles (the interquartile range), the center line indicates the median (50th percentile). The whiskers extend from the bottom and top line up to 1.5-times the interquartile range. The points beyond the 1.5-times the interquartile range from the bottom and top line are labeled as dots.

3.3.5 Synthesis of phenotypic and genomic analyses

The across-environment clonal mean heritability was generally very high in the evaluated traits, the value being close to one for harvest date and red over color and not lower than 0.80 for all the other traits with the exception of full flowering (0.74), end of flowering (0.79) and water core grade (0.79, Figure 6, Supplementary Table 6). The genomic heritability, which is the proportion of phenotypic variance explained by the markers, was larger than 0.80 for harvest date, floral emergence, green color and red over color, the value was not lower than 0.40 for all the other traits with the exception of bitter bit frequency (0.33) and grade (0.39, Figure 6, Supplementary Table 6).

The effects of genotype and significantly associated markers together explained a substantial part of the phenotypic variance of traits, the largest sums of these genotypic effects were observed for harvest date (82.8%) and red over color (74.6%, Figure 6, Supplementary Table 6). Altogether, the sum of the genotypic effects explained a very low proportion of the total variance for floral emergence (13.1%), flowering intensity (11.4%), trunk diameter (10.9%) and trunk increment (8.7%). The major proportion of the phenotypic variance was explained by the effect of environment for floral emergence (73.9%) and trunk diameter (66.3%). The lowest impact of environment was found for traits measured at only one location over two or three years such as fruit diameter or water core frequency, both showing an effect of environment (i.e., year) below 1%. The effect of G×E was the most pronounced for productivity traits, i.e., flowering intensity (23.7%), weight of fruits (20.8%) and number of fruits (21.6%). The proportion of the G×E effect was the lowest for harvest date (4.7%), floral emergence (5.2%), red over color (5.9%) and trunk diameter (4.2%) among the traits measured at more than one location and for end of flowering (5.7%), fruit volume (5.9%) and green color (3.9%) among the traits measured at one location. A high proportion of the phenotypic variance remained unexplained by the model parameters for flowering intensity (47.5%), bitter pit grade (53.4%) and trunk increment (55.1%).

Hierarchical clustering of the phenotypic variance components revealed three clusters of traits (Figure 6). A strong genotypic effect and a comparably low effect of environment and G×E was observed for 13 traits assigned to the cluster one. Most of the phenotypic variance was explained by the effect of environment in floral emergence and trunk diameter, which were grouped in cluster two. Finally, 15 traits with a pronounced effect of environment and/or G×E were grouped in cluster three.

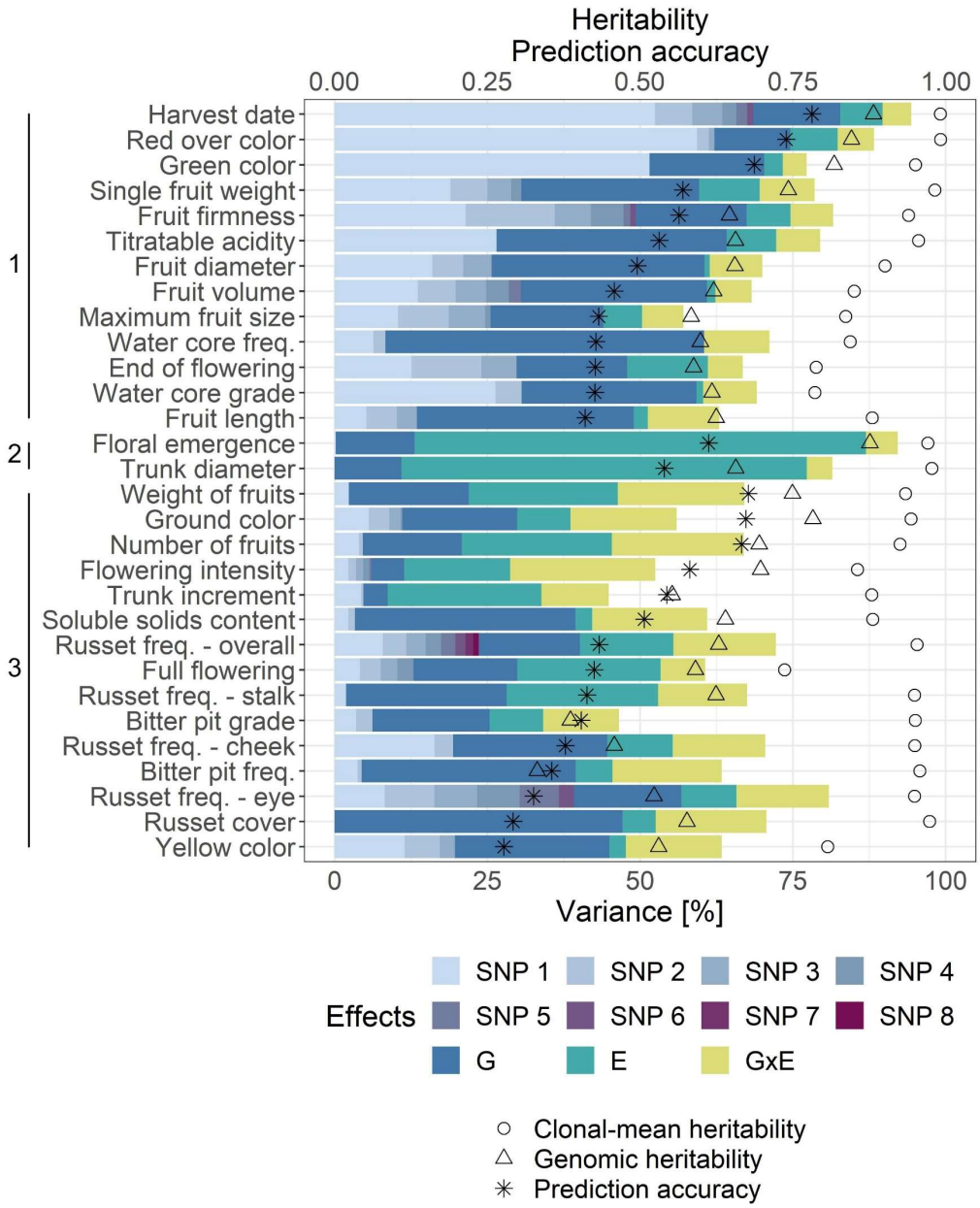


Figure 6: Synthesis of phenotypic and genomic analyses. Across-environment clonal mean heritability, genomic heritability, average prediction accuracy ($\bar{r}_t$) for the single-environment G-BLUP and the proportion of phenotypic variance explained by the effect of each significantly associated marker (SNP 1–8), genotype (G), environment (E) and genotype by environment interaction (G×E). The significantly associated markers corresponded to results of the global GWAS. Phenotypic variance components were used to estimate clusters of traits outlined along the vertical axis. Within each cluster, the traits were sorted according to $\bar{r}_t$.

3.4 Discussion

3.4.1 Discovered loci overlap between association studies and traits

Our GWAS permitted to enlighten the architecture of analyzed traits as well as the identification of numerous marker-trait associations stable across, and specific to, the locations of the apple REFPOP. The particular design of the experiment, including the diversity of the plant material used (accessions and small progeny groups), multiple locations, and multiple years of evaluation, resulted in about two thirds of the discovered associations being novel when compared with the loci published in studies spanning more than two decades. Our study design also allowed us to replicate the identification of many previously known loci associated with the studied traits.

The association of one locus with two or more seemingly independent traits (i.e., caused by pleiotropy) and linkage disequilibrium between loci associated with different traits are frequent for complex traits[63]. The GWAS performed in this study showed several marker-trait associations at identical or close positions for different traits. The interdependency between harvest date and fruit firmness, which can be also observed empirically for early cultivars that soften more, may be an example of pleiotropy or linkage disequilibrium between loci. Harvest date and fruit firmness are known to be regulated by ethylene production[64] and associated with loci present on chromosomes 3 (*NAC18.1*), 10 (*Md-ACO1*, *Md-PG1*), 15 (*Md-ACS1*) and 16[22,65-68].

In this work, closely located (distance <100 kb) associations with both harvest date and fruit firmness were found on chromosome 3. Migicovsky et al.[22] reported an overlap between associations with harvest time and fruit firmness on chromosome 3 falling within the coding region of *NAC18.1*. The authors hypothesized that the lack of associations on other chromosomes was likely due to low SNP density around the causal loci (the study used a GBS-derived 8K SNP dataset). The larger number of associations reported here might be a result of the high SNP density (303K SNPs) deployed in GWAS, however, not all previously reported loci were re-discovered.

The SNPs associated with harvest date and fruit firmness on chromosome 10 were further apart (~6 Mb). For harvest date, one of the associations on chromosome 10 was stable across locations and several associations were location specific. However, the association on chromosome 10 with fruit firmness was found for the Italian location only. It has been shown that chromosome 10 contains more than one QTL controlling fruit firmness[65-67], but stable across-location association with fruit firmness on chromosome 10 was missing in our study. One of the known loci on chromosome 10, the *Md-PG1* gene, is responsible for the loss of fruit firmness after storage[67,69]. In apple REFPOP, fruit firmness was measured within one week after the harvest date and this very short storage period might have contributed to the less pronounced effect of the locus *Md-PG1* in our GWAS.

Two associations with harvest date measured in Italy but no association with fruit firmness were found on chromosome 15. Although a marker for *Md-ACS1* related to ethylene production was previously mapped on chromosome 15[66], and QTL for fruit firmness was discovered on the same chromosome[65], these markers did not co-locate, but rather, mapped at the opposite extremes of chromosome 15[65,66]. Likewise, the connection between harvest date and fruit firmness on chromosome 15 could not be confirmed here.

Our GWAS showed associations with harvest date and fruit firmness on chromosome 16, which were located 38 kb apart. In the past, loci associated with harvest date and fruit firmness have been reported in the same region on chromosome 16[26,65]. The role of this locus in the regulation of harvest date and fruit firmness remains unknown and requires further research.

In practice, ripeness of fruit (harvest date) is decided based on ground color and starch content. The GWAS results showed that the association on chromosome 3 was not only found for harvest date and nearby markers associated with fruit firmness, but also corresponded to associations with ground color and soluble solids content. This might be explained by the fact that these traits are used to define ripeness and thus harvest date. Further, the association of the *NAC18.*1 locus on chromosome 3 with overall russet frequency would support the known enhanced expression of *NAC* transcription factors in russet skin[70].

Co-localizations between associations found for different measures of bitter pit on chromosome 16, russet on chromosomes 12 and 17, fruit size on chromosomes 1, 8 and 11, and skin color on chromosome 9 are likely the result of relatedness among trait measurements. The measures that are easiest to score can be used in future to phenotype these traits.

3.4.2 Signs of selection in marker-trait associations of large effect

The design of apple REFPOP allowed for the discovery of major marker-trait associations and for the analysis of changes in allele frequency between 30 ancestral accessions and 265 progeny included in the apple REFPOP. Comparing ancestors with the progeny, higher frequencies of the alleles associated with later harvest date and increased flowering intensity, titratable acidity, fruit firmness and trunk increment were found for the progeny. Of these traits, harvest date and fruit firmness are correlated, probably due to pleiotropy or linkage disequilibrium of causal loci, as it was shown in this and previous studies[22]. Consequently, the consistently higher frequency of alleles contributing to later harvest and firmer apples in the progeny is because the softening of harvested apples is undesirable and likely selected against[71]. Signs of selection for increased firmness were also recently found in USDA germplasm collection[5]. Our study also showed fixation of the late-harvest and high-firmness alleles on chromosome 3 in the whole progeny group, which suggests a loss of genetic diversity in the modern breeding material at this locus. For flowering intensity, a trait positively correlated with apple yield, a new locus was discovered on chromosome 14. The increased frequency of the allele contributing to higher flowering intensity in the progeny, its presence in all parental genotypes, and fixation in some parental combinations may be the result of breeding for high yield. The major locus found for acidity on chromosome 16 was consistent with the *Ma* locus frequently detected in various germplasm[8,11]. The total number of the high-acidity alleles for *Ma* and *Ma3*, which is another regularly detected acidity locus, was shown to be higher in parents of a European breeding program (Better3fruit, Belgium) than in parents used in the USDA breeding program[11,72]. The desired acidity level might depend on local climate of the breeding program and market preferences[72]. The increase in frequency of the allele contributing to higher acidity in the progeny may indicate a current preference towards more acidic apples in European breeding, but further investigation is needed to clarify the trend. The last locus of large effect showing allele frequency dynamics between

generations was found for trunk increment. The allele associated with an increase in trunk increment may have been selected in the progeny due to its potential impact on productivity suggested by moderate positive correlations between tree vigor (trunk diameter and increment) and yield-related traits. Additionally, the marker associated with trunk increment was 1.8 Mb apart from a SNP marker associated with *Rvi6* gene responsible for resistance against apple scab[10]. These two markers (AX-115183752 for trunk increment and AX-115182989 (also called Rvi6_42M10SP6_R193) for apple scab) showed a correlation of 0.15 and occurred within a region of increased linkage disequilibrium between markers (Supplementary Figure 10). All accessions were homozygous for the reference allele of AX-115183752 associated with decreased trunk increment (Figure 6c) except for 'Prima' and X6398, which were heterozygous. The scab-resistant accessions 'Prima' and X6398 (which is a second-generation offspring of 'Prima'[73]) but also 'Priscilla-NL' (known to be heterozygous for *Rvi6*[74]), were also heterozygous for AX-115182989. All other accessions were homozygous for the reference allele not associated with *Rvi6*. The allele on chromosome 1 associated with increased trunk increment may have been co-selected with the *Rvi6* locus responsible for resistance against apple scab.

Signs of intense selection for red skin were recently detected in the USDA germplasm collection when compared with progenitor species of the cultivated apple[5]. Our results show that the associations with red over color and green color, which phenotypically mirrored red over color and was associated with the same marker, did not show changes in allele frequency between ancestors and progeny included in the apple REFPOP. Some parental combinations showed almost exclusively the allele increasing red skin color, other parental combinations exhibited a lack of the allele. This uneven distribution of the alleles in the progeny group pointed to different directions of selection for fruit skin color in the European breeding programs (Figure 4b).

3.4.3 Performance of the single-environment univariate genomic prediction models

Single-environment univariate genomic prediction models were applied to individual traits after accounting for environmental effects and averaging across locations and/or years. The observed small differences between genomic prediction accuracies of various models (Figure 5a) were in accordance with previous model comparisons where distinctions among models were negligible[39,75]. The largest extremes in prediction accuracy between traits were found with random forest, which allowed for the overall highest prediction accuracy among all compared models for red over color. The explanation for the striking performance of random forest for red over color might be found in the results of our GWAS. This trait of oligogenic architecture was associated with a few low-effect loci and one locus of large effect explaining 61% of the red over color phenotypic variance measured in the apple REFPOP. High correlations between many markers, i.e., linkage disequilibrium, were found in the vicinity of the large-effect locus (Supplementary Figure 10). Random forest is known to assign higher importance to correlated predictor variables (here the markers) in the tree building process[76], which may have contributed to the particularly high prediction accuracy found for red over color with random forest.

The prediction accuracy for red over color reached ~0.4 in several former prediction studies[22,23,29,34] and was approximately doubled in our work, which demonstrated the potential of the current study design for accurate genomic predictions. For harvest date, the currently reported prediction accuracy of 0.78 was only slightly higher than the accuracy of 0.75 obtained with the initial apple REFPOP dataset measured during one year[36], but these accuracies showed a considerable improvement over other accuracies of approximately 0.5–0.6 reported elsewhere[22,23,29]. As shown before[36], these results underline the suitability of apple REFPOP design for the application of genomic prediction.

Prediction accuracy for traits such as yellow color or russet cover were on the opposite side of the spectrum when compared to harvest date and red over color. The prediction accuracy of yellow color and russet cover was low, although the genotypic effects explained 45% and 47% of the phenotypic variance, respectively. The across-environment clonal-mean heritability of russet cover was high (0.97), while the heritability for yellow color was slightly lower (0.81, Figure 6). Yellow color showed a moderate phenotypic correlation of 0.55 with russet cover, suggesting that the phenotyping device might have classified some russet skin as yellow color. Symptoms of powdery mildew could have been misinterpreted as russet skin. The decreased performance of genomic prediction models might stem from inaccurate phenotyping methods, insufficient SNP density in the associated regions, or other factors, all of which could not be explained in this work.

3.4.4 Role of genotype by environment interactions in multi-environment univariate genomic prediction

The multi-environment univariate genomic prediction models either directly estimated environmental effects (across-environment G-BLUP, called here G-BLUP.E) or additionally borrowed genotypic information across environments and thus considered the G×E (marker by environment interaction G-BLUP, called here G-BLUP.E.G×E)[42]. The average accuracy of the G-BLUP.E.G×E model across traits was only slightly higher than the accuracy of the G-BLUP.E. In contrast, the G-BLUP.E.G×E model had substantially greater prediction accuracy than the G-BLUP.E model when applied in wheat[42]. In the latter study, a productivity trait was measured under simulated conditions of mega-environments and the effect of G×E explained up to ~60% of the phenotypic variance[42]. Our work only focused on European environments and the largest proportion of phenotypic variance assigned to G×E was 24% for a productivity trait (flowering intensity). Furthermore, the average proportion of G×E across traits was approximately 12%, which may explain the mostly negligible differences between the G-BLUP.E and G-BLUP.E.G×E models. Our results were in line with the low interaction of additive genetic effects with location of up to ~6% obtained for apple fruit quality traits measured at two locations in New Zealand[33], and the limited G×E reported for fruit maturity timing in sweet cherry across continents[77]. For approximately half of the tested traits, the G-BLUP.E.G×E did not outperform G-BLUP.E. For these traits, the G-BLUP.E ignoring G×E may be sufficient to account for the environmental effects across European sites because it is computationally simpler and therefore less demanding. Traits such as flowering intensity, soluble solids content, trunk increment or traits related to fruit size and russet showed an improved performance under G-BLUP.E.G×E when compared to G-BLUP.E. For

traits positively responding to G-BLUP.E.G×E, the G×E should be considered when making predictions across environments. The highest improvement of prediction accuracy with G-BLUP.E.G×E when compared to G-BLUP.E was found for flowering intensity, the difference between the models amounting to 0.07 (Figure 5b). This result might be explained by the highest contribution of G×E to the phenotypic variance of flowering intensity among all traits (Figure 6). A comparably high contribution of G×E was also found for weight of fruits and number of fruits, though no improvement with G-BLUP.E.G×E model was observed for these traits. When comparing the relative contributions of variance components to the phenotypic variance of flowering intensity, weight of fruits and number of fruits, the genotype explained 11%, 22% and 21%, the environment 17%, 24% and 25%, and the G×E 24%, 21% and 22%, respectively. Although the proportions of G×E were similar in the three compared traits, the effects of genotype and environment explained a higher proportion of the variance for weight of fruits and number of fruits than for flowering intensity. This may have contributed to the surprisingly lower accuracy of the G-BLUP.E.G×E model when compared with G-BLUP.E for weight of fruits and number of fruits, but additional investigations may be needed to clarify this result in the future.

The G-BLUP.E.G×E model assumes positive correlations between environments and is therefore mostly suitable for the joint analysis of correlated environments[42,78]. As shown by Lopez-Cruz et al.[42] and in our study, this assumption of G-BLUP.E.G×E resulted in the best model performance for traits showing high positive correlations between environments (here harvest date and red over color) and the worst performance for traits exhibiting low correlations between environments (here flowering intensity and trunk increment, Figure 5b, Supplementary Table 2, Supplementary Figure 3). For flowering intensity and trunk increment, bivariate prediction of the environments or prediction with a different G×E model not assuming positive correlations between environments might be more appropriate than the currently applied approach[42,79].

3.4.5 Multivariate models as a useful element in the genomic prediction toolbox

Multivariate (also called multi-trait) models were shown to be useful for predicting traits that are costly to phenotype when a correlated trait less expensive to phenotype was available[45]. In our study, when the prediction accuracy of the single-environment multivariate model MTM.UN was compared with the baseline model G-BLUP, several combinations of related and unrelated traits led to increased accuracy. For the related traits with a high phenotypic correlation (Figure 1a), prediction of traits measured at one location were often improved when a related trait measured across different locations was included. This was the case for the combination of floral emergence with full flowering and end of flowering and for single fruit weight combined with fruit diameter, fruit length, maximum fruit size and fruit volume. Inclusion of soluble solids content in MTM.UN resulted in increased prediction accuracy for russet cover, although the traits showed only a moderate correlation and no obvious explanation for this result could be found. Our study supports the potential of multivariate models to borrow information that correlated traits provide about one another and identified trait combinations that can be successful under the multivariate setup.

In place of the correlated traits, environments of a single trait can be implemented in a multivariate model[46]. The average prediction accuracy over all traits was ~0.04 lower in the multi-environment multivariate (MTM.FA) than in the multi-environment univariate genomic prediction models (G-BLUP.E and G-BLUP.E.G×E). Compared to G-BLUP.E and G-BLUP.E.G×E, the MTM.FA showed the potential to perform equally well for six (CV1) and three traits (CV2) and was able to outperform both models for two (CV1) and five traits (CV2). In cases where MTM.FA outperformed G-BLUP.E and G-BLUP.E.G×E, a very limited increase in prediction accuracy of 0.01 was found for all traits but trunk increment, for which the increase was equal to 0.07 under the second cross-validation scenario. Except for the noticeable increase in prediction accuracy for trunk increment that could not be explained by our analyses, the performance of MTM.FA was similar to G-BLUP.E and G-BLUP.E.G×E, which establishes the multivariate model as a useful tool for multi-environment genomic prediction in apple.

3.4.6 Two approaches to genomic prediction addressed with cross-validation scenarios.

The cross-validation scenarios CV1 and CV2 were applied with multi-environment genomic prediction models to test two genomic prediction approaches typically faced in breeding. The CV1 imitated evaluation of breeding material that was yet untested in field trials. The CV2 was implemented to simulate incomplete field trials where breeding material was evaluated in some but not all target environments. More specifically, the CV2 investigated a situation where the breeding material has been evaluated at one location (the breeding site, in this case Switzerland) and the material's potential over other European sites was predicted without its assessment in a multi-environment trial, which may increase selection efficiency at latter stages of evaluation. As CV2 provided more phenotypic information to the models than CV1, a higher genomic prediction accuracy was found under CV2 when compared with CV1, which was anticipated[33,42]. The CV2 was tested by calibrating the model with Swiss observations only. The application of CV2 could be extended to other apple REFPOP locations to provide useful information for the breeding programs located at these sites. The choice of cross-validation scenario did not affect the general ranking of the average genomic prediction accuracies estimated for the evaluated traits.

3.4.7 Implications for apple breeding

Phenotypic variance decomposition into genetic, environmental, G×E and residual effects was compared with the results of GWAS and genomic prediction as well as heritability estimates. The comprehensive comparison indicated three classes of traits with contrasting genetic architecture and prediction performance. Characteristics of these trait classes and proposals for their efficient prediction strategies are described in the following paragraphs.

The first class included harvest date and red over color that showed a few loci of large effect and some additional loci of low effect, the highest prediction accuracies, and the highest across-environment clonal-mean heritability among all traits. Both traits showed a very high proportion of the genotypic effect explaining ~75% of the phenotypic variance. For harvest date and red over color, the marker with the largest effect explained 52% and 59% of the phenotypic variance and all marker effects in genomic prediction

captured together 88% and 85% of the phenotypic variance (i.e., genomic heritability of 0.88 and 0.85), respectively. Selection for these traits exhibiting a strong genetic effect of one locus could be done using marker-assisted selection, although only a part of the variance would be explained by a single marker. Better results can be achieved using genomic prediction, as this was able to explain a substantially larger amount of the phenotypic variance. Other traits such as fruit firmness, titratable acidity, end of flowering or traits related to fruit size and water core were grouped in the same cluster as harvest date and red over color (Figure 6). These traits showed a strong genotypic effect and a comparably low effect of environment and G×E, suggesting that selection for the traits would be efficient when performed using single-environment genomic prediction models rather than multi-environment prediction.

The second class of traits was represented by floral emergence and trunk diameter displaying a high proportion of the environmental effect (~70%) and a similar ratio of variance explained by genotypic effects compared to variance explained by G×E effects (~2.5). The genomic prediction accuracy did not considerably deviate from the average accuracy over all traits. Several marker associations with these traits were identified using location-specific GWAS. However, in the across-location GWAS, only one association explaining a very small part of phenotypic variance (floral emergence) or no association (trunk diameter) were discovered. Consequently, such traits predominantly driven by the effect of environment can be successfully selected based on genomic prediction, but the lack of associations stable across environments limits the applicability of marker-assisted selection to this class of traits.

In the third class, the productivity traits (flowering intensity, weight of fruits and number of fruits) showed the largest proportion of variance explained by G×E (~20%), with similar amounts of variance explained by genotypic effects for weight of fruits and number of fruits, but half as much variance explained by genotypic effects for flowering intensity (Figure 6). As a consequence, only flowering intensity showed higher prediction accuracy with G-BLUP.E.G×E than G-BLUP.E model. As shown above, the G×E should be considered when making predictions across environments for traits responding positively to the G-BLUP.E.G×E model, but G-BLUP.E may be sufficient for other traits to account for the environmental effects. To our knowledge, this is the first report of genomic prediction for apple yield components and our results can aid the establishment of productivity predictions in apple breeding. Other traits falling within the same cluster as the productivity traits, namely full flowering, ground color, yellow color, soluble solids content, trunk increment, and traits related to bitter pit and russet, showed a pronounced effect of environment and/or G×E (Figure 6). Multi-environment genomic prediction models can be efficient when applying genomic selection to these traits.

The decision to apply either marker-assisted or genomic selection can be based on genetic architecture of traits of interest and resources available in a breeding program. For breeding of yet genetically unexplored traits, variance decomposition of historical phenotypic data prior to genomic analyses may help describe trait architecture, assign traits to one of the three classes described in the previous paragraphs, and finally determine the most appropriate method of genomics-assisted breeding. From all traits explored in this study, the marker-trait associations with large and stable effects across environments found for harvest date, flowering intensity, green color, red over color,

titratable acidity, fruit firmness and trunk increment could be implemented into DNA tests for marker-assisted selection. These tests would allow for a reduction of labor costs in a breeding program when removing inferior seedlings without phenotyping[7]. Although generally requiring more statistical competences than marker-assisted selection, genomic selection can make use of both large- and low-effect associations between markers and traits when accommodating thousands of marker effects in a single genomic prediction model. For all studied traits, our results showed that marker effects estimated in genomic prediction were able to capture a larger proportion of the phenotypic variance than individual markers associated with the traits. Therefore, genomic selection should become the preferred method of genomics-assisted breeding for all quantitative traits explored in this study to ultimately increase their breeding efficiency and genetic gain.

3.5 Conclusion

This study laid the groundwork for marker-assisted and genomic selection across European environments for 30 quantitative apple traits. The apple REFPOP experimental design facilitated identification of a multitude of novel and known marker-trait associations. Our multi-environment trial provided accurate genomics-estimated breeding values for apple genotypes under various environmental conditions. Limited G×E detected in this work suggested consistent performance of genotypes across different European environments for most studied traits. Utilizing our dataset, more efficient selection of traits related to yield may lead to higher productivity and increased genetic gain in the future[37]. Improved fruit quality would appeal to consumers and tree phenology could be synchronized with current and future climates to secure production. The genomic prediction models developed here can be readily used for selecting germplasm in breeding programs, thus providing breeders with tools increasing selection efficiency. Beside the apple REFPOP, one other large multi-environment reference population for fruit trees, the PeachRefPop[80], was designed in Europe. Application of our study design to other horticultural crops such as peach can promote broader use of genomics-assisted breeding in the future.

Acknowledgements

The authors thank the field technicians and staff, especially Sylvain Hanteville, at INRAe IRHS and Experimental Unit (UE Horti), Angers, France, and technical staff at other apple REFPOP sites for the maintenance of the orchards and phenotypic data collection. We thank Dr. Graham Dow for English language editing. Phenotypic data collection was partially supported by the Horizon 2020 Framework Program of the European Union under grant agreement No 817970 (project INVITE: "Innovations in plant variety testing in Europe to foster the introduction of new varieties better adapted to varying biotic and abiotic conditions and to more sustainable crop management practices"). This work was partially supported by the project RIS3CAT (COTPA-FRUIT3CAT) financed by the European Regional Development Fund through the FEDER frame of Catalonia 2014-2020 and by the CERCA Program from Generalitat de Catalunya. We acknowledge financial support from the Spanish Ministry of Economy and Competitiveness through the “Severo Ochoa Programme for Centres of Excellence in R&D” 2016-2019 (SEV-20150533) and 2020-2023 (CEX2019-000902-S). C.D. was supported by "DON CARLOS ANTONIO LOPEZ" Abroad Postgraduate Scholarship Program, BECAL-Paraguay. We dedicate this paper to Prof. Edward Zurawicz of the National Institute of Horticultural Research in Skierniewice, Poland who co-promoted this study, but sadly recently passed away.

References

1 FAOSTAT (Food and Agriculture Organization of the United Nations, 2019).

2 Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I. & Gladieux, P. The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**, 57-65, doi:10.1016/j.tig.2013.10.002 (2014).

3 Way, R. D. *et al.* Apples (*Malus*). *Acta Horticulturae*, 3-46, doi:10.17660/ActaHortic.1991.290.1 (1991).

4 Muranty, H. *et al.* Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. *BMC Plant Biology* **20**, 2, doi:10.1186/s12870-019-2171-6 (2020).

5 Migicovsky, Z. *et al.* Genomic consequences of apple improvement. *Horticulture Research* **8**, 9, doi:10.1038/s41438-020-00441-7 (2021).

6 Urrestarazu, J. *et al.* Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biology* **16**, 130, doi:10.1186/s12870-016-0818-0 (2016).

7 Wannemuehler, S. D. *et al.* A cost–benefit analysis of DNA informed apple breeding. *HortScience horts* **54**, 1998, doi:10.21273/hortsci14173-19 (2019).

8 Maliepaard, C. *et al.* Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97**, 60-73, doi:10.1007/s001220050867 (1998).

9 Kenis, K., Keulemans, J. & Davey, M. W. Identification and stability of QTLs for fruit quality traits in apple. *Tree Genetics & Genomes* **4**, 647-661, doi:10.1007/s11295-008-0140-6 (2008).

10 Jänsch, M. *et al.* Identification of SNPs linked to eight apple disease resistance loci. *Molecular Breeding* **35**, 45, doi:10.1007/s11032-015-0242-4 (2015).

11 Verma, S. *et al.* Two large-effect QTLs, *Ma* and *Ma3*, determine genetic potential for acidity in apple fruit: breeding insights from a multi-family study. *Tree Genetics & Genomes* **15**, 18, doi:10.1007/s11295-019-1324-y (2019).

12 Baumgartner, I. O. *et al.* Development of SNP-based assays for disease resistance and fruit quality traits in apple (*Malus* × *domestica* Borkh.) and validation in breeding pilot studies. *Tree Genetics & Genomes* **12**, 35, doi:10.1007/s11295-016-0994-y (2016).

13 Iezzoni, A. F. *et al.* RosBREED: Bridging the chasm between discovery and application to enable DNA-informed breeding in rosaceous crops. *Horticulture Research* **7**, 177, doi:10.1038/s41438-020-00398-7 (2020).

14 Chagné, D. *et al.* Validation of SNP markers for fruit quality and disease resistance loci in apple (*Malus* × *domestica* Borkh.) using the OpenArray® platform. *Horticulture Research* **6**, 30, doi:10.1038/s41438-018-0114-2 (2019).

15 Velasco, R. *et al.* The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42**, 833-839, doi:10.1038/ng.654 (2010).

16 Daccord, N. *et al.* High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics* **49**, 1099-1106, doi:10.1038/ng.3886 (2017).

17 Zhang, L. *et al.* A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications* **10**, 1494, doi:10.1038/s41467-019-09518-x (2019).

18 Sun, X. *et al.* Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nature Genetics* **52**, 1423-1432, doi:10.1038/s41588-020-00723-9 (2020).

19 Broggini, G. A. L. *et al.* Chromosome-scale de novo diploid assembly of the apple cultivar 'Gala Galaxy'. *bioRxiv*, 2020.2004.2025.058891, doi:10.1101/2020.04.25.058891 (2020).
20 Bianco, L. *et al.* Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLOS ONE* **9**, e110377, doi:10.1371/journal.pone.0110377 (2014).
21 Bianco, L. *et al.* Development and validation of the Axiom®Apple480K SNP genotyping array. *The Plant Journal* **86**, 62-74, doi:10.1111/tpj.13145 (2016).
22 Migicovsky, Z. *et al.* Genome to phenome mapping in apple using historical data. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0113 (2016).
23 McClure, K. A. *et al.* A genome-wide association study of apple quality and scab resistance. *The Plant Genome* **11**, 170075, doi:10.3835/plantgenome2017.08.0075 (2018).
24 Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95-108, doi:10.1038/nrg1521 (2005).
25 Kumar, S. *et al.* Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genomics* **14**, 393, doi:10.1186/1471-2164-14-393 (2013).
26 Urrestarazu, J. *et al.* Genome-wide association mapping of flowering and ripening periods in apple. *Frontiers in Plant Science* **8**, 1923, doi:10.3389/fpls.2017.01923 (2017).
27 Larsen, B. *et al.* Genome-wide association studies in apple reveal loci for aroma volatiles, sugar composition, and harvest date. *The Plant Genome* **12**, 180104, doi:10.3835/plantgenome2018.12.0104 (2019).
28 Hu, Y. *et al.* ERF4 affects fruit firmness through TPL4 by reducing ethylene production. *The Plant Journal* **103**, 937-950, doi:10.1111/tpj.14884 (2020).
29 Minamikawa, M. F. *et al.* Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. *Horticulture Research* **8**, 49, doi:10.1038/s41438-021-00485-3 (2021).
30 Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819 (2001).
31 Meuwissen, T. Genomic selection: marker assisted selection on a genome wide scale. *Journal of Animal Breeding and Genetics* **124**, 321-322, doi:10.1111/j.1439-0388.2007.00708.x (2007).
32 Kumar, S. *et al.* Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). *PLOS ONE* **7**, e36674, doi:10.1371/journal.pone.0036674 (2012).
33 Kumar, S. *et al.* Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. *G3 Genes|Genomes|Genetics* **5**, 2711-2718, doi:10.1534/g3.115.021105 (2015).
34 Muranty, H. *et al.* Accuracy and responses of genomic selection on key traits in apple breeding. *Horticulture Research* **2**, 15060, doi:10.1038/hortres.2015.60 (2015).
35 Roth, M. *et al.* Genomic prediction of fruit texture and training population optimization towards the application of genomic selection in apple. *Horticulture Research* **7**, 148, doi:10.1038/s41438-020-00370-5 (2020).
36 Jung, M. *et al.* The apple REFPOP—a reference population for genomics-assisted breeding in apple. *Horticulture Research* **7**, 189, doi:10.1038/s41438-020-00408-8 (2020).

37 García-Ruiz, A. *et al.* Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proceedings of the National Academy of Sciences* **113**, E3995-E4004, doi:10.1073/pnas.1519061113 (2016).

38 Duan, N. *et al.* Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications* **8**, 249, doi:10.1038/s41467-017-00336-7 (2017).

39 Howard, R., Carriquiry, A. L. & Beavis, W. D. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. *G3: Genes|Genomes|Genetics* **4**, 1027-1046, doi:10.1534/g3.114.010298 (2014).

40 Cooper, M. & DeLacy, I. H. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* **88**, 561-572, doi:10.1007/BF01240919 (1994).

41 Snape, J. W. *et al.* Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. *Euphytica* **154**, 401-408, doi:10.1007/s10681-006-9208-2 (2007).

42 Lopez-Cruz, M. *et al.* Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3: Genes|Genomes|Genetics* **5**, 569-582, doi:10.1534/g3.114.016097 (2015).

43 Jarquín, D. *et al.* A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics* **127**, 595-607, doi:10.1007/s00122-013-2243-1 (2014).

44 Tsai, H.-Y. *et al.* Use of multiple traits genomic prediction, genotype by environment interactions and spatial effect to improve prediction accuracy in yield data. *PLOS ONE* **15**, e0232665, doi:10.1371/journal.pone.0232665 (2020).

45 Lado, B. *et al.* Resource allocation optimization with multi-trait genomic prediction for bread wheat (Triticum aestivum L.) baking quality. *Theoretical and Applied Genetics* **131**, 2719-2731, doi:10.1007/s00122-018-3186-3 (2018).

46 Gianola, D. & Fernando, R. L. A multiple-trait Bayesian LASSO for genome-enabled analysis and prediction of complex traits. *Genetics* **214**, 305-331, doi:10.1534/genetics.119.302934 (2020).

47 Laurens, F. *et al.* An integrated approach for increasing breeding efficiency in apple and peach in Europe. *Horticulture Research* **5**, 11, doi:10.1038/s41438-018-0016-3 (2018).

48 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81**, 1084-1097, doi:10.1086/521987 (2007).

49 Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A. & Eilers, P. H. C. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* **23**, 52-71, doi:10.1016/j.spasta.2017.10.003 (2018).

50 Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67** (2015).

51 Gabriel, K. R. The biplot graphic display of matrices with application to principal component analysis. *Biometrika* **58**, 453-467, doi:10.1093/biomet/58.3.453 (1971).

52 Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**, doi:10.1093/gigascience/giy154 (2018).
53 Tang, Y. *et al.* GAPIT version 2: An enhanced integrated tool for genomic association and prediction. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0120 (2016).
54 Ward, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* **58**, 236-244, doi:10.1080/01621459.1963.10500845 (1963).
55 Breiman, L. Random forests. *Machine Learning* **45**, 5-32, doi:10.1023/A:1010933404324 (2001).
56 Habier, D., Fernando, R. L., Kizilkaya, K. & Garrick, D. J. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* **12**, 186, doi:10.1186/1471-2105-12-186 (2011).
57 de los Campos, G., Gianola, D., Rosa, G. J. M., Weigel, K. A. & Crossa, J. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research* **92**, 295-308, doi:10.1017/S0016672310000285 (2010).
58 VanRaden, P. M. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414-4423, doi:10.3168/jds.2007-0980 (2008).
59 Wright, M. N. & Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software* **77**, 1-17, doi:10.18637/jss.v077.i01 (2017).
60 Pérez, P. & de los Campos, G. Genome-wide regression and prediction with the BGLR statistical package. *Genetics* **198**, 483-495, doi:10.1534/genetics.114.164442 (2014).
61 de los Campos, G., Sorensen, D. & Gianola, D. Genomic heritability: What is it? *PLOS Genetics* **11**, e1005048, doi:10.1371/journal.pgen.1005048 (2015).
62 Lehermeier, C., de Los Campos, G., Wimmer, V. & Schön, C. C. Genomic variance estimates: With or without disequilibrium covariances? *J Anim Breed Genet* **134**, 232-241, doi:10.1111/jbg.12268 (2017).
63 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nature Genetics* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).
64 Johnston, J. W., Gunaseelan, K., Pidakala, P., Wang, M. & Schaffer, R. J. Co-ordination of early and late ripening events in apples is regulated through differential sensitivities to ethylene. *Journal of Experimental Botany* **60**, 2689-2699, doi:10.1093/jxb/erp122 (2009).
65 Chagné, D. *et al.* Genetic and environmental control of fruit maturation, dry matter and firmness in apple (Malus × domestica Borkh.). *Horticulture Research* **1**, 14046, doi:10.1038/hortres.2014.46 (2014).
66 Costa, F. *et al.* Role of the genes Md-ACO1 and Md-ACS1 in ethylene production and shelf life of apple (Malus domestica Borkh). *Euphytica* **141**, 181-190, doi:10.1007/s10681-005-6805-4 (2005).
67 Costa, F. *et al.* QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (Malus x domestica Borkh.). *Journal of experimental botany* **61**, 3029-3039, doi:10.1093/jxb/erq130 (2010).
68 Longhi, S., Moretto, M., Viola, R., Velasco, R. & Costa, F. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (*Malus* x

domestica Borkh.). *Journal of Experimental Botany* **63**, 1107-1121, doi:10.1093/jxb/err326 (2012).

69 Longhi, S. *et al.* A candidate gene based approach validates Md-PG1 as the main responsible for a QTL impacting fruit texture in apple (*Malus* x *domestica* Borkh.). *BMC Plant Biology* **13**, 37, doi:10.1186/1471-2229-13-37 (2013).

70 Legay, S. *et al.* Apple russeting as seen through the RNA-seq lens: Strong alterations in the exocarp cell wall. *Plant Molecular Biology* **88**, 21-40, doi:10.1007/s11103-015-0303-4 (2015).

71 Johnston, J. W., Hewett, E. W. & Hertog, M. L. A. T. M. Postharvest softening of apple (Malus domestica) fruit: A review. *New Zealand Journal of Crop and Horticultural Science* **30**, 145-160, doi:10.1080/01140671.2002.9514210 (2002).

72 Rymenants, M. *et al.* Detection of QTL for apple fruit acidity and sweetness using sensorial evaluation in multiple pedigreed full-sib families. *Tree Genetics & Genomes* **16**, 71, doi:10.1007/s11295-020-01466-8 (2020).

73 van de Weg, E. *et al.* Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. *Molecular Breeding* **38**, 5, doi:10.1007/s11032-017-0755-0 (2017).

74 Evans, K. M. *et al.* Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: Trueness-to-type of cultivars and their parentages. *Molecular Breeding* **28**, 535-547, doi:10.1007/s11032-010-9502-5 (2011).

75 Heslot, N., Yang, H.-P., Sorrells, M. E. & Jannink, J.-L. Genomic selection in plant breeding: A comparison of models. *Crop Science* **52**, 146-160, doi:10.2135/cropsci2011.06.0297 (2012).

76 Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T. & Zeileis, A. Conditional variable importance for random forests. *BMC Bioinformatics* **9**, 307, doi:10.1186/1471-2105-9-307 (2008).

77 Hardner, C. M. *et al.* Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. *Horticulture Research* **6**, 6, doi:10.1038/s41438-018-0081-7 (2019).

78 Crossa, J. *et al.* Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science* **22**, 961-975, doi:10.1016/j.tplants.2017.08.011 (2017).

79 Cuevas, J. *et al.* Bayesian genomic prediction with genotype × environment interaction kernel models. *G3 Genes|Genomes|Genetics* **7**, 41-53, doi:10.1534/g3.116.035584 (2017).

80 Cirilli, M. *et al.* The multisite PeachRefPop collection: A true cultural heritage and international scientific tool for fruit trees *Plant Physiology* **184**, 632-646, doi:10.1104/pp.19.01412 (2020).

Supplementary methods

All traits were measured at the level of individual trees (genotype replicates). From the traits, 20 were measured at more than one location and ten were measured at one location only (see Supplementary Table 1 for trait-location combinations). Unless stated otherwise, the measurement of traits was performed on the whole crop (further denoted as produced fruits). The fruits fallen to the ground due to, e.g., over-ripeness or strong wind, were also collected unless they were heavily decayed or found too far away from the tree. Fruits that were not fully developed and of a very small size due to, e.g., aphids or second flowering, were discarded and not counted as produced fruits. Weight of fruits and outer fruit traits were scored on a fruit batch consisting of 20 fruits selected randomly or all produced fruits for trees bearing at least five and up to 20 fruits. Trees producing less than five fruits were not scored for traits measured on a fruit batch.

Phenology traits

Floral emergence was estimated as the date when the first 10% of flowers opened. *Full flowering* was estimated as the date when at least 50% of flowers were open. *End of flowering* was the date when all flower petals fell off. On *harvest date*, more than 50% of the produced fruits were physiologically fully mature. The ripeness of fruits for the estimation of the harvest date was determined by expert knowledge. Every date was converted to the number of days since the beginning of year of the measurement.

Productivity traits

Flowering intensity was the percentage of existing flowers from the maximum possible number of flowers. A tree would show the maximum possible number of flowers if flowering at each leafy shoot. The used scale was: grade 1 corresponded to 0%, grade 3 to 25%, grade 5 to 50%, grade 7 to 75% and grade 9 to 100%. To measure the *number of fruits*, all produced fruits were counted on harvest date. *Weight of fruits* in kilograms was measured with scales or a sorting machine (Greefa iQS4 v.1.0) from the full set of produced fruits or from the fruit batch. In case the weight of a fruit batch was measured, the weight of all fruits was calculated using the average fruit weight estimated from the batch multiplied by the number of all produced fruits.

Fruit size

Single fruit weight in grams was obtained by dividing the weight of all fruits in grams by the number of fruits. *Fruit diameter*, *fruit length*, *maximum fruit size* and *fruit volume* were estimated with the sorting machine for each produced fruit. To measure fruit diameter in mm, the sorting machine made twelve images per fruit, estimated the diameter from each image (along the horizontal axis), excluded the lowest and the largest value and averaged the remaining ten estimates. Fruit length in mm was obtained similarly to fruit diameter but using the vertical axis instead of the horizontal axis. Maximum fruit size in mm was derived from the fruit diameter estimates obtained from twelve images as an average of the second and the third largest values. Fruit volume was estimated using a ready-made algorithm of the sorting machine. Fruit diameter, fruit length and fruit volume were averaged across all produced fruits. For the maximum fruit size, the maximum value among all produced fruits was used.

Outer fruit

Ground color was visually estimated for a fruit batch assigning a grade between zero and two (Figure SM 1). *Red over color* represented a percentage of red fruit skin estimated visually. The assessment of red over color was performed on a fruit batch, where a grade between zero and five was assigned to the whole batch (Figure SM 2). *Yellow color* and *green color* were estimated automatically using the sorting machine as an average percentage of these colors on the skin of produced fruits.

green (0) | green-yellow yellow-green (1) | yellow (2)

Figure SM 1: Scale for the assessment of ground color with the corresponding grades in brackets

absent (0) | 1–20% (1) | 21–40% (2) | 41–60% (3) | 61–80% (4) | 81–100% (5)

Figure SM 2: Scale for the assessment of red over color with the corresponding grades in brackets

Bitter pit is a disorder characterized by discrete pits in the fruit flesh, which turn brown and desiccate over time[1]. *Bitter pit frequency* was measured as the number of apples in a batch that were showing symptoms of bitter pit. The frequencies were transformed into percentages, which represented proportions of the batch size. *Bitter pit grade* was determined for a fruit batch assigning a grade between zero and two (Figure SM 3).

no symptoms (0) | medium symptoms (1) | very high symptoms (2)

Figure SM 3: Scale for the assessment of bitter pit grade with the corresponding grades in brackets, the scale was adapted from Buti et al.[1].

Russet is a fruit characteristic caused by small cracks in fruit skin, which develop during fruit growth and are replaced by cork tissue leading to rough brown skin. *Russet cover*

was estimated as a percentage of fruit surface covered by the russet, estimated for a fruit batch assigning a grade between zero and five (Figure SM 4). A small patch of russet deep in the stalk cavity, which reached up to 50% of the depth of the cavity and in case of no other signs of russet on the fruit skin was considered an absent russeting (grade 0). *Overall russet frequency* was measured as the number of apples in a batch that were showing russet. *Russet frequency in the stalk basin*, *russet frequency on the cheek* and *russet frequency in the eye basin* were equal to the number of apples in a batch showing russet in the given area. All frequencies were transformed into percentages as proportions of the batch size.

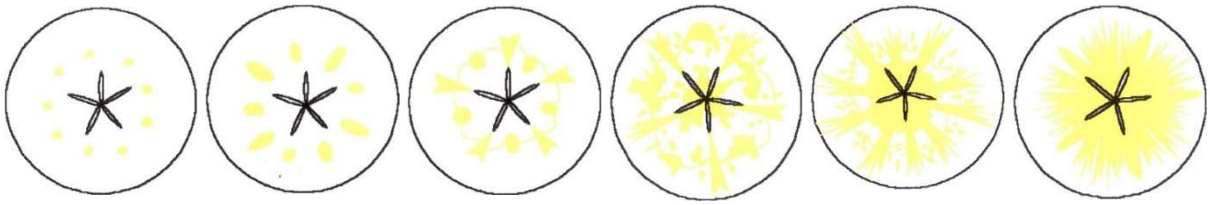
absent (0)	1–20% (1)	21–40% (2)	41–60% (3)	61–80% (4)	81–100% (5)

Figure SM 4: Scale for the assessment of russet cover with the corresponding grades in brackets

Inner fruit

A sample of ten fruits was analyzed with an automated instrument Pimprenelle (Setop, France) within one week after the harvest date. A tree was omitted from the analysis with Pimprenelle in case it produced less than ten fruits (or less than five fruits in 2018 due to young age of trees linked with low production). *Fruit firmness* in g/cm^2 was measured on individual fruits of the sample using a penetrometer. Each fruit was then pressed to extract the juice, which passed onto a refractometer to measure *soluble solids content* as a refraction index in degrees Brix. Mean values of fruit firmness and soluble solids content were calculated per fruit sample, the refraction index of the first measured fruit was deleted by the Pimprenelle. All juice samples of the same fruit sample were pooled for the analysis of *titratable acidity*, which was measured in grams of titratable acid per liter of solution.

Water core is a disorder that is characterized by areas of fruit flesh soaked with water appearing as translucent tissue. *Water core frequency* was estimated as the number of apples in a sample of ten fruits that were showing symptoms of water core. *Water core grade* was visually scored from the sample of ten fruits assigning a grade between zero and six, the zero grade being equal to absence of water core in the sample (Figure SM 5).



(1)	(2)	(3)	(4)	(5)	(6)

Figure SM 5: Scale for the assessment of water core grade with the corresponding grades in brackets (developed and applied at the Research Centre Laimburg, South Tyrol, Italy)

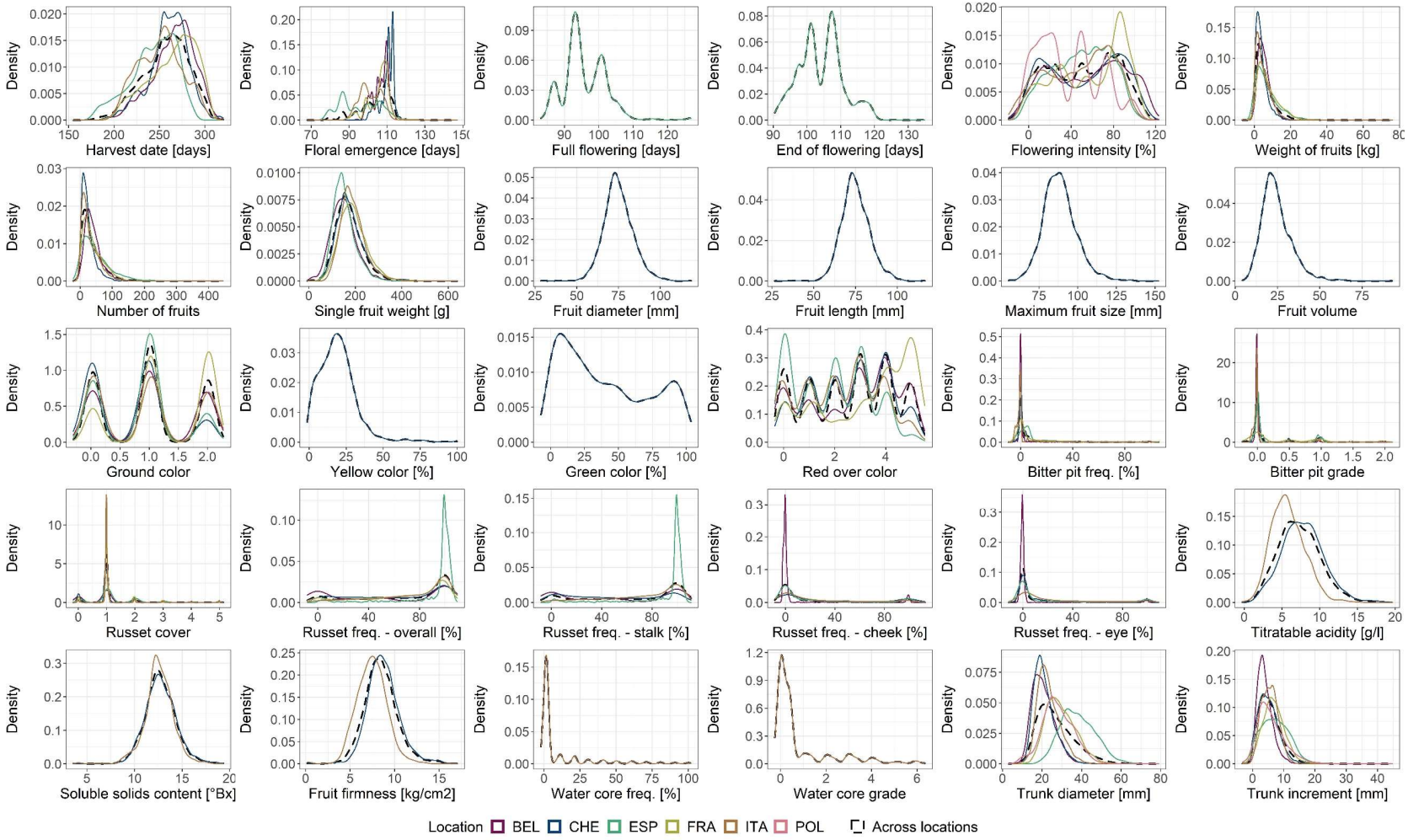
Vigor

Before the onset of flowering, *trunk diameter* in mm was measured 20 cm above the grafting point of trees using a digital caliper. Alternatively, trunk circumference was measured at the Italian site and transformed using the relation of circle's circumference to the diameter. *Trunk increment* in mm was obtained as difference between trunk diameter of the current and previous year.

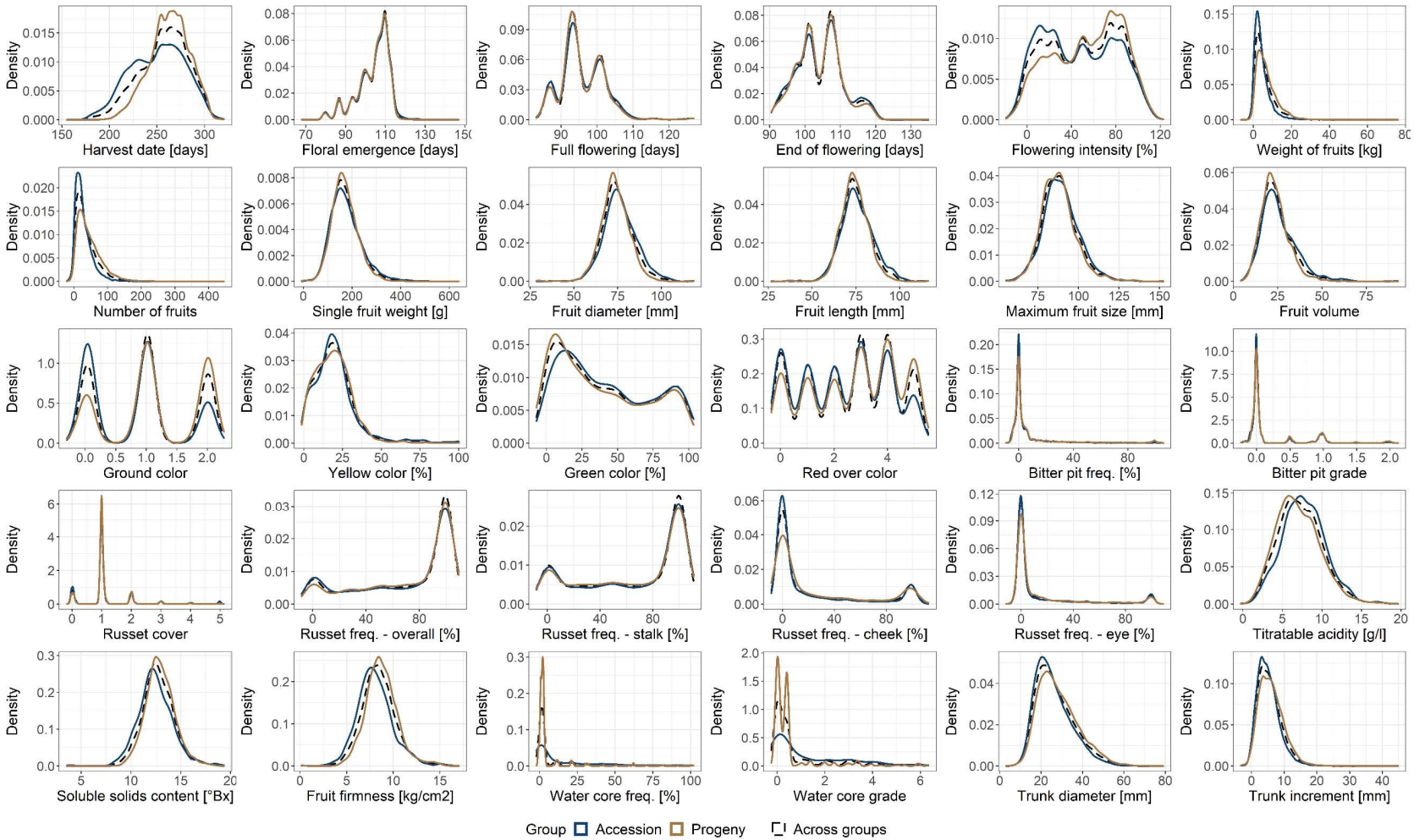
Reference

1 Buti, M. *et al.* Identification and validation of a QTL influencing bitter pit symptoms in apple (*Malus* × *domestica*). *Molecular Breeding* **35**, 29, doi:10.1007/s11032-015-0258-9 (2015).

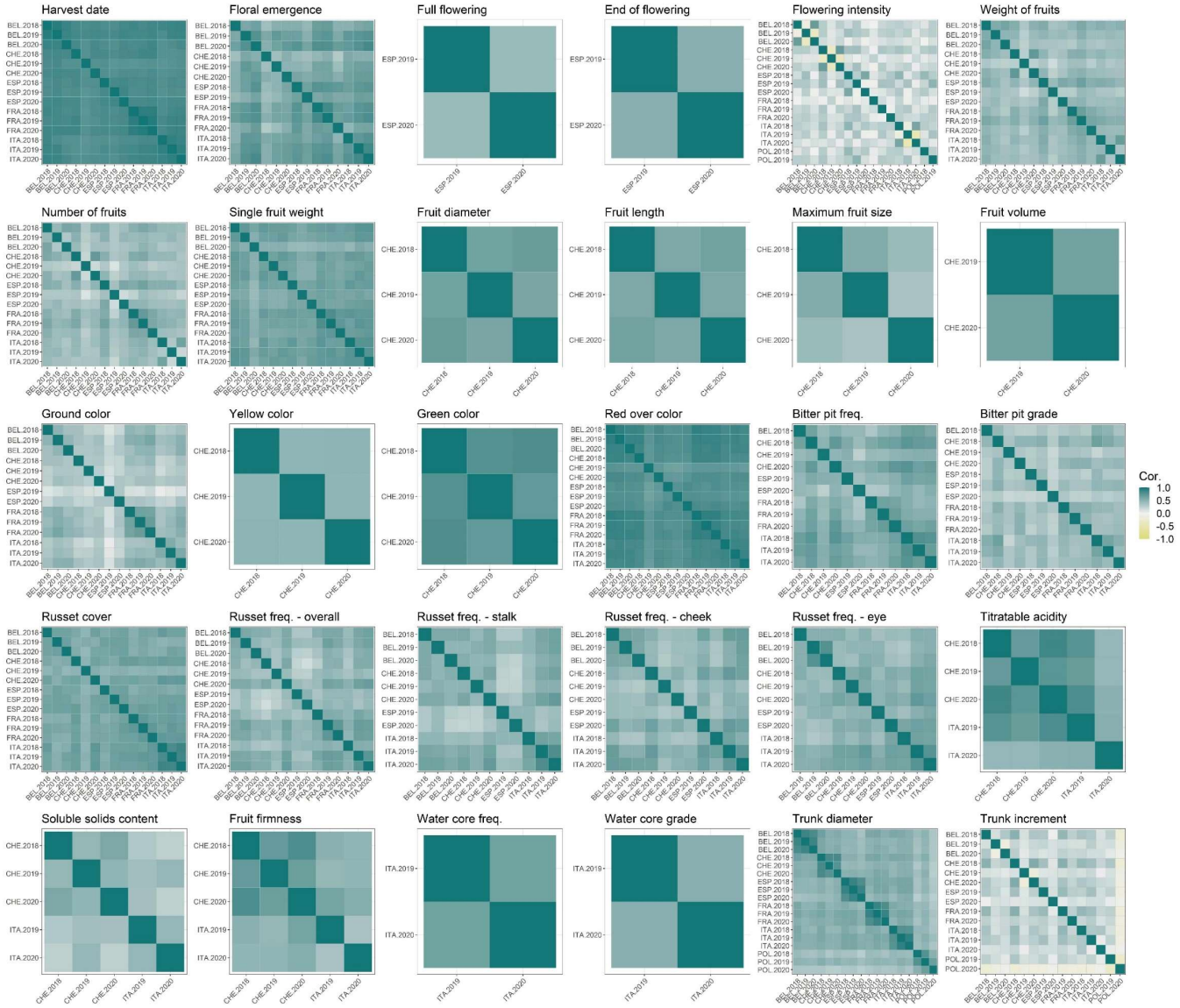
Supplementary figures



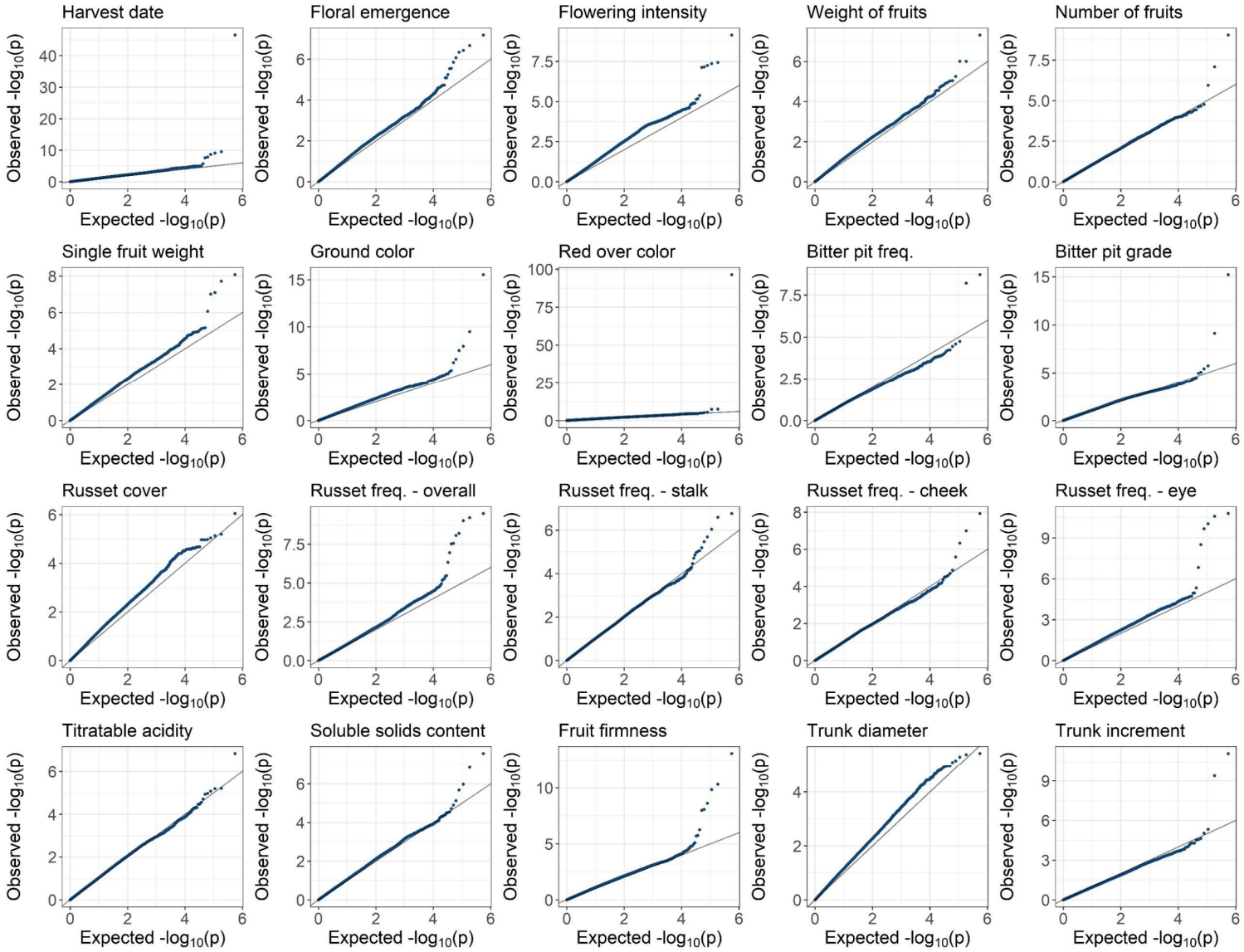
Supplementary Figure 1: Distributions of phenotypic values of traits adjusted for spatial heterogeneity within environments (adjusted phenotypic values of each tree), plotted per location.



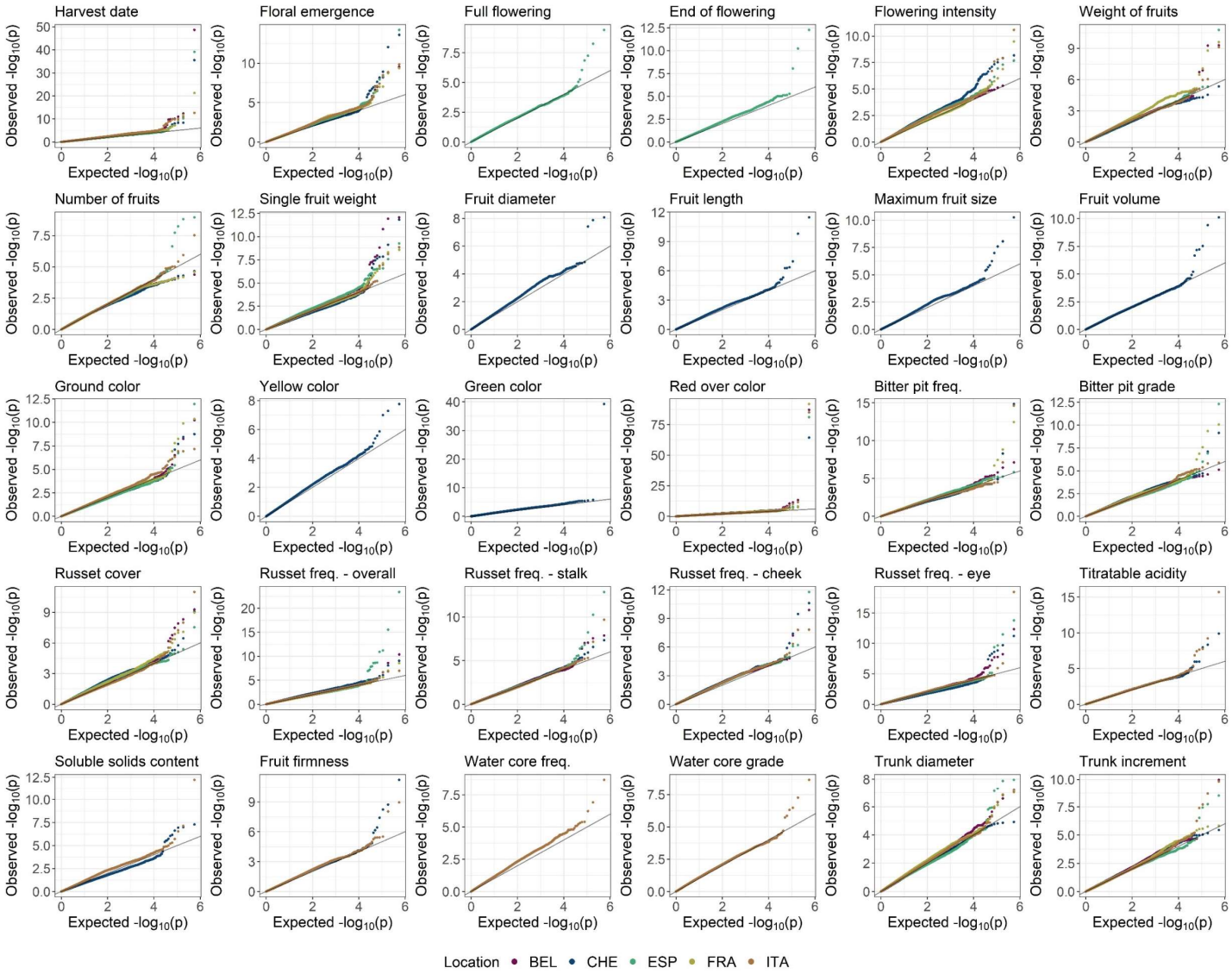
Supplementary Figure 2: Distributions of phenotypic values of traits adjusted for spatial heterogeneity within environments (adjusted phenotypic values of each tree), plotted per apple REFPOP group.



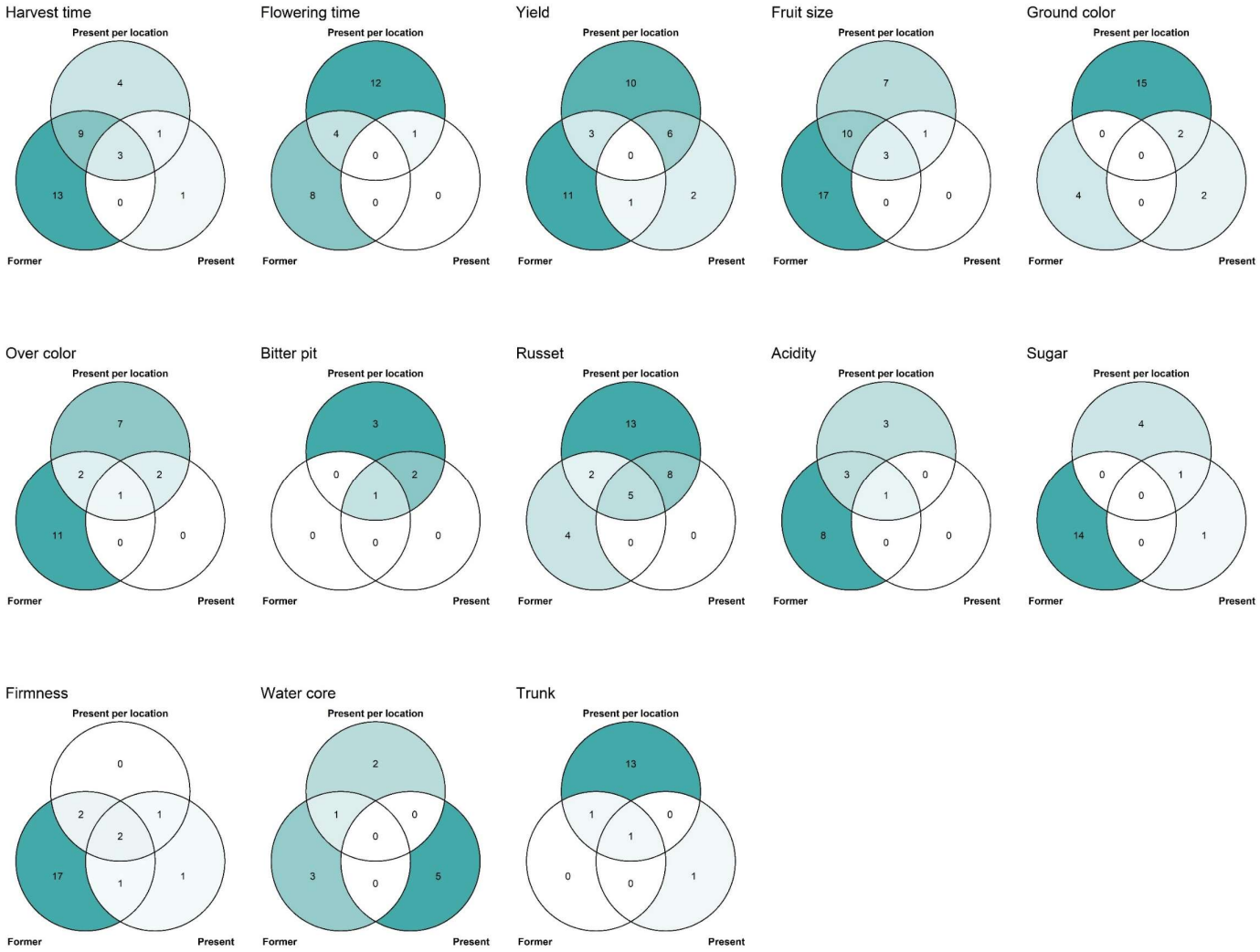
Supplementary Figure 3: Pairwise correlations of the adjusted phenotypic values of each genotype measured in different environments for individual traits.



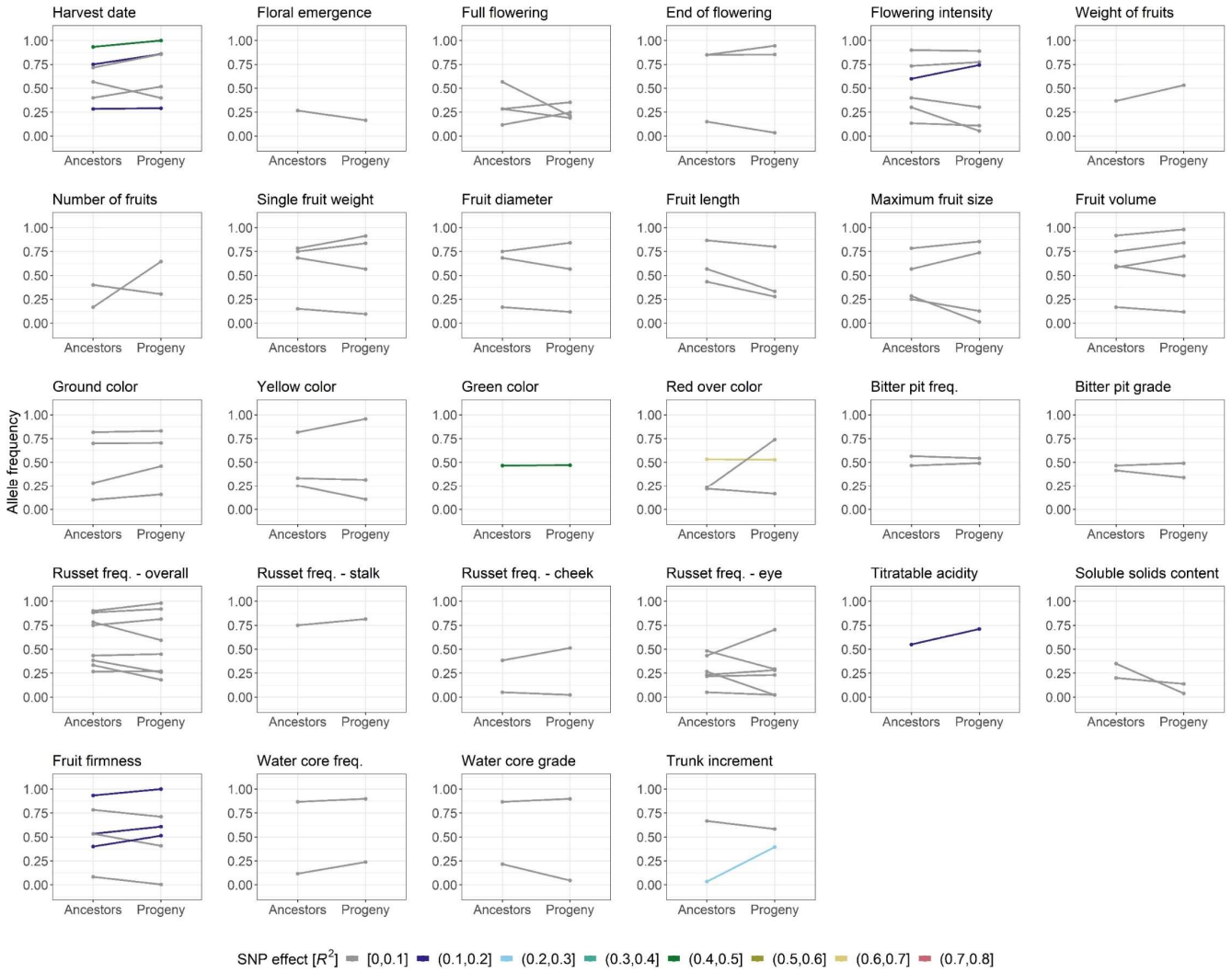
Supplementary Figure 4: QQ plots of the observed versus expected p-values for individual traits from the across-location GWAS.



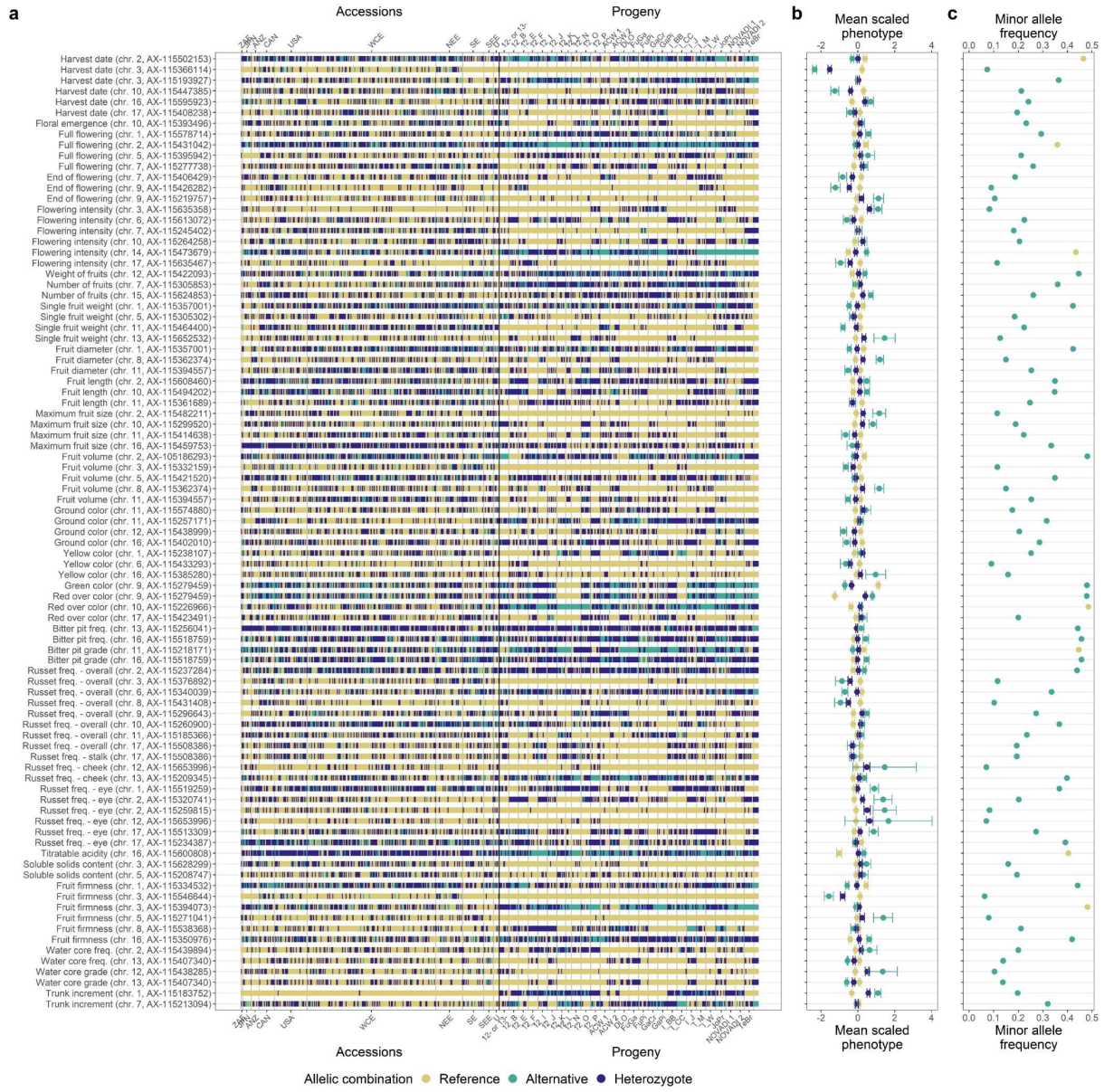
Supplementary Figure 5: QQ plots of the observed versus expected p-values from the location-specific GWAS for individual traits and locations.



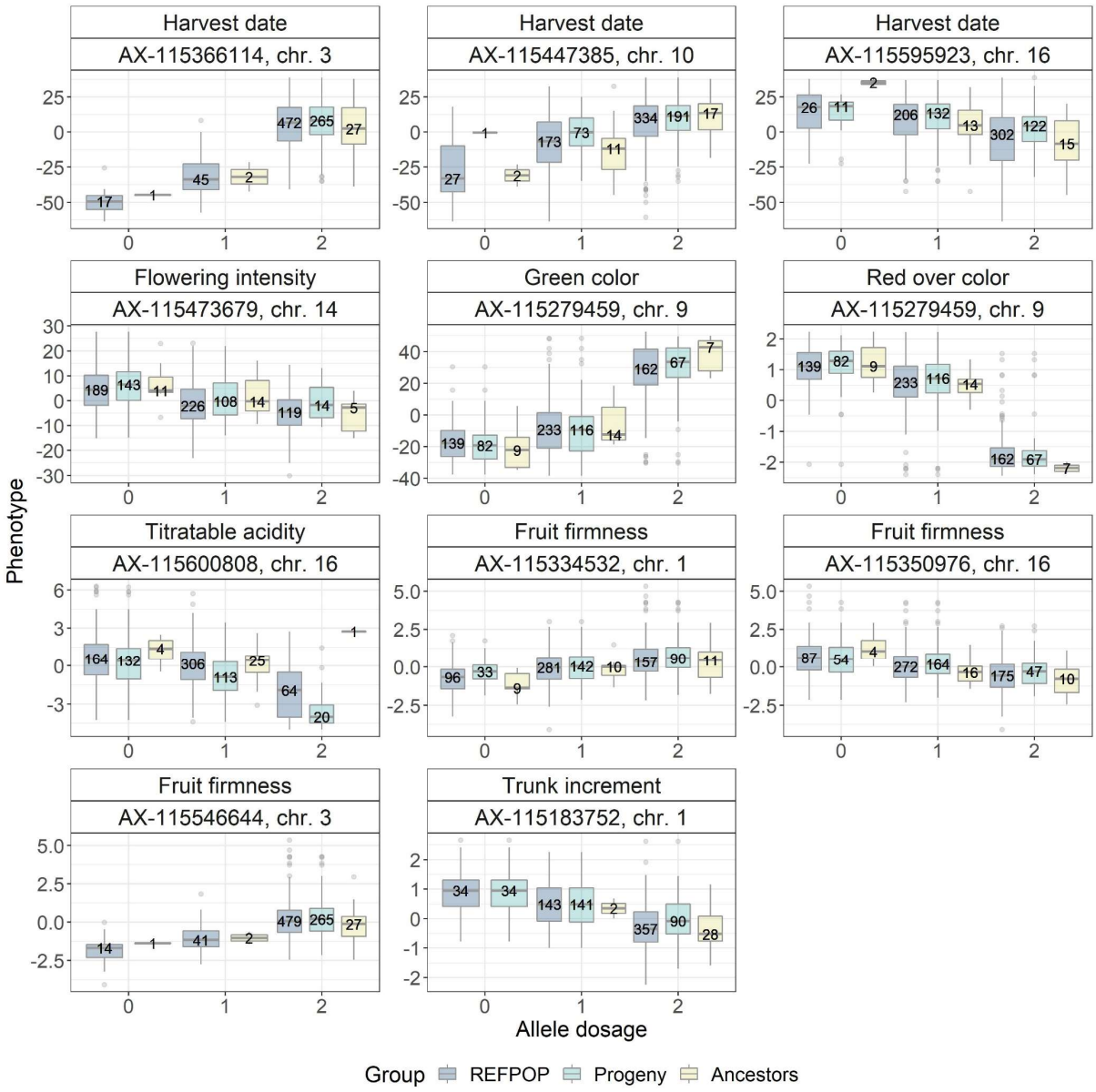
Supplementary Figure 6: Venn diagrams for each trait comparing the number of published associations (former, see also Supplementary Table 4) with the significant marker-trait associations from the across-locations GWAS (present, see also Supplementary Table 3) and location-specific GWAS (present (per location), see also Supplementary Table 3). The traits were assembled into trait groups based on their similarity. Color intensity reflects the number of associations per diagram area. The associations were assigned to chromosome segments (top, center, and bottom of a chromosome).



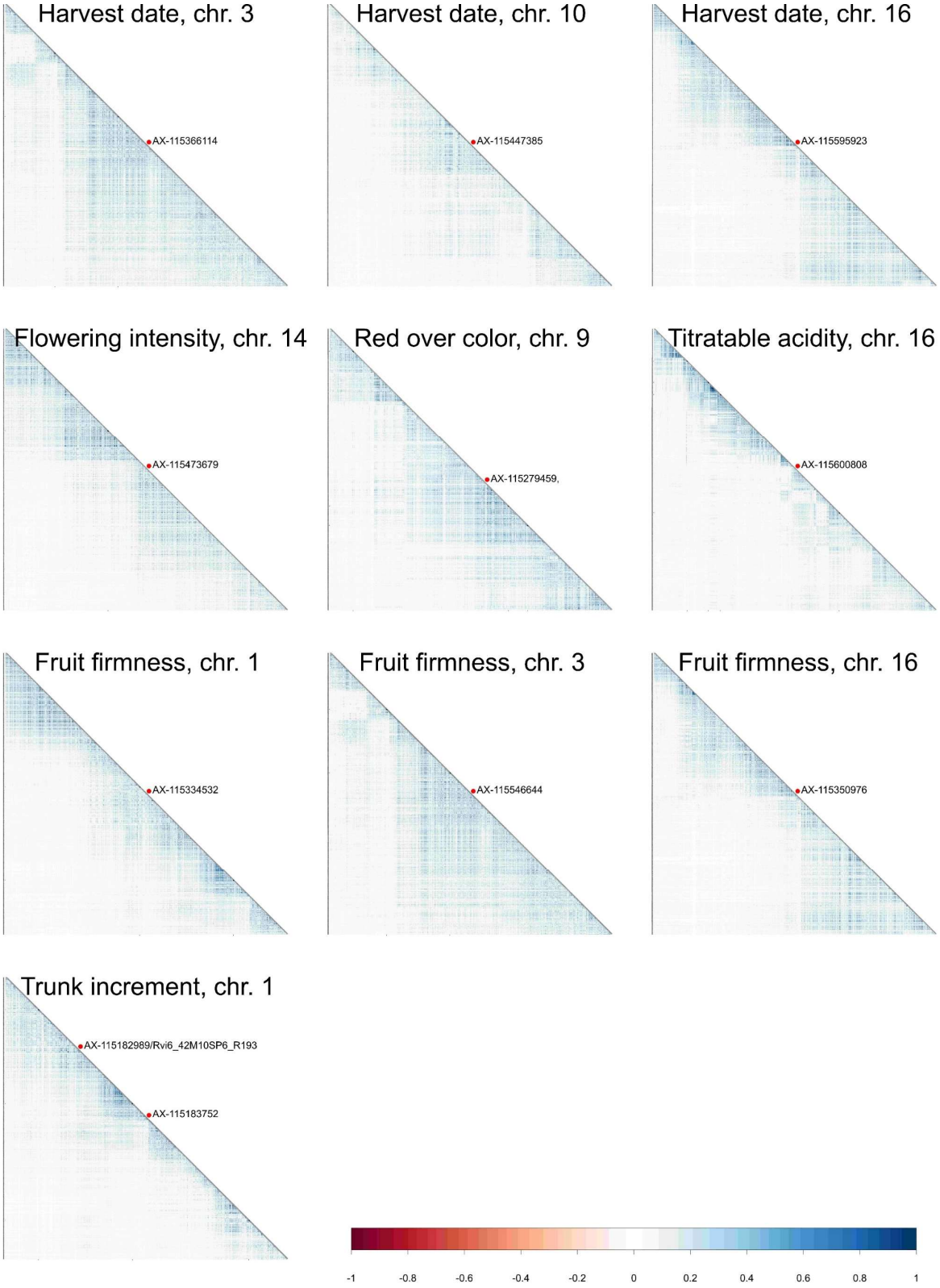
Supplementary Figure 7: Frequencies of alleles associated with increased phenotypic value for all significant marker-trait associations from the global GWAS. For the apple REFPOP progeny group (progeny) and its five ancestor generations (ancestors), the allele frequencies are shown as points connected with a line. Out of all known ancestors, the allele frequency was estimated for 30 accessions included in the apple REFPOP. Colors of the regression lines correspond to the part of phenotypic variance (R^2) explained by the associated SNPs.



Supplementary Figure 8: a Allelic combinations carried by the apple REFPOP genotypes, sorted according to geographic origin of accessions (269) and affiliation of progeny (265) to parental combinations (the x-axis was labeled according to Supplementary Table 1 and 2 in Jung et al. 2020). **b** Mean scaled and centered phenotypic BLUPs of traits and their standard error for each allelic combination. **c** Frequency of the minor allele in the whole apple REFPOP. **a-c** The legend and y-axis are shared between plots. In c, the color of an allelic combination corresponds to an allele of the same name. Presented are associations from the global GWAS (across-location GWAS with the addition of location-specific GWAS for traits measured at a single location only).



Supplementary Figure 9: Boxplots of phenotypes (across-location BLUPs) against dosage of the reference allele (0 – reference allele, 1 – heterozygote, 2 – alternative allele) for the major significant marker-trait associations, plotted for all apple REFPOP genotypes (REFPOP), the apple REFPOP progeny group (Progeny) and 30 ancestral accessions of the progeny group included in the apple REFPOP (Ancestors). Number of genotypes of a subgroup is shown in each box. Less than nine boxplots per trait indicate that not all specific allelic combinations were present in the respective group.



Supplementary Figure 10: Linkage disequilibrium estimated as squared Pearson's correlations in a window of 3,000 markers surrounding each of the major significant marker-trait associations. For the association with red over color, which corresponds to green color, only 2,736 markers were used due to the position of the association towards the end of chromosome nine. Position of a marker associated with apple scab resistance (*Rvi6*) is additionally shown in the plot of trunk increment. Physical size of the marker windows ranged between 4.1 Mb (harvest date, chromosome 10) and 5.8 Mb (trunk increment, chromosome 1).

Supplementary tables

Supplementary table 1

Trait	BEL. 2018	BEL. 2019	BEL. 2020	CHE. 2018	CHE. 2019	CHE. 2020	ESP. 2018	ESP. 2019	ESP. 2020	FRA. 2018	FRA. 2019	FRA. 2020	ITA. 2018	ITA. 2019	ITA. 2020	POL. 2018	POL. 2019	POL. 2020
Bitter pit freq.	1			1	1	1		1	1	1	1	1	1	1	1			
Bitter pit grade	1			1	1	1	1	1	1	1	1	1	1	1	1			
End of flowering								1	1									
Floral emergence	1	1	1	1	1	1	1	1		1	1	1	1	1	1			
Flowering intensity	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Fruit diameter				1	1	1												
Fruit firmness				1	1	1								1	1			
Fruit length				1	1	1												
Fruit volume					1	1												
Full flowering								1	1									
Green color				1	1	1												
Ground color	1	1	1	1	1	1		1	1	1	1	1	1	1	1			
Harvest date	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Maximum fruit size				1	1	1												
Number of fruits	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Red over color	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Russet cover	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Russet freq. - cheek	1	1	1	1	1	1		1	1				1	1	1			
Russet freq. - eye	1	1	1	1	1	1		1	1				1	1	1			
Russet freq. - overall	1	1	1	1	1	1		1	1	1	1	1	1	1	1			
Russet freq. - stalk	1	1	1	1	1	1		1	1				1	1	1			
Single fruit weight	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Soluble solids content				1	1	1								1	1			
Titratable acidity				1	1	1								1	1			
Trunk diameter	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Trunk increment	1	1	1	1	1	1		1	1		1	1	1	1	1		1	1
Water core freq.														1	1			
Water core grade														1	1			
Weight of fruits	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
Yellow color				1	1	1												

Supplementary table 2

Trait	Phenotypic correlation of environments		
	Mean	**Min**	**Max**
Harvest date	0.82	0.73	0.95
Floral emergence	0.62	0.30	0.84
Full flowering	0.42	0.42	0.42
End of flowering	0.52	0.52	0.52
Flowering intensity	0.18	-0.49	0.68
Weight of fruits	0.43	0.16	0.70
Number of fruits	0.42	0.10	0.69
Single fruit weight	0.63	0.43	0.79
Fruit diameter	0.63	0.59	0.67
Fruit length	0.59	0.57	0.62
Maximum fruit size	0.45	0.37	0.55
Fruit volume	0.62	0.62	0.62
Ground color	0.35	0.03	0.67
Yellow color	0.48	0.47	0.49
Green color	0.75	0.71	0.79
Red over color	0.80	0.62	0.92
Bitter pit freq.	0.55	0.26	0.78
Bitter pit grade	0.39	0.19	0.64
Russet cover	0.61	0.45	0.86
Russet freq. - overall	0.48	0.15	0.74
Russet freq. - stalk	0.48	0.23	0.79
Russet freq. - cheek	0.50	0.30	0.72
Russet freq. - eye	0.52	0.32	0.76
Titratable acidity	0.64	0.45	0.85
Soluble solids content	0.40	0.25	0.57
Fruit firmness	0.56	0.36	0.80
Water core freq.	0.63	0.63	0.63
Water core grade	0.52	0.52	0.52
Trunk diameter	0.52	0.28	0.91
Trunk increment	0.16	-0.31	0.55

Supplementary table 3

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115279459	Red over color	9	33799120	3.38387E-97	across-location	0.613	CAGCTGGACCCGCATTTCATTCTTGAGTCGGAAAG[A/G]CTCTCAATTGAAAAACATTTCATATGAATTGAAAC	A
AX-115366114	Harvest date	3	30681581	2.80176E-47	across-location	0.410	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115183752	Trunk increment	1	29560923	7.76978E-12	across-location	0.230	AGAGAGTTTCGTACTATAAATCGGTGAATAATCAG[A/C]TGGGAAACCCTGGAAGGTATAGAAATATTTTGGAC	A
AX-115447385	Harvest date	10	38871831	3.18069E-10	across-location	0.188	AAAATTTCAAAAA[C/T]CCCTCAAAGATTAGAAGAGCATGATGAACTGGGTA	C
AX-115473679	Flowering intensity	14	4229082	7.46136E-08	across-location	0.143	TTCAATGACAAAAAT[A/G]AGAAATTTGCGTCCGAAAACGAATTGAGAGGCGCA	G
AX-115595923	Harvest date	16	9023861	2.58471E-08	across-location	0.131	CCCATGATCCTCACCAAAAGGATCAGGAGAGGATC[C/T]TGTTGGATTATAATGGTCTCTAAACTTTACCCCTA	T
AX-115600808	Titratable acidity	16	3161268	1.35905E-07	across-location	0.126	CAAAGTAACCTCGGTTCTAATCCTCTCGTAATTCC[A/C]GTTATTC	A
AX-115546644	Fruit firmness	3	30590166	7.69832E-14	across-location	0.125	AGAGTATAACTCATGATCACACTAGTATTCCCGCC[A/C]AGTGTCCTC	A
AX-115350976	Fruit firmness	16	8985888	1.02258E-08	across-location	0.121	ACTTCCATGCAAAGTTTGAATGCTCGTTATTGACT[A/G]CATTCGAGGCCAAGGCCGGAGTGAATTCTTTTCAA	G
AX-115334532	Fruit firmness	1	26863095	2.46502E-09	across-location	0.120	CTCTGCTCAGACTCGACACAGTCTCTCCAAAATTG[C/T]ACCGGATCGGTCATCCAACCTGACTCGAC	T
AX-115624853	Number of fruits	15	32620209	8.35595E-08	across-location	0.094	TGTTTCTTAGGTTTTGACAACATTTCGAACACAGA[G/T]CTTTGAGGTGTTGGCGTGTTAAGCCTTTTGAAGAC	T
AX-115320741	Russet freq. - eye	2	29976071	9.13249E-11	across-location	0.084	AAAAATATATCTATGCTATGTTTCAAATGATTGT[A/G]AA	A
AX-115635358	Flowering intensity	3	31449413	4.38919E-08	across-location	0.082	GCTAAAACACTACAAAATTGTCAAGAAT[A/G]CTACACACTCCAAAAAGACTTGCCAGACTCACATC	A
AX-115653996	Russet freq. - eye	12	23013281	2.56834E-11	across-location	0.074	TTTGAAAC[A/C]GACCAGAGTTTTGAAGTGATCCACAAGTCCTCCCG	C
AX-115431408	Russet freq. - overall	8	8497351	3.00236E-08	across-location	0.073	AAGAAGACTTTCTTTGCTGCTGTATCCATTAGCT[C/T]ATTGTATAGTA	C
AX-115296643	Russet freq. - overall	9	10144814	3.38218E-10	across-location	0.069	GAAACAAATAAACAAAAAATCTAGGGACAGATG[A/G]TGAGATAAGAGACTAGGACATGTGACATGCCA	A
AX-115259815	Russet freq. - eye	2	19901873	3.16001E-09	across-location	0.066	CCTTCT[G/T]GATTTATATACATAGAAGAAGAAAATATAATAATA	T
AX-115376892	Russet freq. - overall	3	29745788	6.38774E-09	across-location	0.065	TGCTGAATAAACTTTGGACTGCTGATTTTGGGAAA[A/C]A	C
AX-115513309	Russet freq. - eye	17	5391171	1.64272E-11	across-location	0.064	ATGCTTCGCACCGTGCGTCGATGAAAACACAATTG[C/T]AACTTGCAAATCA	C
AX-115340039	Russet freq. - overall	6	25003040	8.6769E-09	across-location	0.062	GTCGAATTTGACATGTTACTAGTACAATTGCAGAT[G/T]TTAGAGCCAAGTTTTTTCTTATATGCATTCTAGAC	G

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115402010	Ground color	16	9500454	2.83902E-16	across-location	0.060	CTCTATGTGGTCACTGAACACTGAATACTATTGTC[C/T]A	T
AX-115652532	Single fruit weight	13	6291369	1.86455E-08	across-location	0.060	GTATA[C/T]CACTAAACGGTAATATAAGTGACATAAATTATTCT	T
AX-115357001	Single fruit weight	1	18679105	8.09772E-09	across-location	0.060	ACTT[G/T]GCTGAGATATTTGGACTTACTCAGATCTGA	G
AX-115613072	Flowering intensity	6	8916804	5.70475E-08	across-location	0.059	AATTCCTCTTAGTCATGATGGCCAAGCTAAGGTTC[A/G]CT	G
AX-115635467	Flowering intensity	17	10711135	7.13694E-10	across-location	0.059	GTCTTAAATCTTCTCCAGTAGATGGATTTGGAGTT[A/C]TCC	C
AX-115519259	Russet freq. - eye	1	28377243	2.13108E-10	across-location	0.058	ACTTCACTGTGAAGTGGGTTCTGGTTGAAAAGTGA[A/C]GCAGCGGAGGAGTGTGAGGAAGGTGATGATGATAG	C
AX-115422093	Weight of fruits	12	19704684	4.7469E-08	across-location	0.058	TTATTCGTTCAGAAGGTGACTACTT[C/T]TGCATTTAATACATGTTGCTTGGTACTGTTTTGCT	C
AX-115502153	Harvest date	2	13463592	1.60585E-08	across-location	0.057	AGGCCTCTGA[C/T]TAACATGCCGGCCATTAATCTTCCCCATTTTGCAC	T
AX-115218171	Bitter pit grade	11	6028089	5.88924E-16	across-location	0.049	TTTTGTAGGGTGGATATGTCGATTTATGCC[A/G]TGGTGTTTGGAGAATTCAGGGGTCTCTAGCTGTAT	A
AX-115209345	Russet freq. - cheek	13	3622205	1.00725E-07	across-location	0.049	ACTT[A/G]TTTGACATTTATAAGCATGATGTCATTATATTCGA	G
AX-115438999	Ground color	12	29469694	3.20499E-08	across-location	0.048	ACCAGTCCTACCTTGTAACTGTGGTACCTAAACCC[A/G]ATTTGTATAAC	A
AX-115653996	Russet freq. - cheek	12	23013281	1.19196E-08	across-location	0.044	TTTGAAAC[A/C]GACCAGAGTTTTGAAGTGATCCACAAGTCCTCCCG	C
AX-115508386	Russet freq. - overall	17	27249890	2.81705E-08	across-location	0.043	GTTAAGAAGTTCGAAGCAGACTC[A/G]AGTAGTTACATGGATCCTCTGGAATTGATTGCGAA	G
AX-115518759	Bitter pit grade	16	7681416	7.60953E-10	across-location	0.043	TACTTTTTAAAGGCCTCCCTTAGAGTCTCGCACAG[A/G]A	G
AX-115464400	Single fruit weight	11	14369160	7.9415E-08	across-location	0.042	ACTAACAAGTGAGCCTTAGTTAACTTCCATAGAGC[G/T]CACAAATGGCCCTCTA	T
AX-115226966	Red over color	10	39368708	4.61703E-08	across-location	0.041	CCTTATCATAAATAACAAGGCCTCCATGCT[C/T]TCCA	T
AX-115574880	Ground color	11	26298977	3.28094E-10	across-location	0.041	GTTGTAAGAATACAATCAAGCCGACTA[G/T]AAAATCTCAAACTTGAGAAGCTGTCTGTCACGGTC	T
AX-115260900	Russet freq. - overall	10	7885378	6.21121E-10	across-location	0.040	AAATAAATAATTACCGTATTCCACTGTTTGCATGA[C/T]AGATATTTAGTGCCAGACCTATGGAATACTCACTC	C
AX-115264258	Flowering intensity	10	40673251	3.64999E-08	across-location	0.040	TGTGTTC[G/T]TCAAGGATTGGAGCTCTTCTGTCTTTCTGCTGTGT	Heterozygote
AX-115508386	Russet freq. - stalk	17	27249890	1.72702E-07	across-location	0.039	GTTAAGAAGTTCGAAGCAGACTC[A/G]AGTAGTTACATGGATCCTCTGGAATTGATTGCGAA	G
AX-115518759	Bitter pit freq.	16	7681416	6.22548E-09	across-location	0.039	TACTTTTTAAAGGCCTCCCTTAGAGTCTCGCACAG[A/G]A	G
AX-115237284	Russet freq. - overall	2	28672556	9.74138E-10	across-location	0.035	TCCAAAGTCATTGAGTTCATCCCAATCTGGTCATA[C/T]GCTAAGCTGATTACGTATTTTTGGTTGGTCTGCCC	C

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115271041	Fruit firmness	5	4845617	4.90696E-11	across-location	0.032	GACTGAAATTTCAAGACCGCCCAAACAACATGA[C/T]GGAAACCAATCTCAGGCCAGA	T
AX-115408238	Harvest date	17	8129327	2.21559E-09	across-location	0.026	TGAGAATATCCTATAAATCGTCTTACGACTCACTA[A/G]CTATGTCGCTC	A
AX-115305302	Single fruit weight	5	23369304	9.63738E-08	across-location	0.023	AAAAAGGAGGACTGCATAAACTTAAGCGTCAACCT[A/C]CATGGGATTACAAAAACAAGTAAACTCAAACTGTT	A
AX-115628299	Soluble solids content	3	6160476	1.38586E-07	across-location	0.019	TGAGGTGAAAAAGCCGATTGAAAAAATTGTAAATG[C/T]GTCAACAAAAGCTACAGAGGTTGCTCCCGGATTGA	C
AX-115423491	Red over color	17	28350928	3.25396E-08	across-location	0.017	T[C/T]GTTCACTAGTTTTCCTTTCTTCAATTTGGCTTTGA	C
AX-115234387	Russet freq. - eye	17	645399	1.50672E-07	across-location	0.016	ATGCAC[C/T]ATTTGTTGTACATATAATTTAGTCTTGAGGGTCAC	T
AX-115208747	Soluble solids content	5	34132408	2.51884E-08	across-location	0.014	A[A/G]TGCTCAACCAGGCCTTTCTTTAATTGATGTCAAAA	A
AX-115393496	Floral emergence	10	22780819	6.65644E-08	across-location	0.012	CCACCAGGTAT[C/T]GAAACAATAGATCCATGGCGAGATGCAGGGACACC	T
AX-115257171	Ground color	11	4167945	1.12428E-08	across-location	0.009	CACATTTGGAGTTGATGCCGGACCCTCTTACCTGT[C/T]TCTCAATTAAG	T
AX-115538368	Fruit firmness	8	26763841	1.45446E-10	across-location	0.006	CCCAA[G/T]GAGCCGTTGATCATAAGATGCGTCAGATATGGTTA	G
AX-115185366	Russet freq. - overall	11	13011946	1.12257E-07	across-location	0.004	CCTGAGCACTTTTACCTGATAACTTAGCGAACTTT[A/C]GAGACTCTGCATCCTTAAAAGTCCCTGTATGTAGT	C
AX-115394073	Fruit firmness	3	30587378	8.60654E-09	across-location	0.003	TT[G/T]TTTACACCATATGCCACATGCATTTGTTGTCGTTG	Heterozygote
AX-115213094	Trunk increment	7	33577065	4.11476E-10	across-location	0.003	ACCAGCAAGCTTGGTGCTAATTCACCTGCCCTTGT[A/G]CCAGA	G
AX-115256041	Bitter pit freq.	13	3358666	1.9109E-09	across-location	0.002	GATTTGAAGTTCTTTGTTCTTTCTCTTAAGCTCCA[C/T]GACTTCCACCTCTAATTGATTCACAGC	C
AX-115305853	Number of fruits	7	25080162	9.14433E-10	across-location	0.002	ACGTGTGATCATTGTTCTTAAGAGTAGATTTCTC[C/T]A	Heterozygote
AX-115193927	Harvest date	3	25345071	8.14357E-10	across-location	0.000	TATGCAACAATGGCTTCTGATTTCAATGGTGTAAA[G/T]AAATTAGAGAATTATTAATTTATATATT	Heterozygote
AX-115245402	Flowering intensity	7	7854014	7.04279E-08	across-location	0.000	ATGGCTTTAGATATAATGGGGAAATTTTTGTGACC[A/G]GAATGGATCCTCAGGCTTTCATTAGGTTGTCAATA	G
AX-115279459	Red over color	9	33799120	7.85705E-88	location-specific BEL	0.613	CAGCTGGACCCGCATTTCATTCTTGAGTCGGAAAG[A/G]CTCTCAATTGAAAAACATTTCATATGAATTGAAAC	A
AX-115366114	Harvest date	3	30681581	2.38916E-49	location-specific BEL	0.410	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115447385	Harvest date	10	38871831	1.41573E-10	location-specific BEL	0.188	AAAATTTCAAAAA[C/T]CCCTCAAAGATTAGAAGAGCATGATGAACTGGGTA	C
AX-115546271	Harvest date	3	30728730	4.84518E-13	location-specific BEL	0.107	CAGTCCATTACTACTTATAAACACGTTGATGACAC[C/T]TTCTCCAGATCCACCTTGTTATAATATTTCTAAGC	C
AX-115505214	Weight of fruits	2	16341726	5.03608E-10	location-specific BEL	0.107	TGACTATCACCAAACTGAACTTTCGTCATAAGGCC[A/G]CTGGAGTAATTTTTTTTGCTTTATGTGTGTGTCCG	G

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115348723	Russet freq. - stalk	10	40798690	2.73345E-08	location-specific BEL	0.087	GATGAAAATTTAAACT[A/G]TCGAGAATATGTAATTACGAAACTTTGTTTGG	A
AX-115244829	Harvest date	16	9226272	2.07635E-10	location-specific BEL	0.085	GATCTTACCTAATTGAGTTACAAGCA[C/T]CTTGCATAAGTTCTTAATACTCAAAGCTTAT	T
AX-115348723	Russet freq. - overall	10	40798690	2.6032E-09	location-specific BEL	0.080	GATGAAAATTTAAACT[A/G]TCGAGAATATGTAATTACGAAACTTTGTTTGG	A
AX-115445253	Trunk increment	13	11177816	1.05772E-10	location-specific BEL	0.079	TTACA[C/T]ATTGATACAAGTTATCTACTTGAGCATATTGAAGT	C
AX-115214473	Red over color	10	36633322	5.67993E-08	location-specific BEL	0.074	TTGCATGACAACTTGAACTTATTAAAACCTTGAA[C/T]TCC	C
AX-115440975	Russet cover	9	11929448	3.30482E-08	location-specific BEL	0.072	TCCAAATCATTCTGCCTCTAATGATGTTTCTGTGG[C/T]ATCAGCTACTAT	T
AX-115266151	Harvest date	16	19381178	3.69172E-08	location-specific BEL	0.070	CGATCAACCTAGAGAAGAAGATGGAGGGAATATTA[A/G]ATGATAAAATTGAGATAAAATGGAAATGACGGATG	G
AX-115392999	Russet freq. - overall	10	16357902	4.27185E-11	location-specific BEL	0.068	TAATCGTTGTGGTTGTGTGATTGATGCAGAAGTTG[G/T]TTCTAGGGTTTGTGAAAG	T
AX-115248731	Floral emergence	2	12380892	3.03186E-10	location-specific BEL	0.067	AATTTGTTTGCA[G/T]TGAAGACTAATTGGTTGATCTTCACACCAAGTCCT	T
AX-115259815	Russet freq. - eye	2	19901873	1.9113E-08	location-specific BEL	0.066	CCTTCT[G/T]GATTTATATACATAGAAGAAGAAAATATAATAATA	T
AX-115198592	Floral emergence	9	1078953	7.04397E-09	location-specific BEL	0.065	TGCATCCGATGGTAATGGAAGGAGTTGATCATATA[C/T]AAACAACTAGAAACAATCGCTACAAAAAGGA	T
AX-115426282	Floral emergence	9	9761358	1.00091E-07	location-specific BEL	0.065	TGTATTTGCTTCAAAAATCTTGGTGTCGACCTTGA[A/G]GCAAAA	A
AX-115566226	Russet freq. - cheek	14	13868530	6.08932E-08	location-specific BEL	0.064	CACC[A/G]TGCCAAATGTTGGTGTCATACCCCAGATTTACCCA	A
AX-115260895	Trunk diameter	10	7884107	6.65339E-08	location-specific BEL	0.062	TGTAAGTTGCTCAGGTGTGTTTTAATGTTGGCCAC[C/T]GTTTTCTCCTCATAATATATCCTTGAGGAT	T
AX-115558433	Floral emergence	16	15243294	1.55141E-09	location-specific BEL	0.061	TAGTTGTGACAATGCAAAAAAGTTCTTAAATACT[A/G]TAGAACAAAACGT	A
AX-115519259	Russet freq. - eye	1	28377243	6.12113E-09	location-specific BEL	0.058	ACTTCACTGTGAAGTGGGTTCTGGTTGAAAAGTGA[A/C]GCAGCGGAGGAGTGTGAGGAAGGTGATGATGATAG	C
AX-115394557	Single fruit weight	11	18521895	1.23964E-12	location-specific BEL	0.056	TAACATATAGATGTACATAATAAATAAACGTCTTA[A/C]CTGGAGAAAGATAAAGAAAATTAGAAAAATTTAGG	C
AX-115361793	Red over color	1	27555798	3.99637E-11	location-specific BEL	0.055	GAT[A/C]TGAATTAAAGAATTTTGTGGATTTGGTAGAAAAAA	A
AX-115362374	Single fruit weight	8	2005502	4.02526E-08	location-specific BEL	0.053	AGTGATAATAAT[A/G]TGCATCCGTGTACTAAGACTTGAAAGCTAATACTT	A
AX-115392999	Russet freq. - stalk	10	16357902	8.76077E-08	location-specific BEL	0.051	TAATCGTTGTGGTTGTGTGATTGATGCAGAAGTTG[G/T]TTCTAGGGTTTGTGAAAG	T
AX-115359367	Single fruit weight	10	31588217	1.0401E-07	location-specific BEL	0.049	ATATAGGCCACTGTT[A/G]GCATCAAAGCTACAGACATTTGCCCTTCGATTTCC	G
AX-115619939	Single fruit weight	12	27138898	1.54541E-08	location-specific BEL	0.047	GCTTCCATTGAACACTTAACAAGAAGAATAGAAAA[G/T]ATTGCTTCTGGTTTTCTTTATTCTTTCAATAATTA	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115332159	Single fruit weight	3	30527005	1.55285E-09	location-specific BEL	0.045	ACTTTTATGGCAATTGGCAAAAAGACAATATATAC[G/T]CATT	T
AX-115514706	Russet freq. - eye	9	12020921	4.72091E-13	location-specific BEL	0.041	ATTGACTCGAACCCATTTTTTATGACTCGGGTTGG[A/C]A	C
AX-115318661	Flowering intensity	6	7687982	2.16184E-08	location-specific BEL	0.040	ATTGCACTGTGGAGGTATAATTACGCAGACACAAA[G/T]AGAGGAAGAGAAGGCTAGCAGAAA	G
AX-115446281	Russet freq. - cheek	9	5791395	1.3303E-10	location-specific BEL	0.039	TACTTTCCAGCTGGAGAATCCTGATAATATCGTGT[C/T]GGATTACGGGAATTAGCATTTTGATTACGTTTGGA	C
AX-115490433	Harvest date	5	23224014	8.95261E-11	location-specific BEL	0.039	TGAACAGCTGCCCCAGCCAATAGTTTCACAAGTTT[A/G]AAAGTAAAACTGTTTGCAGGTATTTAGTTCAATCA	A
AX-115254039	Single fruit weight	4	845431	8.77584E-13	location-specific BEL	0.037	TCCATAGCTGTATTTAGTTGTTTAAAACAATCTT[G/T]TAC	G
AX-115224250	Russet freq. - stalk	11	2859820	1.3436E-08	location-specific BEL	0.037	TCAAAACCCTAACCCTAAACCTGAAAAGTAAAGAA[G/T]ATGAGGTTTATAAGGAAGGGGCCTGTTTGGTTTGC	T
AX-105189206	Harvest date	2	13525162	1.25992E-11	location-specific BEL	0.036	GCTTGGGTTTTGGAACCAAAACACCTTAGCGTGTG[C/T]TGGGTGCTTTTCGCTGGAGAAGAACGCCAGACCTG	C
AX-115226972	Red over color	10	39365479	3.02293E-08	location-specific BEL	0.035	CCTACCTGTGAAAAAAGTACATAAACAGATAATGG[C/T]GAGTTCTTAAATTATTAGAACTAATACAATATCTA	C
AX-115348723	Russet cover	10	40798690	5.02782E-09	location-specific BEL	0.035	GATGAAAATTTAAACT[A/G]TCGAGAATATGTAATTACGAAACTTTGTTTGG	A
AX-115657805	Bitter pit freq.	9	5169418	7.63845E-08	location-specific BEL	0.035	AAAGTTATTT[C/T]TCTAGGCTTTTTATCTGATATGACAAATTGTAATA	Heterozygote
AX-115321581	Ground color	16	9205152	5.77271E-11	location-specific BEL	0.034	CCCAATCGCAACCAGAACTGCGTCGTT[C/T]ACCGGACTGGTGAAATGAAGAGACAGAATAGTGGG	T
AX-115464340	Weight of fruits	9	6693742	1.30981E-07	location-specific BEL	0.034	TGAGGAAAGCTTGATTTGTGGAGAAGGTTACGCTA[C/T]TCTGCTGA	T
AX-115407736	Russet freq. - cheek	12	22271380	1.55468E-08	location-specific BEL	0.033	TTGGGGGTTAAAGCTGCA[C/T]TATCTCTACTCAAATTCTACAGAAGTAAGTTTCT	T
AX-115519250	Single fruit weight	11	37619351	5.56817E-08	location-specific BEL	0.028	AGTGGCATGAAAGAGTTTCTATTAAACCAGCAAAA[A/G]TGAAAATTCTTAGCCGTT	Heterozygote
AX-115429780	Single fruit weight	16	8477513	1.60297E-11	location-specific BEL	0.028	AGTTGGGTTTTTAGAATTTTGTGATTTTATCATCT[A/G]AGTTAAAATGTCGGAAG	A
AX-115329383	Russet freq. - stalk	12	28831881	9.73611E-08	location-specific BEL	0.025	GAGGATTTAGAGGAGGATTACTGTTCTTCTTTGGA[C/T]GGAGATGCGTCGGAAGAGAGCTTTGGGAGTTTTGT	T
AX-115331167	Russet cover	2	29383575	5.29282E-10	location-specific BEL	0.024	GTTTTGCTGGTATTAATTGATAGTTGTTACATACA[A/G]GAGTTATGTTGTTAGATAAAAATAATAGATAGAAC	G
AX-115424668	Floral emergence	5	5989381	6.53172E-09	location-specific BEL	0.022	GGGGATGCTTGTGTGTTTGAGTTGATTAATAAGGG[C/T]A	Heterozygote
AX-115633382	Single fruit weight	1	27848521	1.09353E-08	location-specific BEL	0.018	CTCTAATTTATACACTTCTGGTGCATGTATGATCA[C/T]TTACACATATATTACTTTGATTTCCTTTGCAAGTA	C
AX-115452523	Red over color	17	3769755	5.00088E-12	location-specific BEL	0.018	GTGAGTACTGA[A/G]AAAATTCATGTGCATTATTATTTCCCACAATTTAC	A
AX-115413715	Trunk increment	15	43700335	2.02007E-09	location-specific BEL	0.016	CTACCA[A/G]AAAGAGTGAAAGGATAATGATCAAAACGGAATATC	A

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115550974	Red over color	17	28309860	5.04183E-14	location-specific BEL	0.015	TCGTGGATTCTGTTTTACAATTAAGTTGTTTTGAA[G/T]TATATACTTCTCTCTACTA	T
AX-115192200	Single fruit weight	2	3679888	6.01167E-08	location-specific BEL	0.012	TGAAAGGGAACCAAATCAAATGACGAAAAAATACA[A/G]AGGCACACAGAAGATAAAAATTATATATATGAAAT	G
AX-115454123	Russet cover	5	38269446	1.25419E-08	location-specific BEL	0.011	CGATCATCTCTCGCTAGATGA[A/C]TTTGAGCATTAAAGATTCGGCTTAATAGCTGCGAA	C
AX-115236078	Red over color	3	4366924	4.0507E-09	location-specific BEL	0.010	TGTCCCTTCTTGTATCATTGCGTATAATTGCTC[C/T]CTGACCTAATCACTTTCTTGATGAATTATTGAAGC	T
AX-115502263	Russet cover	15	43411698	1.75362E-07	location-specific BEL	0.010	TATGTGTAAAAAAGTTAGATT[A/G]TCAGTAAATGTGCTCTTCCTAATATTTTGTGCCAT	Heterozygote
AX-115298101	Russet freq. - eye	2	20019589	1.70723E-08	location-specific BEL	0.009	TCTGAATCATCCAAAGGGCAGCAGCTCCATTGGCT[A/G]ATTCCTCTAGGTTCTGGGAGACGAAGAGAATTGGC	G
AX-115524456	Harvest date	1	20772639	1.41145E-10	location-specific BEL	0.008	ATT[A/G]GTTGAAATAATATACTCAGGTGGCTGATGTGTGGA	G
AX-115484911	Ground color	6	1521834	1.67E-07	location-specific BEL	0.007	TAATTCGAAAAGTGAGTAGTAACGGTAGAAAGAAA[A/C]CCAGAAACAAT	C
AX-115205135	Weight of fruits	12	2651443	5.46007E-10	location-specific BEL	0.001	A[A/G]TTCTTGGAAAAAGTTGATCCAAAAACTAATTAAAA	A
AX-115241162	Russet freq. - eye	3	22353748	5.65441E-08	location-specific BEL	0.001	AAGCAAGGTTTGAGCTCTTAATTTCAATTTCTTTT[C/T]ATTAATTGTTTTTCAATCTTATATTGAA	Heterozygote
AX-115649072	Ground color	15	33950748	5.64765E-09	location-specific BEL	0.000	TTCTTTTGATTTGATCTGTGTAAGCTGTAATCGAT[C/T]GTCATCTTTTGAATAATGGTGCGTTTACTAATC	T
AX-105213720	Red over color	9	33801013	4.2956E-65	location-specific CHE	0.615	TGACTATTGTGGAGCAGGAGGCGAACCATAA[C/T]GGGCCGGTTTTCTTGGACCGGTTCACGGAGGCATT	T
AX-115279459	Green color	9	33799120	6.05613E-40	location-specific CHE	0.480	CAGCTGGACCCGCATTTCATTCTTGAGTCGGAAAG[A/G]CTCTCAATTGAAAAACATTTCATATGAATTGAAAC	G
AX-115366114	Harvest date	3	30681581	3.24954E-36	location-specific CHE	0.410	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115549406	Harvest date	10	38832477	2.91137E-08	location-specific CHE	0.188	GCCC[A/G]GCTTGGGTGAGTCTCACAATTGATAAAATTAAATT	A
AX-115473679	Flowering intensity	14	4229082	4.53847E-08	location-specific CHE	0.143	TTCAATGACAAAAAT[A/G]AGAAATTTGCGTCCGAAAACGAATTGAGAGGCGCA	G
AX-115492821	Harvest date	16	8925608	5.64452E-09	location-specific CHE	0.132	TTGGTTGTCATGGCTGCCATTATTCATCATTTTCG[C/T]GGAAGAAGACAAGGGTTTTGTGACAGA	C
AX-115239448	Titratable acidity	8	7415559	1.27433E-10	location-specific CHE	0.121	GCAA[C/T]TTTCACGGATAAGTAATCGAACTCTTTGAAACCAT	C
AX-115350976	Fruit firmness	16	8985888	1.90718E-09	location-specific CHE	0.121	ACTTCCATGCAAAGTTTGAATGCTCGTTATTGACT[A/G]CATTCGAGGCCAAGGCCGGAGTGAATTCTTTTCAA	G
AX-115334532	Fruit firmness	1	26863095	4.9927E-12	location-specific CHE	0.120	CTCTGCTCAGACTCGACACAGTCTCTCCAAAATTG[C/T]ACCGGATCGGTCATCCAACCTGACTCGAC	T
AX-115250644	Ground color	11	9490564	3.68214E-09	location-specific CHE	0.105	TAATTCTGCTCCGACTTCGACGGCTTCAGATTCTC[A/G]TTGAACAGAGCGAAGATGTAGGTCTCGAACGTCCG	G
AX-115630047	Floral emergence	9	530386	8.31084E-09	location-specific CHE	0.102	ATTAGTCGCGTTTTTCCTTTAATTTTGAGATTGTA[G/T]AGTA	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115325408	Ground color	10	41258346	1.79748E-09	location-specific CHE	0.100	GGCACTTTATCATGAC[A/G]CCATGAGCTTGCTTCCCATGCTCCAGTGTTGCTAA	G
AX-115320293	Floral emergence	11	11333668	9.56406E-13	location-specific CHE	0.094	AAATTAGTCTCTCCACCAAAGCA[A/G]TTATAGGCAATAATATCAAAATCAATGTAAGAAGT	A
AX-115344079	Floral emergence	8	10638158	1.16329E-07	location-specific CHE	0.087	TTTCT[A/C]CTTATATCCCCGAAAAAGACACATTTCGTCTTGAT	C
AX-115348723	Russet freq. - stalk	10	40798690	4.58324E-08	location-specific CHE	0.087	GATGAAAATTTAAACT[A/G]TCGAGAATATGTAATTACGAAACTTTGTTTGG	A
AX-115320741	Russet freq. - eye	2	29976071	4.35339E-08	location-specific CHE	0.084	AAAAATATATCTATGCTATGTTTCAAATGATTGT[A/G]AA	A
AX-115562961	Bitter pit freq.	16	3471935	1.46037E-15	location-specific CHE	0.082	TATAGTTT[C/T]GTCCAACCCCAAATAATAATGGCAAATAGGAGTTT	T
AX-115421795	Titratable acidity	8	12452779	1.139E-08	location-specific CHE	0.082	AATTTGAGGAACTTTGAT[G/T]ATTTCTTGTCAAATTTGGATGGGAAAACTAGAAGG	T
AX-115328653	Flowering intensity	11	42193859	1.19273E-07	location-specific CHE	0.077	TGCAAACAAACCCTACAACCCCCGCCTCTCTGCCT[A/C]AAGAGCATCAAAGGAGCAGCCGTAATGGTAATCTC	C
AX-115562961	Bitter pit grade	16	3471935	7.36441E-10	location-specific CHE	0.069	TATAGTTT[C/T]GTCCAACCCCAAATAATAATGGCAAATAGGAGTTT	T
AX-115534786	Russet freq. - cheek	14	25600768	2.46561E-11	location-specific CHE	0.069	CGCGGAAACACGATGGAGGACGCAATTGATGCTGA[C/T]GGAGAAAAGGG	C
AX-115357001	Fruit diameter	1	18679105	1.32998E-08	location-specific CHE	0.066	ACTT[G/T]GCTGAGATATTTGGACTTACTCAGATCTGA	G
AX-115494202	Fruit length	10	26445213	1.08112E-07	location-specific CHE	0.062	ACATGCTTTGCAACTGATAAGCAAACTTCTTCTAC[C/T]GAAATCAGGTCTTCCCTTTTCCAGT	C
AX-115653957	Russet freq. - eye	12	22692673	6.12335E-12	location-specific CHE	0.061	TCAACTTGCATTAGATGTGTTTTTTCTTTAAATAT[A/G]GAATGCTA	A
AX-115617155	Soluble solids content	13	41925221	1.04094E-07	location-specific CHE	0.061	TAATTTTGAGGTCTACA[A/G]TGATGCTTCCATAAATGGTCTGAGTTGTGTGTTGA	A
AX-115608460	Fruit length	2	29824044	3.46481E-12	location-specific CHE	0.060	GCCCCATATCAAAGTAAGGATCAATATGTGGCAGC[A/G]GCGGAGAAACTCTTGACCCAGAATCTGCATAACGC	A
AX-115357001	Single fruit weight	1	18679105	1.48324E-08	location-specific CHE	0.060	ACTT[G/T]GCTGAGATATTTGGACTTACTCAGATCTGA	G
AX-115299520	Maximum fruit size	10	35060486	2.64106E-08	location-specific CHE	0.059	TCACAAATACAATTCCTCAGCTCAACAATACTGTA[C/T]CCGGGA	T
AX-115519259	Russet freq. - eye	1	28377243	1.03039E-08	location-specific CHE	0.058	ACTTCACTGTGAAGTGGGTTCTGGTTGAAAAGTGA[A/C]GCAGCGGAGGAGTGTGAGGAAGGTGATGATGATAG	C
AX-115362374	Fruit diameter	8	2005502	3.82136E-08	location-specific CHE	0.058	AGTGATAATAAT[A/G]TGCATCCGTGTACTAAGACTTGAAAGCTAATACTT	A
AX-115361689	Fruit length	11	3622193	1.62096E-10	location-specific CHE	0.057	TAAGAAATATGTGTAGTTTTACATATCAATACTCG[A/G]A	G
AX-115394557	Single fruit weight	11	18521895	1.56364E-12	location-specific CHE	0.056	TAACATATAGATGTACATAATAAATAAACGTCTTA[A/C]CTGGAGAAAGATAAAGAAAATTAGAAAAATTTAGG	C
AX-115361689	Single fruit weight	11	3622193	2.32619E-08	location-specific CHE	0.055	TAAGAAATATGTGTAGTTTTACATATCAATACTCG[A/G]A	G

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115384077	Soluble solids content	15	6120745	1.16028E-07	location-specific CHE	0.054	AGGCAATTGATGAGATTTGAGACATTTTTTAATTT[A/C]A	C
AX-115457259	Harvest date	16	8703849	4.70625E-09	location-specific CHE	0.053	TACAAGACTTTCTAGATGCCTAGGTCCCTCAGAAA[C/T]AGTGTGATCAGAATAAGAAATACAAACATAACATT	T
AX-115647524	Russet freq. - eye	2	19952878	2.53686E-09	location-specific CHE	0.053	TCTGCAATAAGCTTGTCCACTTGTACTTCAAGTTT[C/T]CTTGACAAACTTTTGTAATCAAACTCTTCCTTAAT	C
AX-115322641	Soluble solids content	5	46842414	5.17134E-08	location-specific CHE	0.051	CATGCCCAAATTTCCCCTT[G/T]TAAATATAATTTGTTAGCTGATGTTTATTGCCTTA	T
AX-115362374	Fruit volume	8	2005502	6.0375E-08	location-specific CHE	0.051	AGTGATAATAAT[A/G]TGCATCCGTGTACTAAGACTTGAAAGCTAATACTT	A
AX-115422758	Floral emergence	13	14014367	2.62679E-14	location-specific CHE	0.050	GAAACTGAATATTGATATTATGACTATTTCTTAAA[C/T]TGGACATTGATGCAAATTTCCCTTAATAATATGAT	T
AX-115266971	Russet freq. - cheek	1	28595992	3.45861E-10	location-specific CHE	0.049	CACATTACACCAAGTATGAAGAATAGACAGTAATC[A/C]AATGCTAAAAG	C
AX-115264729	Red over color	15	54411796	6.39673E-08	location-specific CHE	0.046	CTTATCCAGTTGATTTTCATATATGTTCTTCAGAG[A/G]GAATCTGCATTGCCTAAATCAGCAGAGATGATCGG	A
AX-105186293	Fruit volume	2	10322930	2.96643E-08	location-specific CHE	0.046	CGAGAGCTTCAACATTTTCTCTCCCTCTCCTCACG[C/T]TCTCTCTCTAGATTGAAGGAAGAGATCTATAGTGA	T
AX-115206819	Red over color	11	265336	1.30008E-07	location-specific CHE	0.045	TTAATTCAATTTAAAGTTTTGCAATTATTTATAGA[C/T]TACTAAAAGAAAACAATTGTTCCAACTTCAAATGC	C
AX-115316232	Flowering intensity	5	1987082	6.60491E-09	location-specific CHE	0.044	C[A/C]TTTCCCCTTTCTCTCTTCCGTTAATGGAGAC	A
AX-115508386	Russet freq. - overall	17	27249890	8.82466E-09	location-specific CHE	0.043	GTTAAGAAGTTCGAAGCAGACTC[A/G]AGTAGTTACATGGATCCTCTGGAATTGATTGCGAA	G
AX-115414638	Maximum fruit size	11	17878463	5.68812E-11	location-specific CHE	0.042	GTAATGTTATTTCCATTTGATGCAGACATGATTAC[C/T]CATCTAATAAAATCTCCTTTGAACTAGTCAGCTAT	C
AX-115482211	Maximum fruit size	2	21759081	1.03901E-07	location-specific CHE	0.040	AAGTTTCTAATATTTAAGGATCATCTTACATAAAG[A/G]CA	G
AX-115433293	Yellow color	6	24024503	5.22773E-08	location-specific CHE	0.039	CAGAATGTACAGCCCTTTAGACCCAAAAATAATGG[A/C]ACATTCAGTTTCTAGATGTAGAAGCAATATCCTAA	C
AX-115650478	Single fruit weight	5	17738796	8.71768E-08	location-specific CHE	0.039	CAAAATGGTCATTTGACCATTTATTTAGAAAGCTC[C/T]AGCTTTGCTGAGTAGCCATCTGGGGTGCCA	T
AX-115391701	Floral emergence	5	10353882	1.7818E-07	location-specific CHE	0.036	TG[A/G]TGTTTCATTGAATTTAAAACTTCGTTGTACAAAC	G
AX-115292619	Russet freq. - overall	1	29612376	8.60933E-10	location-specific CHE	0.036	TTTTCACAATAATGTTCAG[C/T]TACTATTGAAATGATTAATGTGAAGCATATTCTTT	C
AX-115394557	Fruit volume	11	18521895	7.80037E-11	location-specific CHE	0.035	TAACATATAGATGTACATAATAAATAAACGTCTTA[A/C]CTGGAGAAAGATAAAGAAAATTAGAAAAATTTAGG	C
AX-105206365	Russet freq. - cheek	5	36728628	4.00466E-08	location-specific CHE	0.035	AGCATTATGTTAAACAAAAAATCTAAATGGGAAAA[C/T]A	T
AX-115238107	Yellow color	1	16168825	1.73499E-08	location-specific CHE	0.035	TCATA[A/C]CATGTTGCTGGAATTCACAAATGAAGGCAGCGGTT	Heterozygote
AX-115332159	Fruit volume	3	30527005	6.88782E-08	location-specific CHE	0.033	ACTTTTATGGCAATTGGCAAAAAGACAATATATAC[G/T]CATT	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115394557	Fruit diameter	11	18521895	8.44122E-09	location-specific CHE	0.031	TAACATATAGATGTACATAATAAATAAACGTCTTA[A/C]CTGGAGAAAGATAAAGAAAATTAGAAAAATTTAGG	C
AX-115485291	Single fruit weight	2	10051576	3.74468E-08	location-specific CHE	0.030	GGTAACTATCTTTACCCAAAATAG[A/G]TAACTAACAGTACTTACGGTTTCTGCATTCTGTGA	G
AX-115461265	Flowering intensity	10	18725252	1.20489E-08	location-specific CHE	0.030	TGTTTTTACCAAGTGGCTGATAGAGCAAAGATGCA[A/G]TTTTGTGATTCTTAACGGTATGACTTGAGGTGGAA	G
AX-115321105	Fruit firmness	8	25362682	5.85875E-09	location-specific CHE	0.025	TTTACTGTTTAACTTAAAACCAACCATGAG[C/T]TAA	T
AX-115508423	Russet freq. - eye	17	27235275	1.47726E-09	location-specific CHE	0.025	CCATGAA[A/G]GGAATCGTTTTCCAGAAAGTGAGAGAAAATTAGGT	G
AX-115353120	Russet freq. - eye	5	45018392	1.99931E-10	location-specific CHE	0.024	GAAAATCAAAGAGACATTACCCATCTGT[C/T]AAAGAGGCATTTAAGGACGAGCCATCAGCCGGCAG	T
AX-115551098	Red over color	6	2585444	2.12949E-08	location-specific CHE	0.023	TACTTGCAACAGTACAAGGTTATTGTAATATAGCA[A/G]CTCTAG	Heterozygote
AX-115385280	Yellow color	16	27863229	1.00246E-07	location-specific CHE	0.021	ACTTGTCAATTATAACCATTAGTAAAATAATGACA[C/T]GTGGACACA	T
AX-115632716	Floral emergence	9	4822731	8.62531E-08	location-specific CHE	0.021	AAATTGTAGATTAATCTTTCTA[C/T]GGTTTACTCTACTGTAAATTCTAGTGTACGGGAA	T
AX-115421520	Single fruit weight	5	32741279	7.6222E-10	location-specific CHE	0.021	CCAAAGGATTCTGCAGTCT[A/G]CAGAATTATTCTTTCGAAAACATGCACAAATAATA	G
AX-115421520	Fruit volume	5	32741279	3.83782E-10	location-specific CHE	0.020	CCAAAGGATTCTGCAGTCT[A/G]CAGAATTATTCTTTCGAAAACATGCACAAATAATA	G
AX-115400167	Bitter pit grade	9	2402403	8.29252E-08	location-specific CHE	0.020	TTTGGGTGCCACTGAGTAGAGTTGAAGGAAGAATA[G/T]GAAGAAAGTTAAGACT	T
AX-115359037	Titratable acidity	10	20136814	4.66794E-09	location-specific CHE	0.020	GAAGTCCTGACCCAATTCTTGAGGTCAGATTTTGC[A/G]AAAGTCCAAATGTATTATTGATACTTTGCAACCCT	A
AX-115440105	Floral emergence	5	31580677	1.27202E-09	location-specific CHE	0.018	AAAAGATTACATGTTCAAATTAAATAAATTGTTGA[C/T]ATTATCAACTTGGTATGTTATGTTAAACTA	Heterozygote
AX-115220987	Russet freq. - eye	15	26250511	5.08065E-09	location-specific CHE	0.017	AACATCTTTTTAAATAACAGTGCATTATAAGTACA[C/T]CACG	T
AX-115245532	Ground color	9	14186556	7.97171E-08	location-specific CHE	0.017	AAAAATTTGCTCAAC[A/G]CTTTCAGGCTTTAACCATAACAGAGTGAATATCTA	Heterozygote
AX-115508423	Russet cover	17	27235275	8.7208E-10	location-specific CHE	0.017	CCATGAA[A/G]GGAATCGTTTTCCAGAAAGTGAGAGAAAATTAGGT	G
AX-115192200	Single fruit weight	2	3679888	1.62045E-08	location-specific CHE	0.012	TGAAAGGGAACCAAATCAAATGACGAAAAAATACA[A/G]AGGCACACAGAAGATAAAAATTATATATATGAAAT	G
AX-115210600	Fruit firmness	6	3080366	3.88214E-08	location-specific CHE	0.009	TGCATTATATCATTTCGTATAACTCCCCCAAACCC[A/G]CCTGTACTGTAGTTTGATCTTAATGGTTGAGGCGC	A
AX-115459753	Maximum fruit size	16	3380983	8.90471E-09	location-specific CHE	0.006	GTTCTAAAAAGGGTGTTTTTGAGGATGTTGATGAT[A/G]GGGTAGTGGATGCTGGATTGAGGCAGGAAAGCGTC	A
AX-115219752	Floral emergence	9	7577508	2.422E-08	location-specific CHE	0.006	CTACCACTAAAGAGTCCTTCATAACTTTAGAGGTG[A/C]GACTCTTGATCTGC	Heterozygote
AX-115398287	Flowering intensity	7	8114233	8.11334E-08	location-specific CHE	0.004	GTTAAAAA[A/G]GTGATAAGCAAGATGGTGAATGGGAAAGTGC	G

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115253823	Harvest date	14	28673903	1.07016E-08	location-specific CHE	0.003	A[G/T]TTTTGAGTTAATTGTTTGGGCGGGAAAACTGTAGG	T
AX-115256041	Bitter pit freq.	13	3358666	5.07365E-09	location-specific CHE	0.002	GATTTGAAGTTCTTTGTTCTTTCTCTTAAGCTCCA[C/T]GACTTCCACCTCTAATTGATTCACAGC	C
AX-115469898	Ground color	14	19788647	1.96457E-08	location-specific CHE	0.000	TGGGTGGAGGAAGACAAGTCATCATCAGAGATGGC[A/G]CAGTTGTGCAAGA	Heterozygote
AX-115324071	Flowering intensity	7	7605109	2.97438E-08	location-specific CHE	0.000	TT[C/T]GTTTAAACTTTAAGCAGCCATAATGATTTTCTTCT	T
AX-115279459	Red over color	9	33799120	8.14423E-82	location-specific ESP	0.613	CAGCTGGACCCGCATTTCATTCTTGAGTCGGAAAG[A/G]CTCTCAATTGAAAAACATTTCATATGAATTGAAAC	A
AX-115366114	Harvest date	3	30681581	7.68241E-40	location-specific ESP	0.410	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115200955	Number of fruits	16	4105954	5.89526E-09	location-specific ESP	0.227	AATCGTGAAGCTTATGACGGTCGTCATTTTTCTGA[C/T]TTATGAATAAATTACCTGTCAGCAACGGCCTGATA	T
AX-115435991	Trunk increment	6	34315336	1.79335E-08	location-specific ESP	0.141	TTCTCCCTCCAATTCATGAACATTGTTGTAATTTT[G/T]GGTTGAATCCATCGCC	T
AX-115350952	Harvest date	16	8932940	6.39473E-08	location-specific ESP	0.116	TCTGTTCAAACTCTTCCTCGACGCGGTGATAGACG[A/G]GATGAAGTCTACACTTGTGCACATCAA	G
AX-115267846	Trunk increment	13	8355795	3.17591E-09	location-specific ESP	0.107	ACTCGACGCGAACCTCGTGCGCCGACTCGCCCTGC[G/T]CCACAATCTCCGTCTTCTTCGACTTCAAGCACACC	G
AX-115465870	Harvest date	11	29305015	5.90036E-11	location-specific ESP	0.077	TTTCTACCACAAAGAA[A/C]TTAAATCGAATCGAGCTTTGTCGAAGCTAATTTGA	A
AX-115317393	Russet cover	9	12381129	2.96633E-08	location-specific ESP	0.076	C[A/C]ACTTACATATAGTTCCGTTAACTAGTGCACTCCTA	A
AX-115653996	Russet freq. - eye	12	23013281	2.21058E-10	location-specific ESP	0.074	TTTGAAAC[A/C]GACCAGAGTTTTGAAGTGATCCACAAGTCCTCCCG	C
AX-115350543	Trunk diameter	3	28264461	4.9953E-08	location-specific ESP	0.072	TTT[A/G]TGTTATATGGAAGTGGACATGTACATAAAAGACAC	A
AX-115406429	End of flowering	7	2735255	9.09215E-09	location-specific ESP	0.068	TGACAATATGAGAATGCTATGTTCAAAAAATTTAG[C/T]AACAGGTTTCAGATTCAGATGAATAAAAATTTCGA	C
AX-105193183	Russet freq. - eye	2	29512625	3.33788E-12	location-specific ESP	0.066	ACACCCATTCAAGCTGGAAATTTGGAGAAGTAGCG[C/T]ATACCT	C
AX-115605665	Trunk diameter	6	22538503	1.34142E-08	location-specific ESP	0.064	AACATCTTCATTTTATGATCCTCTTCATCAGTTTC[C/T]A	C
AX-115426282	End of flowering	9	9761358	5.53595E-13	location-specific ESP	0.063	TGTATTTGCTTCAAAAATCTTGGTGTCGACCTTGA[A/G]GCAAAA	A
AX-115405485	Number of fruits	9	20188498	1.5627E-09	location-specific ESP	0.059	TTGGTCAGGATCAAAGAAATCAACTGTAATGCTCC[G/T]GTAAATTGAGTCACCAAGACTTGCCCTCCCATTCT	G
AX-115422195	Russet freq. - overall	10	39892373	2.57664E-09	location-specific ESP	0.056	CTTTGCC[C/T]GTTGGCGCAAGGCAGGACTTAAATTGCTCGAGCAT	C
AX-115190668	Bitter pit grade	16	3891590	5.00671E-13	location-specific ESP	0.055	GCAAGTTTCCATTAAAATGGCACAAATCACTCGCA[A/C]GGTACTTTATTTATGGCACTTTCCAAACTCGTTGG	A
AX-115392559	Russet freq. - cheek	17	3472654	1.67124E-12	location-specific ESP	0.055	GCTCAATGTGGAGATTTGGATTTCATTGAACTTGT[G/T]TATTGCTGGCGACCT	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115310232	Russet freq. - eye	17	5278335	1.64206E-14	location-specific ESP	0.055	TCTTAA[C/T]CAACTAAATGACATGCTCATTGCTAATGTTC	T
AX-115460774	Floral emergence	4	29028596	1.22823E-08	location-specific ESP	0.054	ATGGCAATGCACATATGATTAAT[A/G]TTACTGTCATAAGATTTACCTTTTGATCAGATCTA	A
AX-115277738	Full flowering	7	12526666	3.74904E-10	location-specific ESP	0.052	AATGTGTCA[C/T]GAAAATAACAGACATTCTTGAGATGAAAGCAAAGT	C
AX-115519258	Russet freq. - eye	1	28377604	4.40075E-10	location-specific ESP	0.051	AAAATTTTCACATGATTCACATCAATTCAATCACA[A/G]TGAACGGAAAGGTTACCCGAAGC	A
AX-115578714	Full flowering	1	12852725	1.41476E-07	location-specific ESP	0.051	AATTGCAAGCATGGAATAACTATAGTGTATAAAAA[A/G]ATCACAAGAAAAAGAAAAAAGAAAGCTAGGTTAAA	A
AX-115366114	Russet freq. - overall	3	30681581	4.05545E-24	location-specific ESP	0.049	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115340041	Russet freq. - overall	6	25003502	1.70799E-08	location-specific ESP	0.045	ATCTTAGGACCGATAAGGGAAAATTTCCAACAAAA[A/G]A	G
AX-115420425	Trunk diameter	1	12348192	7.61444E-08	location-specific ESP	0.039	ATGTTCTCCA[A/G]TATTCTATCACAGGTTCTTTGGTCATTACATC	G
AX-115211784	Trunk diameter	6	34467548	1.10106E-08	location-specific ESP	0.036	CTGAAAGCAGGCAAAGGTAAATTAACATACAAGTC[C/T]GATCACATATACTTAATCAGGAAACTCAAAAAATG	T
AX-115606670	Floral emergence	5	4250031	6.20474E-15	location-specific ESP	0.036	TGGGGACGAGT[A/C]TAAACACTATCCACTATGTTTGTAATTGCA	C
AX-115464325	Weight of fruits	9	6684500	1.66478E-11	location-specific ESP	0.034	AGGGCAATTTTGTCCTCTAATGTTTTGAGCGAATT[A/G]GGGGCAGGTTTGGCTTCGT	G
AX-115488925	Floral emergence	14	1877265	1.68091E-09	location-specific ESP	0.033	GTCTCAATAATGTGTACATGAAGAAACCAATGTC[C/T]GGCTTGCAGATTTATGGAGATCCAGAAGCAGCAGA	C
AX-115219757	End of flowering	9	7575625	6.00271E-11	location-specific ESP	0.032	TC[A/G]TCATAACACATATGAATATGCACTCATATTCTGCA	G
AX-115220326	Russet freq. - overall	1	12028810	6.48226E-12	location-specific ESP	0.028	TTTTAAGTTGCTATATGTTCTGTTGTCTTATTTCA[G/T]ATGTGGATGATAACG	G
AX-115395942	Full flowering	5	4298159	5.43154E-09	location-specific ESP	0.027	TTCTTTCTTCGGTGGTTGCGGTTT[C/T]GTTTTGTGTGGTGCTTGTTGGAGGATGATGGTGCT	T
AX-115405753	Flowering intensity	11	828296	5.0209E-08	location-specific ESP	0.026	TTTTCCTAATCCATATCATTTCATTACTCCA[A/G]CAGAATTCTCTATTGACAAGAAC	G
AX-115431042	Full flowering	2	13018540	5.58681E-08	location-specific ESP	0.026	GGACAAGTCTTGGCTGTAAAGAACACTAAAGAAGA[C/T]GTCACGGAGGAGCTAAATATTCTGCAGAAGGTAAA	C
AX-115395677	Harvest date	2	32965721	3.93565E-10	location-specific ESP	0.026	ATTTGCTAGTAAATTAATAAGGTCGACCAATTGCA[C/T]GGCGAAAGCAAACATCGGTTCCTTGACGTGCTTAT	C
AX-115231925	Flowering intensity	7	4995472	1.15661E-07	location-specific ESP	0.025	AACTGCATGTTGCAATTGTACGGATGAATCTTCAC[G/T]TATATGTAATTTTTACTAAAATATTGTCGTTAGAA	G
AX-115658243	Russet freq. - stalk	12	4194237	5.35654E-11	location-specific ESP	0.024	GCATCTTGGGTCCGCCTATTGTTTTAGTAAGGAGC[C/T]CACCAGTGGAT	C
AX-115405061	Russet freq. - stalk	13	2184895	1.46781E-07	location-specific ESP	0.023	GAAGAGAAATGTGGCAATCGTTTGGGATTGCTTTT[G/T]T	G
AX-115658243	Russet freq. - overall	12	4194237	2.53277E-09	location-specific ESP	0.019	GCATCTTGGGTCCGCCTATTGTTTTAGTAAGGAGC[C/T]CACCAGTGGAT	C

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115499827	Number of fruits	2	7997682	1.91694E-08	location-specific ESP	0.019	TA[C/T]GGTTCATAACAATTTGAATCAATTTCTGTAATTAC	T
AX-115444998	Russet freq. - overall	9	8534515	1.28163E-08	location-specific ESP	0.018	AGAAGAAGTGCAAGAGCTCGAAGATGCCTCTGCTT[A/G]CAGACTCAGGCCCCTCCCTCCAGAACTCAGGTACC	G
AX-115220326	Russet freq. - stalk	1	12028810	1.30083E-13	location-specific ESP	0.016	TTTTAAGTTGCTATATGTTCTGTTGTCTTATTTCA[G/T]ATGTGGATGATAACG	G
AX-115448235	Russet freq. - overall	7	20343306	1.00415E-07	location-specific ESP	0.016	ATTTTTACTATCTCTCTAATAAAGACGGGAGCCAC[A/G]AACAA	G
AX-115444998	Russet freq. - stalk	9	8534515	5.75937E-09	location-specific ESP	0.015	AGAAGAAGTGCAAGAGCTCGAAGATGCCTCTGCTT[A/G]CAGACTCAGGCCCCTCCCTCCAGAACTCAGGTACC	G
AX-115412093	Red over color	13	3461966	8.5955E-09	location-specific ESP	0.013	GATTTATGGTTTGATGCAGAACTAAACAATTGCAG[C/T]TTTAAGAGGATAATTAGATGTCACCTGTGCACGAG	Heterozygote
AX-115515804	Russet freq. - overall	12	2054295	1.21981E-11	location-specific ESP	0.012	TGAGAAAAAACCTTAAAATTTCAATTGGTTCACTT[C/T]CCGATTCTTTTTGTGTGT	T
AX-115191874	Red over color	17	28349586	1.07897E-08	location-specific ESP	0.012	ATGAATCGCGTTCTCCATAGGGGATAATTATTGCA[G/T]TCAAGTTTGATCGTGAGGAAGTTGTAACATTAGAG	T
AX-115272912	Russet freq. - overall	2	6306321	1.20006E-07	location-specific ESP	0.010	TTGGTGGTGGAGGAGAATTTGTTGGTGGCACAGAT[A/G]GTGCATGACCTTCAACACTTGGTGGTGGAGGAGCA	A
AX-115429980	Harvest date	6	6509938	1.09121E-07	location-specific ESP	0.009	A[A/G]TATAACTAAAACAGGGTAGCCTTCATTTGT	A
AX-115545984	Russet freq. - overall	15	36931867	6.34819E-08	location-specific ESP	0.007	GACGAGTAGCA[C/T]ACTACTTAAATAAGTCAAAAACAAAACACACAACA	C
AX-115333828	Bitter pit grade	8	4081206	1.27221E-07	location-specific ESP	0.005	AAGAAATAAAATTCTG[C/T]AAACTACAGATTTACCTCTCTCAAATTGATTGTAG	T
AX-115431939	Single fruit weight	12	26039343	2.71667E-08	location-specific ESP	0.005	TACGGATATATACAAATTGATCGT[A/G]AATATGCGTTTAAATAGCCAACGGTGAGGGTTCTA	G
AX-115278179	Single fruit weight	4	28092945	5.69469E-10	location-specific ESP	0.004	TCCACATAATTTACTCTCTTCGAGGATTCCAAATC[C/T]A	C
AX-115220019	Russet freq. - overall	12	12269553	2.87893E-16	location-specific ESP	0.001	GGAGAAGGTTCGAATCTATTATAATAAATTA[A/G]GAGAGAGAGCATCCAACTCA	A
AX-115271052	Flowering intensity	5	4849992	2.02911E-08	location-specific ESP	0.001	TTTATTCGGCGACTAGAATTCTGT[C/T]CGGTTTTAATTATTCTTGTCCGGCTTTTGTGATTG	Heterozygote
AX-115366114	Ground color	3	30681581	1.11315E-12	location-specific ESP	0.000	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	G
AX-115302680	Russet freq. - overall	15	52325798	2.1881E-09	location-specific ESP	0.000	TAATCGTTGAGTTTATCCAACCTCACAAATTCAC[C/T]TTTTTCCAAAAG	Heterozygote
AX-115395944	Floral emergence	5	4299711	5.03158E-09	location-specific ESP	0.000	CTTGAAAACTTCAAAGGCACAACTACATAGT[A/C]TAAATCCATC	C
AX-115415461	Number of fruits	9	7112818	1.12672E-09	location-specific ESP	0.000	TAGCATTACCCATATGAATAG[A/G]AAGAATGACTAAATTGGCTCAGTCCTTCGTTGCTT	A
AX-115519732	Trunk increment	8	6653442	3.62754E-08	location-specific ESP	0.000	AGTTGTCGAATTCCCAAGCTGCTTTATAAATCCAA[A/G]A	Heterozygote
AX-105207792	Ground color	9	35639881	1.17753E-07	location-specific ESP	0.000	CCCTTTTTTCACAGAACAAGTTAGGAATCTGGTCG[A/G]GTACCGTCACCGTAGAGTTGAAGACCACC	Heterozygote

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115279459	Red over color	9	33799120	1.58899E-92	location-specific FRA	0.613	CAGCTGGACCCGCATTTCATTCTTGAGTCGGAAAG[A/G]CTCTCAATTGAAAAACATTTCATATGAATTGAAAC	A
AX-115366114	Harvest date	3	30681581	5.92683E-22	location-specific FRA	0.410	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115294359	Harvest date	10	33341151	6.23512E-08	location-specific FRA	0.171	GTTAAATTTCTATTTAAGATGTATTTGACTT[A/G]ATATATCACATTAGCGAAA	A
AX-115331889	Weight of fruits	7	10150663	8.6429E-08	location-specific FRA	0.134	CATGCATAAGTTCAGATAACGGAACAAT[A/C]GTGAAAGTTCTGCCTTTAGCATATGGAAACTGG	A
AX-115492821	Harvest date	16	8925608	2.243E-12	location-specific FRA	0.132	TTGGTTGTCATGGCTGCCATTATTCATCATTTTCG[C/T]GGAAGAAGACAAGGGTTTTGTGACAGA	C
AX-115476635	Floral emergence	11	11277966	1.00316E-07	location-specific FRA	0.107	TACAGTATTCTCATTGAAGGAATGTGCATAGCTGG[A/C]AAAATTGAAGATGCCAGAAAACTCTTT	C
AX-115385023	Trunk diameter	9	9180796	7.44125E-08	location-specific FRA	0.093	ATTTTGTCAAGCACTTCTAAAAGAACCCATCTGCA[C/T]GCACAA	T
AX-115429214	Ground color	10	2291372	1.27625E-10	location-specific FRA	0.091	TTGATATGATATTCATATGAT[A/C]AAGTAAGATTGTTCATATTACAAAGTATTCAATCA	C
AX-115487572	Flowering intensity	17	4118713	1.32899E-07	location-specific FRA	0.084	GCAGTTTCCTTAATCAATTGGCTTTCTCAAAGATC[A/G]TGAAGAAATATGACAAGGTTTGATCGCTTTGTTTC	G
AX-115316099	Weight of fruits	7	28721665	2.44114E-10	location-specific FRA	0.080	CAATGAGGTCGATTTTCTCCCTTATGAGGTATATT[C/T]TCGAACTTCACATTCTCTGAGGTTT	T
AX-115659049	Floral emergence	9	1311208	1.90979E-09	location-specific FRA	0.079	AAGCCATCTGGTCTGAGTATCCTATCCATTTCAAT[C/T]AATAGATCTTCTGCATTGCAA	T
AX-115440975	Russet cover	9	11929448	1.06195E-09	location-specific FRA	0.072	TCCAAATCATTCTGCCTCTAATGATGTTTCTGTGG[C/T]ATCAGCTACTAT	T
AX-115357001	Single fruit weight	1	18679105	4.97625E-09	location-specific FRA	0.060	ACTT[G/T]GCTGAGATATTTGGACTTACTCAGATCTGA	G
AX-115251529	Ground color	8	14484388	4.29878E-11	location-specific FRA	0.060	AAACTGTAAGTGTTCTACTTTTGGTATGATATTAT[A/G]AGTGCATATTTGTTTGTCACATTTCAATACATTTC	G
AX-115252388	Floral emergence	16	15218544	4.50856E-10	location-specific FRA	0.059	CAATAGAGAACAACGTAATGCAGATTCACGAGCAC[A/G]TC	A
AX-115628003	Single fruit weight	14	2902993	2.66708E-09	location-specific FRA	0.056	AATAATATAATAATCAAGATCTAATTATTAAAGGG[A/G]TGTG	G
AX-115182443	Bitter pit grade	16	3409834	4.5727E-10	location-specific FRA	0.053	AAGAGTGATTCTTATCCCATAAATTTGCAATACTA[C/T]TAAGCTTTTACATTAACAAGTAGTTGACTAAGACA	T
AX-115522199	Russet freq. - overall	5	34611609	1.04921E-07	location-specific FRA	0.050	CAGGAGCAGTGTGTGCA[G/T]CAACTCTCGCCAGAAATGGAGTGTCTGTAACTTTG	T
AX-115218171	Bitter pit grade	11	6028089	9.02679E-11	location-specific FRA	0.049	TTTTGTAGGGTGGATATGTCGATTTATGCC[A/G]TGGTGTTTGGAGAATTCAGGGGTCTCTAGCTGTAT	A
AX-115244803	Ground color	16	9245659	5.62044E-09	location-specific FRA	0.049	ACGGGGCCGGTTGCCGGGACTCAGTATCTGTCGAC[C/T]GCTAGACTGGCTTTTTGCAGCGATCGTCCTCTGAC	T
AX-115185265	Ground color	15	5967264	1.21038E-07	location-specific FRA	0.048	ACT[A/G]TTAGCATCCATATAGATCATACAAATTGACCGTTT	A
AX-115200006	Harvest date	5	47253922	1.78809E-07	location-specific FRA	0.048	ATTCCAAACCTTTTGGGATAAACCCAATTTCAATA[C/T]CTTCCCACCACATTCATCACTTCCTTGCTTGGAAA	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115434303	Harvest date	2	13514493	2.90236E-10	location-specific FRA	0.048	GGTGAAGAAAGTTAGAGAGCGAGTTTTTGCTGTGA[A/G]GGGG	A
AX-115515978	Single fruit weight	8	1233262	7.39986E-08	location-specific FRA	0.046	TAATCGTACAGATTTTACGCGATCCAATTGCAGAG[A/C]GACTCTCTGTGTGTCGCGAAACGGCGATCGCAATC	A
AX-115609732	Floral emergence	8	14015091	9.77098E-08	location-specific FRA	0.042	ACTTATATTATTTAAATTTACTATGAACAGCGACA[A/G]CCATATCCTAATTGATATCATCCTAGTCATGAACT	A
AX-115260900	Russet freq. - overall	10	7885378	2.71141E-09	location-specific FRA	0.040	AAATAAATAATTACCGTATTCCACTGTTTGCATGA[C/T]AGATATTTAGTGCCAGACCTATGGAATACTCACTC	C
AX-115274446	Ground color	8	7166705	1.61781E-08	location-specific FRA	0.037	TATGTTGTCGGCTACCTCTCTGATAATTTCCCAAT[C/T]GAAGTCAAACTAATTTTATCGTATCTTTTGTTTAT	C
AX-115572381	Bitter pit grade	16	8211656	2.70494E-08	location-specific FRA	0.037	AATTAGTGGTAATCTATAGTATTCTGGTTCAATTT[C/T]AG	T
AX-115575219	Bitter pit freq.	16	7634599	1.48058E-09	location-specific FRA	0.037	ACA[C/T]AATACTTTATAACGTTAGCCTGATAATTAACCTTA	C
AX-115218171	Bitter pit freq.	11	6028089	3.52883E-13	location-specific FRA	0.034	TTTTGTAGGGTGGATATGTCGATTTATGCC[A/G]TGGTGTTTGGAGAATTCAGGGGTCTCTAGCTGTAT	A
AX-115259815	Russet cover	2	19901873	8.40102E-08	location-specific FRA	0.032	CCTTCT[G/T]GATTTATATACATAGAAGAAGAAAATATAATAATA	Heterozygote
AX-115232910	Bitter pit grade	12	23632017	1.37403E-07	location-specific FRA	0.031	A[G/T]AGAGAGAAATGGACGACATGCCAAATATGTCTCCT	G
AX-115367875	Trunk diameter	13	25396311	1.35172E-07	location-specific FRA	0.014	GAATGAAGCCTTGAATTACCTATCCCACTGACCAT[A/G]TATCTCATAAAGAAGCTCGAATGTTGTTGAGTCAA	A
AX-115189018	Flowering intensity	14	25160350	3.25269E-10	location-specific FRA	0.013	ATAAGATCAAATCATTGATCTCCTAGACTACTAAG[C/T]GAT	C
AX-115395987	Bitter pit freq.	8	2983123	3.56866E-08	location-specific FRA	0.010	AGACGGTTTGCTTT[A/G]CAACAGAAAATTGTAAGGTTATTATCGTCTTTTC	G
AX-115414774	Red over color	14	25356468	1.39788E-08	location-specific FRA	0.010	TCTTCCATG[C/T]CTTCTCGAAGAGAAGTAATCCTTCAAAATAATATG	T
AX-115484495	Russet cover	4	29711863	9.97358E-08	location-specific FRA	0.001	CTGAAAAGAAAACTGCCTCTTCTGCTGGAGCATCC[A/G]GTTCTGGAAAGAGGATCCCAAGATCATTGAAGGGA	G
AX-115375158	Weight of fruits	12	5540975	1.62434E-09	location-specific FRA	0.000	GCTGTCGAGCAGTAAGCAATCGTATCAGTGGAGCT[A/G]AGAGACGAGTGAAACAGTGAAAGAGAAGGACAGCG	A
AX-105213720	Red over color	9	33801013	1.11675E-85	location-specific ITA	0.615	TGACTATTGTGGAGCAGGAGGCGAACCATAA[C/T]GGGCCGGTTTTCTTGGACCGGTTCACGGAGGCATT	T
AX-115366114	Harvest date	3	30681581	1.05131E-10	location-specific ITA	0.410	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115183752	Trunk increment	1	29560923	1.9473E-08	location-specific ITA	0.230	AGAGAGTTTCGTACTATAAATCGGTGAATAATCAG[A/C]TGGGAAACCCTGGAAGGTATAGAAATATTTTGGAC	A
AX-115366114	Soluble solids content	3	30681581	5.83747E-13	location-specific ITA	0.173	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	A
AX-115213187	Harvest date	10	38906935	1.86344E-10	location-specific ITA	0.165	ACTAACGGCTATAATATTATTTATAGATTGCATTG[A/G]TCTTGAAACATTGCATT	A
AX-115359775	Weight of fruits	12	24539575	7.71923E-10	location-specific ITA	0.151	CGGCTCAATTCAATTCTTGTCAAAAG[C/T]TGATTTTTTTGGTCAACAAACACAAATTTTTTTTT	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115473679	Flowering intensity	14	4229082	1.15414E-08	location-specific ITA	0.143	TTCAATGACAAAAAT[A/G]AGAAATTTGCGTCCGAAAACGAATTGAGAGGCGCA	G
AX-115454577	Titratable acidity	8	11673886	1.75892E-08	location-specific ITA	0.139	CAAAAATCAAATACAATTACTCAGTAAAATCATTG[A/G]A	G
AX-105196472	Harvest date	16	8953044	2.37313E-13	location-specific ITA	0.136	TAATTTATGTGTAAATCTCCGTTTCATGATAGTCT[C/T]GTATAGAGGTTTAGAACACTCGATGTAAAGCATCT	T
AX-115600808	Titratable acidity	16	3161268	1.91244E-16	location-specific ITA	0.126	CAAAGTAACCTCGGTTCTAATCCTCTCGTAATTCC[A/C]GTTATTC	A
AX-115546273	Harvest date	3	30726252	1.39492E-08	location-specific ITA	0.121	TCTTTTATTCTAAGATCACAGTGCTGGAAGTGCCA[A/G]CTATATTTCCAAAAATACTTCTAC	A
AX-115476635	Floral emergence	11	11277966	1.39874E-09	location-specific ITA	0.107	TACAGTATTCTCATTGAAGGAATGTGCATAGCTGG[A/C]AAAATTGAAGATGCCAGAAAACTCTTT	C
AX-115495298	Soluble solids content	4	29188324	7.63158E-08	location-specific ITA	0.106	AGCTTGACCTATGGAATCAGTTGA[A/C]ATATGTACTTCCTGCAGGTGTAAGGTCATGATCCA	C
AX-115452498	Titratable acidity	17	3760667	3.90017E-08	location-specific ITA	0.099	TTCAGTCGTTTCCGAGCTCCAAGTCCTCAGCTCCG[A/G]CAATCCCATCCTTTCTCTCACAGATTTTTTGGGGA	G
AX-115248540	Harvest date	11	13523287	1.61052E-09	location-specific ITA	0.088	TTTTTCAAATGGATGTCAAAACTGCATT[A/C]CTTAATGGATATCTTAATGAGAAGGTCTATGTGAA	C
AX-115438285	Water core grade	12	1055508	5.48842E-08	location-specific ITA	0.085	TTATAC[C/T]ATGCTAAATTAAAAAGAAGGTATTACTGACACACC	C
AX-115320741	Russet freq. - eye	2	29976071	3.51347E-19	location-specific ITA	0.084	AAAAATATATCTATGCTATGTTTCAAATGATTGT[A/G]AA	A
AX-115562961	Bitter pit freq.	16	3471935	2.43617E-15	location-specific ITA	0.082	TATAGTTT[C/T]GTCCAACCCCAAATAATAATGGCAAATAGGAGTTT	T
AX-115635358	Flowering intensity	3	31449413	2.62352E-11	location-specific ITA	0.082	GCTAAAACACTACAAAATTGTCAAGAAT[A/G]CTACACACTCCAAAAAGACTTGCCAGACTCACATC	A
AX-115407736	Russet cover	12	22271380	1.01931E-11	location-specific ITA	0.080	TTGGGGGTTAAAGCTGCA[C/T]TATCTCTACTCAAATTCTACAGAAGTAAGTTTCT	T
AX-115511827	Fruit firmness	16	9042108	9.42022E-09	location-specific ITA	0.071	TCTCACATTGAAACGTACCTATGCTAAATTAAAAT[A/G]TATTCATACCAAATAAAAATGTATGCTTACTAAAT	A
AX-115534786	Russet freq. - cheek	14	25600768	1.62633E-08	location-specific ITA	0.069	CGCGGAAACACGATGGAGGACGCAATTGATGCTGA[C/T]GGAGAAAAGGG	C
AX-115271245	Floral emergence	2	13180576	1.45448E-10	location-specific ITA	0.068	A[A/G]GTTTTGGGTTTATTAAATTTAATTAAGCATA	A
AX-115227682	Flowering intensity	13	4967068	1.6217E-08	location-specific ITA	0.066	ACCCAAAGCTCATTTTTAGGGCAAATTTAGACGGG[C/T]TG	C
AX-115407371	Russet freq. - overall	10	40116376	1.07634E-07	location-specific ITA	0.066	A[C/T]GGTTGTAATCAACAACTGTTTGATGTAATTTTTAG	T
AX-115320550	Fruit firmness	10	32825957	1.11616E-09	location-specific ITA	0.065	CATTGTACGCATACACATCAGGCTCATACCCTGCT[C/T]CTTGCATCTGCTCAAAAATTTCTTCCGCTTTCTCA	C
AX-115592097	Red over color	1	26173468	2.57019E-12	location-specific ITA	0.064	ATATCCTAACGTCC[C/T]ATATCATGAGGTGAGTCCCAAACTCCCAATACTAA	C
AX-115652532	Single fruit weight	13	6291369	1.42921E-09	location-specific ITA	0.060	GTATA[C/T]CACTAAACGGTAATATAAGTGACATAAATTATTCT	T

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115504430	Russet freq. - overall	10	11679436	1.55446E-07	location-specific ITA	0.059	TCATA[A/G]TATAAATGTTATGTGGGCCCTAGCATAAAATCAAG	G
AX-115514879	Titratable acidity	9	31771721	1.45616E-07	location-specific ITA	0.058	TTTAGGCTTTTCGGGTTATCAACC[A/G]TGAATGACTTCTTAACTAACTGACTAACTAATAAA	G
AX-115551270	Floral emergence	12	6310447	3.32907E-09	location-specific ITA	0.053	TGATTCGTCTCCACCCAATATTTTCTATGACTCTA[C/T]ATCTTCAATCAAAGGTTTGAATTCTACG	C
AX-115249524	Harvest date	1	23157595	1.25635E-08	location-specific ITA	0.052	GTGTAAATTCCGACTAACTTCCAAAAGGCA[A/G]TGCAATTCTCCGAAACTTTCTCAGAGTATTGCTGG	A
AX-115434303	Harvest date	2	13514493	6.79052E-08	location-specific ITA	0.048	GGTGAAGAAAGTTAGAGAGCGAGTTTTTGCTGTGA[A/G]GGGG	A
AX-115407371	Russet freq. - stalk	10	40116376	2.00936E-10	location-specific ITA	0.048	A[C/T]GGTTGTAATCAACAACTGTTTGATGTAATTTTTAG	T
AX-115649770	Harvest date	15	15712727	9.61998E-08	location-specific ITA	0.048	ACAAAGGGCCAAT[C/T]GTGCTCTTTGGCAAATTACATTTTGAGTTTGGCTG	C
AX-115263051	Floral emergence	5	4539872	8.04848E-08	location-specific ITA	0.045	ACTCCACTCTGATTCGGAAATTTTGGTTCAGACTC[G/T]CTTGGAACCTCACAAGATCTCTGTTGATGATGATC	G
AX-115653996	Russet freq. - cheek	12	23013281	1.51296E-08	location-specific ITA	0.044	TTTGAAAC[A/C]GACCAGAGTTTTGAAGTGATCCACAAGTCCTCCCG	C
AX-115378078	Single fruit weight	6	35186920	7.96967E-09	location-specific ITA	0.044	TGTATCACCTCCTACCA[A/G]CAGCTTACACTAAATATTTAATTGCATAGAACGAT	G
AX-115473071	Harvest date	3	30224175	1.28161E-11	location-specific ITA	0.043	AAAATGTTACGCATA[A/G]GCACATAAGACTAAAAATTTAGCCCATTGTACTAC	G
AX-115250222	Trunk increment	12	10812778	2.09279E-09	location-specific ITA	0.043	GTACTTTTACTTTACGCCCAAGACGT[C/T]GCTGGAAAGAATCACCACCATTCACCTTTGTCCTG	Heterozygote
AX-115464400	Single fruit weight	11	14369160	1.45815E-07	location-specific ITA	0.042	ACTAACAAGTGAGCCTTAGTTAACTTCCATAGAGC[G/T]CACAAATGGCCCTCTA	T
AX-115397938	Russet freq. - stalk	3	29393034	6.97544E-08	location-specific ITA	0.040	TTTTGCCATTATAGCACAATCATCGGTTGTTCC[C/T]AACTTTTTTT	C
AX-115439894	Water core freq.	2	391358	1.85835E-09	location-specific ITA	0.034	CGTCCAAA[A/G]CACGGGCAGCAGTCACAGCTTATCATCTTCGGGCG	A
AX-115275407	Single fruit weight	2	17236009	1.14195E-07	location-specific ITA	0.033	GGCTCATTTGGTTTTTGGGGGGCTAAGATTGGAAA[A/G]TTTTTTAGAGCTTTATTTTTCATCT	G
AX-115365302	Russet cover	8	6752807	1.02501E-08	location-specific ITA	0.028	TGCTACTTTTCAGGATAAAGAGAG[C/T]GTTGGAAATCTAGTTAAACTGATAGACAAGAGCAA	Heterozygote
AX-105207382	Trunk increment	9	9562482	1.49663E-10	location-specific ITA	0.027	AGCTTTTCAAGGGAGCCAGGGGAGACAACGTGATC[A/G]TCGTTTGAAAGCTCATTTGAGTTTTGGCTGCAAAA	G
AX-115224142	Flowering intensity	5	1256063	2.537E-08	location-specific ITA	0.027	GAGGGATGTAAGTTTTTTTATTTTTGAAATACTTT[A/C]GCTTGAAACCCACTGGTTGGTTAGTGAGAAACAAG	C
AX-115350787	Floral emergence	6	31638694	1.03236E-07	location-specific ITA	0.024	ACGTGCTGCAGTTCTTGAAGTTGGACGAAATTGTG[C/T]TGGAGAGAAAGTTCTGGGTGAGGTCGATGTCCCAT	T
AX-115258349	Russet cover	15	9251630	1.36109E-07	location-specific ITA	0.023	ATAAGCCCAGAATGGGTATTT[C/T]CAATTTTTATGAATTAAATGGGAACAAGCTTAATG	C
AX-115511599	Harvest date	3	4905406	1.13378E-09	location-specific ITA	0.023	ATTATGATATTACGAGAAATGAATATTAAAGCTGA[A/G]TAGACTAACAAAAATACAAAAA	A

SNP	Trait	Chromo-some	Position	p-value	GWAS	R^2	Sequence	Allele increasing phenotypic value
AX-115407340	Water core grade	13	5278675	1.99444E-09	location-specific ITA	0.022	TTTGAAGAGAAGAAGACCATCCCATTGGTAGTCAT[G/T]GTTTGATGATTAATCAAGCTTTTGCCACTAATCTT	G
AX-115251936	Trunk diameter	14	32332562	9.11097E-08	location-specific ITA	0.021	CTATGTTTATAACAGCATATAAAGTGTGAAACAGT[A/G]GATGCCATGCCGCAGCTTGAGAAAAGGAGATGGTA	Heterozygote
AX-115512336	Harvest date	15	32664847	1.73117E-10	location-specific ITA	0.019	GAAGCTCTTATCATGTGAATATTAACTAGAAAACA[A/C]CAGGATGAAATACGTAATTCCTTTTCTAGTCATGT	C
AX-115407340	Water core freq.	13	5278675	1.21247E-07	location-specific ITA	0.019	TTTGAAGAGAAGAAGACCATCCCATTGGTAGTCAT[G/T]GTTTGATGATTAATCAAGCTTTTGCCACTAATCTT	G
AX-115241088	Ground color	5	37952572	6.85942E-08	location-specific ITA	0.019	T[A/G]CCCTTGTGTATTTCTGGTAATTTAGTAATTACAAA	G
AX-115264145	Trunk diameter	17	7032273	1.54801E-07	location-specific ITA	0.019	AAATGCGAACACATTTCTACATTCGTGAGTGTAAT[A/G]TTACAAA	G
AX-115452523	Red over color	17	3769755	4.6613E-08	location-specific ITA	0.018	GTGAGTACTGA[A/G]AAAATTCATGTGCATTATTATTTCCCACAATTTAC	A
AX-115211543	Number of fruits	2	5257009	3.02505E-08	location-specific ITA	0.018	TAAGATGAAAAAGAGTTACATGGTGCATGGAAGAA[A/C]TGATATATTTTTGGGTTAGAACTCAGAATATACC	C
AX-115216035	Russet cover	15	14787265	5.58708E-08	location-specific ITA	0.017	TTTACAAACCATTTGAGTTTTGATTAAACAAAACC[A/G]TACCTCCCATTTTTTCTCCGAGTTCTAAA	A
AX-115405539	Floral emergence	10	22811106	4.06527E-08	location-specific ITA	0.016	ATATTTACGACTCTTATGGCCATGCATTAGAT[A/G]CAGAGGGTTTGATTTGCAT	A
AX-115267858	Red over color	4	19128105	7.46132E-11	location-specific ITA	0.014	CAAGTGC[A/C]AGGAAGGCTATTGTGGTGTGCAGAGTGATTGCAGG	C
AX-115279896	Red over color	17	28472824	1.85041E-09	location-specific ITA	0.007	CCCTAACATAAAGTTTTCTGTCCTCTTTTTGCACA[A/G]A	A
AX-115206311	Titratable acidity	16	5459760	2.32773E-08	location-specific ITA	0.002	CTTTGCTTTCTGTCCAGCATAAGAATATCAAACAC[A/G]GTTGAACTGTTTAA	A
AX-115256041	Bitter pit freq.	13	3358666	7.47286E-08	location-specific ITA	0.002	GATTTGAAGTTCTTTGTTCTTTCTCTTAAGCTCCA[C/T]GACTTCCACCTCTAATTGATTCACAGC	C
AX-115256226	Titratable acidity	13	2000236	6.32178E-10	location-specific ITA	0.002	GTGTTAAAAGCTTTTTTCTTCGGAGAGGATGCTCC[A/C]CTATCCTTCATCACAGACAGTTTAGAGATTTTAGC	A
AX-115579040	Trunk increment	7	21018992	3.93647E-08	location-specific ITA	0.001	CATGCTAAAAAAACTGCAGGTCTTTCTTTCTTAAA[A/G]CTCCCTCTAACAGATACATTTCCCTCAATCA	Heterozygote
AX-115366114	Ground color	3	30681581	1.20829E-07	location-specific ITA	0.000	ATAAGCTTTGCTAAGGTTTGTGAAGTTTACAC[A/G]AACTC	G
AX-115366123	Ground color	3	30698713	1.20829E-07	location-specific ITA	0.000	GAATA[C/T]GAATCAGAACATTGTATGTCATAATTTAATTGGAG	C

Supplementary table 4

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Acidity	Titratable acidity	10	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Acidity	Titratable acidity	13	bottom	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Acidity	Titratable acidity	15	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Acidity	Titratable acidity	16	top	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Acidity	Malic acid content	16	top	QTL	SSRs + other	1014	progeny	Khan et al. (2013)	Tree Genetics & Genomes
Acidity	Titratable acidity	8	center	GWAS	SNPs	8000	progeny	Kumar et al. (2013)	BMC Genomics
Acidity	Titratable acidity	16	top	GWAS	SNPs	8000	progeny	Kumar et al. (2014)	Plant Genetic Resources: Characterization and Utilization
Acidity	Titratable acidity	8	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Acidity	Titratable acidity	15	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Acidity	Titratable acidity	16	top	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Acidity	Titratable acidity	8	center	QTL	SNPs	1014	accessions	Kunihisa et al. (2016)	Breeding Science
Acidity	Titratable acidity	3	bottom	GWAS	SNPs	12608	accessions	Lee et al. (2017)	Plant Breeding
Acidity	Titratable acid	1	center	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Acidity	Titratable acid	1	bottom	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Acidity	Titratable acid	6	center	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Acidity	Titratable acid	7	center	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Acidity	Malic acid	8	center	QTL	SSRs + SNPs	601	progeny	Ma et al. (2016)	Tree Genetics & Genomes
Acidity	Malic acid	16	top	QTL	SSRs + SNPs	601	progeny	Ma et al. (2016)	Tree Genetics & Genomes
Acidity	Fruit pH	16	top	QTL	SSRs + other	290	progeny	Maliepaard et al. (1998)	Theoretical and Applied Genetics
Acidity	Titratable acid content converted into malic acid weight	1	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Acidity	Titratable acid content converted into malic acid weight	3	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Acidity	Titratable acid content converted into malic acid weight	8	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Acidity	Titratable acid content converted into malic acid weight	8	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Acidity	Titratable acid content converted into malic acid weight	11	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Acidity	Titratable acid content converted into malic acid weight	16	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Acidity	Total acid (HPLC)	8	center	QTL	SNPs	3441	progeny	Sun et al. (2015)	BMC Genomics
Acidity	Titratable acidity	8	top	QTL	SNPs	1344	progeny	Verma et al. (2019)	Tree Genetics & Genomes
Acidity	Titratable acidity	16	top	QTL	SNPs	1344	progeny	Verma et al. (2019)	Tree Genetics & Genomes
Acidity	Titratable acidity	16	top	QTL	SSRs	370	progeny	Xu et al. (2012)	Molecular Breeding
Acidity	Malic acid content	8	center	QTL	SSRs	299	progeny	Zhang et al. (2012)	BMC Genomics
Bitter pit	Five levels of bitter pit	16	top	QTL	SSRs	86	progeny	Buti et al. (2015)	Molecular Breeding
Bitter pit	Five levels of bitter pit	16	top	QTL	SSRs	95	progeny	Buti et al. (2018)	Molecular Breeding
Bitter pit	Bitter pit presence or absence	16	top	GWAS	SNPs	8000	progeny	Kumar et al. (2013)	BMC Genomics
Firmness	Fruit firmness at harvest	1	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	1	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	1	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Firmness	Fruit firmness at harvest	3	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	3	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	6	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	6	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	9	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	10	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	10	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	13	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	14	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	14	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	15	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	16	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	16	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Fruit firmness at harvest	16	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Firmness	Firmness (harvest)	6	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Firmness	Firmness (harvest)	10	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Firmness	Firmness (harvest)	15	bottom	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Firmness	Fruit firmness at harvest	10	top	QTL	SSRs + other	25	progeny	Costa et al. (2010)	Journal of Experimental Botany
Firmness	Four levels of fruit flesh firmness at fruit maturity	3	bottom	GWAS	SNPs	7218060	accessions	Hu et al. (2020)	The Plant Journal
Firmness	Firmness at harvest	10	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Firmness	Firmness at harvest	14	bottom	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Firmness	Fruit flesh firmness at harvest	1	bottom	QTL	SSRs + other	290	progeny	King et al. (2000)	Theoretical and Applied Genetics
Firmness	Fruit flesh firmness at harvest	10	bottom	QTL	SSRs + other	290	progeny	King et al. (2000)	Theoretical and Applied Genetics
Firmness	Fruit firmness	3	bottom	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Firmness	Fruit firmness	10	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Firmness	Fruit firmness	11	top	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Firmness	Flesh firmmness	9	top	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Firmness	Flesh firmmness	10	top	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Firmness	Fruit flesh firmness	3	bottom	GWAS	SNPs	8000	accessions	Migicovsky et al. (2016)	Plant Genome
Firmness	Mean fruit firmness	1	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Mean fruit firmness	3	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Mean fruit firmness	6	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Mean fruit firmness	11	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Mean fruit firmness	13	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Mean fruit firmness	14	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Mean fruit firmness	15	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Firmness	Fruit firmness	11	top	QTL	SNPs	3441	progeny	Sun et al. (2015)	BMC Genomics

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Flowering time	Time between budbreak and the beginning of flowering in growing degree hours	9	top	QTL	SNPs	6849	progeny	Allard et al. (2016)	Journal of Experimental Botany
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	8	center	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	9	top	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	6	top	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	6	center	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	8	top	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	12	center	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	1	top	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	12	center	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Date of flowering budbreak (50% of the flower clusters having the king flower open)	17	top	QTL	SSRs + SNPs	368	progeny	Celton et al. (2011)	New Phytologist
Flowering time	Flowering date	15	bottom	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Flowering time	Flowering date	8	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Flowering time	Flowering date	15	bottom	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Flowering time	Flowering period	4	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Flowering time	Flowering period	9	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Flowering time	Flowering period	11	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Flowering time	Flowering period	12	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Fruit size	Fruit length, diameter or size	2	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	3	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Fruit size	Fruit length, diameter or size	4	top	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	5	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	9	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	8	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	8	center	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	8	top	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	11	top	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	12	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	12	center	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	13	center	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	14	bottom	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	15	top	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit length, diameter or size	17	top	QTL	SSRs	251	progeny	Chang et al. (2014)	Scientia Horticulturae
Fruit size	Fruit diameter	7	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit diameter	8	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit height	1	bottom	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit height	4	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit height	7	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit height	8	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit height	17	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Weight	4	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Weight	7	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Weight	8	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit size (1-5 scale)	7	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit size (1-5 scale)	8	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit size (1-5 scale)	11	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit size (1-5 scale)	15	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit size (1-5 scale)	17	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Fruit size	Fruit size - single fruit weight	8	center	QTL	SSRs + SNPs	176	progeny	Devoghalaere et al. (2012)	BMC Plant Biology
Fruit size	Fruit size - single fruit weight	15	bottom	QTL	SSRs + SNPs	176	progeny	Devoghalaere et al. (2012)	BMC Plant Biology
Fruit size	Fruit size - single fruit weight	15	center	QTL	SSRs + SNPs	176	progeny	Devoghalaere et al. (2012)	BMC Plant Biology
Fruit size	Mean fruit diameter	10	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Fruit size	Mean fruit diameter	17	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Fruit size	Mean fruit fresh weight	10	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Fruit size	Fruit weight	6	center	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Fruit size	Fruit diameter	6	center	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Fruit size	Fruit diameter	17	bottom	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Fruit size	Fruit weight in grams	5	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Fruit size	Fruit weight in grams	5	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Fruit size	Fruit weight in grams	11	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Fruit size	Fruit weight in grams	12	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Fruit size	Fruit weight in grams	13	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Fruit size	Fruit weight in grams	15	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Fruit size	Fruit weight in grams	16	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Fruit size	Fruit diameter	3	bottom	QTL	SSRs	1019	progeny	Potts et al. (2014)	Plant Molecular Biology Reporter
Fruit size	Fruit diameter	5	center	QTL	SSRs	1019	progeny	Potts et al. (2014)	Plant Molecular Biology Reporter
Fruit size	Length	3	bottom	QTL	SSRs	1019	progeny	Potts et al. (2014)	Plant Molecular Biology Reporter
Fruit size	Length	5	center	QTL	SSRs	1019	progeny	Potts et al. (2014)	Plant Molecular Biology Reporter
Fruit size	Weight	3	bottom	QTL	SSRs	1019	progeny	Potts et al. (2014)	Plant Molecular Biology Reporter
Fruit size	Weight	5	center	QTL	SSRs	1019	progeny	Potts et al. (2014)	Plant Molecular Biology Reporter
Fruit size	Fruit weight	3	top	QTL	SNPs	3441	progeny	Sun et al. (2015)	BMC Genomics
Fruit size	Fruit weight	5	center	QTL	SNPs	3441	progeny	Sun et al. (2015)	BMC Genomics
Ground color	Ground color	4	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Ground color	Ground color	5	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Ground color	Ground color	7	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Ground color	Ground color	16	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Harvest time	Harvest date	2	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	3	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	3	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	3	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	5	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	5	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	5	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	6	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	8	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	8	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	9	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	10	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	11	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	14	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	15	bottom	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	15	top	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	15	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date	16	center	QTL	SNPs	264	progeny	Chagné et al. (2014)	Horticulture Research
Harvest time	Harvest date in days after picking the first genotype	16	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Harvest time	Harvest time in weeks	15	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Harvest time	Harvest time in weeks	3	bottom	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Harvest time	Harvest time in weeks	16	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Harvest time	Harvest date in days	16	center	QTL	SNPs	1014	accessions	Kunihisa et al. (2016)	Breeding Science
Harvest time	Harvest date	3	bottom	GWAS	SNPs	15000	accessions	Larsen et al. (2019)	The Plant Genome
Harvest time	Harvest season	3	bottom	GWAS	SNPs	8000	accessions	Migicovsky et al. (2016)	Plant Genome
Harvest time	Harvest time in days from January 1	1	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Harvest time	Harvest time in days from January 1	2	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	3	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	5	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	10	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	10	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	10	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	12	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	15	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	15	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Harvest time in days from January 1	16	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Harvest time	Ripening period	3	bottom	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Harvest time	Ripening period	10	bottom	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Harvest time	Ripening period	13	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Harvest time	Ripening period	15	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Harvest time	Ripening period	16	top	GWAS	SNPs	275223	accessions	Urrestarazu et al. (2017)	Frontiers in Plant Science
Over color	Presence or absence of red skin colour	9	bottom	GWAS	SNPs	52440	accessions	Amyotte et al. (2017)	PLoS ONE
Over color	Percentage overcolor	9	bottom	QTL	SNPs	8000	progeny	Chagné et al. (2016)	Tree Genetics & Genomes
Over color	Overcolor %	9	bottom	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Over color	Six levels of fruit skin color	2	center	GWAS	SNPs	7218060	accessions	Duan et al. (2017)	Nature Communications
Over color	Six levels of fruit skin color	14	top	GWAS	SNPs	7218060	accessions	Duan et al. (2017)	Nature Communications
Over color	Six levels of fruit skin color	13	center	GWAS	SNPs	7218060	accessions	Duan et al. (2017)	Nature Communications
Over color	Six levels of fruit skin color	11	bottom	GWAS	SNPs	7218060	accessions	Duan et al. (2017)	Nature Communications
Over color	Six levels of fruit skin color	10	top	GWAS	SNPs	7218060	accessions	Duan et al. (2017)	Nature Communications
Over color	Six levels of fruit skin color	9	bottom	GWAS	SNPs	7218060	accessions	Duan et al. (2017)	Nature Communications
Over color	Overcolor on zero to nine scale	9	bottom	GWAS	SNPs	8000	progeny	Kumar et al. (2014)	Plant Genetic Resources: Characterization and Utilization
Over color	“Depth” of the red coloration	9	bottom	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Over color	“Depth” of the red coloration	16	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Over color	Skin chroma (color intensity)	9	bottom	GWAS	SNPs	12608	accessions	Lee et al. (2017)	Plant Breeding
Over color	% red blush covering the surface	9	bottom	GWAS	SNPs	55000	accessions	McClure et al. (2018)	Plant Genome
Over color	Overcolor and overcolor intesity	9	bottom	GWAS	SNPs	8000	accessions	Migicovsky et al. (2016)	Plant Genome
Over color	Red versus green	9	bottom	GWAS	SNPs	47925	accessions	Migicovsky et al. (2021)	Horticulture Research
Over color	degree of skin coloration in percent (1-5 scale)	1	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Over color	degree of skin coloration in percent (1-5 scale)	8	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Over color	degree of skin coloration in percent (1-5 scale)	9	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Over color	degree of skin coloration in percent (1-5 scale)	14	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Over color	Overall color intensity	9	bottom	GWAS	SNPs	16010	accessions	Moriya et al. (2017)	Euphytica
Over color	Degree of fruit color cover	5	top	QTL	SSRs + SNPs	1130	progeny	Zheng et al. (2020)	Plant Genome
Over color	Degree of fruit color cover	8	top	QTL	SSRs + SNPs	1130	progeny	Zheng et al. (2020)	Plant Genome
Over color	Degree of fruit color cover	10	top	QTL	SSRs + SNPs	1130	progeny	Zheng et al. (2020)	Plant Genome
Over color	Degree of fruit color cover	11	center	QTL	SSRs + SNPs	1130	progeny	Zheng et al. (2020)	Plant Genome
Over color	Degree of fruit color cover	15	center	QTL	SSRs + SNPs	1130	progeny	Zheng et al. (2020)	Plant Genome
Over color	Degree of fruit color cover	9	bottom	QTL	SSRs + SNPs	1130	progeny	Zheng et al. (2020)	Plant Genome
Russet	Russeting %	1	bottom	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Russet	Russeting %	2	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Russet	Russeting %	3	top	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Russet	Russet coverage %	12	bottom	QTL	SSRs + SNPs	7556	progeny	Falginella et al. (2015)	BMC Plant Biology
Russet	Russet (presence-absence and % cover)	2	center	QTL	SSRs + SNPs	605	progeny	Lashbrooke et al. (2015)	Journal of Experimental Botany
Russet	Russet (presence-absence and % cover)	15	center	QTL	SSRs + SNPs	605	progeny	Lashbrooke et al. (2015)	Journal of Experimental Botany
Russet	area covered by russed (1-5 scale)	1	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	2	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	14	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	16	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	1	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	2	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	6	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	10	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	4	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Russet	area covered by russed (1-5 scale)	4	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Sugar	Soluble solids content	8	bottom	GWAS	SNPs	52440	accessions	Amyotte et al. (2017)	PLoS ONE
Sugar	Degrees Brix	6	bottom	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Sugar	Degrees Brix	8	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Sugar	Degrees Brix	12	center	QTL	SNPs	1289	progeny	Costa (2015)	Tree Genetics & Genomes
Sugar	Soluble solids content	2	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Sugar	Soluble solids content	10	center	QTL	SSRs + other	513	progeny	Kenis et al. (2008)	Tree Genetics & Genomes
Sugar	Degrees Brix	15	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Sugar	Degrees Brix	16	top	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Sugar	Degrees Brix	16	center	QTL	SSRs	542	progeny	Kunihisa et al. (2014)	Breeding Science
Sugar	Soluble solids content	1	bottom	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Sugar	Soluble solids content	7	top	QTL	SSRs + other	280	progeny	Liu et al. (2016)	Open Life Sciences
Sugar	Soluble solids content in degrees brix	2	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Sugar	Soluble solids content in degrees brix	14	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Sugar	Soluble solids content in degrees brix	15	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Sugar	Soluble solids content in degrees brix	15	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Sugar	Soluble solids content in degrees brix	16	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Trunk	Bottom diameter	1	bottom	QTL	SSRs	107	progeny	Segura et al. (2009)	Tree Genetics & Genomes
Trunk	Bottom diameter increment	7	center	QTL	SSRs	107	progeny	Segura et al. (2009)	Tree Genetics & Genomes
Water core	Degree of watercore (4 levels)	14	center	QTL	SNPs	1014	accessions	Kunihisa et al. (2016)	Breeding Science
Water core	degree of watercore (1-4 scale)	2	center	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Water core	degree of watercore (1-4 scale)	2	top	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Water core	degree of watercore (1-4 scale)	14	bottom	GWAS	SNPs	11786	progeny + accessions	Minamikawa et al. (2021)	Horticulture Research
Yield	number of harvested fruit	1	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	1	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	3	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	5	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	8	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	10	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	11	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	11	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	11	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of harvested fruit	13	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	1	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	2	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	3	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	5	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	10	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	11	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	mass of harvested fruit	13	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	1	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	3	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	4	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	7	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany

Grouped trait	Measured trait	Chromo-some	Segment	Analysis	Marker type	Marker number	Plant material	Publication	Journal
Yield	number of inflorescences	8	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	8	center	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	10	bottom	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany
Yield	number of inflorescences	15	top	QTL	SSRs + SNPs	168	progeny	Guitton et al. (2012)	Journal of Experimental Botany

Supplementary table 5

MUNQ	Original Code	Preferred Name
1113	X4194	Red Winter
1125	1978-136	Lady Williams
118	1948-236	Reinette Clochard
1236	X2039	O53T136
131	1947-143	Reinette du Mans
1478	Priscilla_NL	Priscilla-NL
163	1907-002	Cox's Orange Pippin
202	1947-086	Franc Roseau du Valais
2410	X4355	P7 R4A4
2422	X6398	TN R42A60
267	1972-019	Prima
29	1957-190	Borowitsky
30	1999-072	Alexander
3190	1921-089	Grimes Golden
32	1965-025	Ingrid Marie
334	1943-007	Rome Beauty
447	1974-052	Winesap
462	1976-149	Wagener
508	2006-014	McIntosh
522	1950-033	Esopus Spitzenburg
546	1953-133	Rall's Janet
548	1976-145	Granny Smith
57	1979-164	Jonathan
584	1971-046	Kidd's Orange Red
587	1974-264	York Imperial
6	1950-123	Wealthy
65	1971-054	Golden Delicious
739	1976-144	Gala
787	1963-025	Newtown Pippin
81	1947-050	Prinzen Apfel

Supplementary table 6

Trait	RF	BayesCπ	RKHS	G-BLUP	MTM. UN	G-BLUP. E.CV1	G-BLUP. E.GxE.CV1	MTM. FA.CV1	G-BLUP. E.CV2	G-BLUP. E.GxE.CV2	MTM. FA.CV2	H^2	h^2	VAR.G. and.SNP	VAR.E	VAR. GxE	VAR. Residual
Harvest date	0.78	0.78	0.78	0.78	0.78	0.69	0.7	0.55	0.86	0.86	0.84	0.99	0.88	82.77	6.88	4.70	5.64
Floral emergence	0.53	0.61	0.62	0.61	0.6	0.51	0.51	0.49	0.7	0.68	0.71	0.97	0.88	13.06	73.89	5.20	7.85
Full flowering	0.39	0.42	0.43	0.43	0.46	0.35	0.35					0.74	0.59	29.91	23.43	7.31	39.34
End of flowering	0.4	0.42	0.44	0.43	0.47	0.39	0.38					0.79	0.59	47.85	13.24	5.72	33.20
Flowering intensity	0.52	0.58	0.58	0.58	0.59	0.29	0.36	0.27	0.32	0.4	0.33	0.86	0.70	11.40	17.33	23.74	47.52
Weight of fruits	0.65	0.68	0.68	0.68	0.68	0.46	0.43	0.44	0.55	0.52	0.55	0.93	0.75	21.98	24.37	20.75	32.90
Number of fruits	0.66	0.67	0.67	0.67	0.67	0.49	0.46	0.42	0.54	0.51	0.52	0.93	0.69	20.83	24.53	21.59	33.04
Single fruit weight	0.46	0.57	0.58	0.57	0.58	0.48	0.48	0.33	0.7	0.71	0.52	0.98	0.74	59.67	9.93	8.96	21.44
Fruit diameter	0.46	0.49	0.5	0.5	0.51	0.46	0.47	0.47				0.90	0.65	60.54	0.84	8.63	29.99
Fruit length	0.36	0.41	0.42	0.41	0.44	0.38	0.4	0.39				0.88	0.62	48.98	2.27	11.69	37.05
Maximum fruit size	0.37	0.43	0.43	0.43	0.45	0.37	0.38	0.38				0.84	0.58	44.01	6.32	6.75	42.92
Fruit volume	0.38	0.46	0.46	0.46	0.49	0.45	0.45					0.85	0.62	60.95	1.36	5.92	31.77
Ground color	0.65	0.67	0.68	0.67	0.67	0.43	0.43	0.43	0.5	0.48	0.5	0.94	0.78	29.88	8.71	17.40	44.01
Yellow color	0.3	0.28	0.29	0.28	0.27	0.25	0.26	0.25				0.81	0.53	44.93	2.72	15.69	36.65
Green color	0.78	0.69	0.68	0.69		0.64	0.64	0.55				0.95	0.82	70.31	3.06	3.85	22.78
Red over color	0.88	0.74	0.73	0.74	0.74	0.66	0.65	0.65	0.84	0.85	0.85	0.99	0.85	74.57	7.77	5.93	11.72
Bitter pit freq.	0.35	0.36	0.35	0.36	0.34	0.24	0.25	0.17	0.6	0.59	0.4	0.96	0.33	39.46	6.05	17.87	36.63
Bitter pit grade	0.41	0.41	0.4	0.4	0.39	0.25	0.23	0.21	0.46	0.42	0.31	0.95	0.39	25.37	8.75	12.43	53.44
Russet cover	0.18	0.29	0.3	0.29	0.3	0.26	0.27	0.28	0.64	0.65	0.66	0.97	0.58	47.14	5.40	18.15	29.31
Russet freq. - overall	0.42	0.43	0.43	0.43	0.43	0.32	0.34	0.2	0.54	0.58	0.41	0.95	0.63	40.15	15.30	16.76	27.79
Russet freq. - stalk	0.42	0.41	0.42	0.41	0.42	0.31	0.33	0.21	0.56	0.57	0.41	0.95	0.62	28.16	24.80	14.52	32.52
Russet freq. - cheek	0.41	0.38	0.38	0.38	0.35	0.23	0.25	0.17	0.55	0.56	0.44	0.95	0.46	44.64	10.73	15.09	29.54
Russet freq. - eye	0.31	0.32	0.33	0.33	0.29	0.21	0.23	0.13	0.6	0.59	0.39	0.95	0.52	56.77	9.01	15.13	19.09
Titratable acidity	0.67	0.53	0.53	0.53		0.49	0.47	0.49	0.64	0.62	0.62	0.96	0.66	64.18	8.09	7.19	20.54
Soluble solids content	0.53	0.5	0.51	0.51	0.5	0.37	0.4	0.39	0.45	0.48	0.49	0.88	0.64	39.43	2.72	18.81	39.04
Fruit firmness	0.57	0.56	0.56	0.56	0.55	0.47	0.46	0.47	0.55	0.55	0.52	0.94	0.65	67.43	7.19	6.95	18.43
Water core freq.	0.39	0.42	0.44	0.43		0.41	0.4					0.84	0.60	60.46	0.02	10.70	28.82
Water core grade	0.43	0.42	0.44	0.43		0.4	0.39					0.79	0.62	59.20	1.10	8.79	30.91
Trunk diameter	0.54	0.54	0.55	0.54	0.54	0.38	0.38	0.38	0.56	0.55	0.57	0.98	0.66	10.94	66.34	4.20	18.53
Trunk increment	0.54	0.54	0.56	0.54	0.54	0.25	0.28	0.29	0.3	0.29	0.37	0.88	0.55	8.70	25.13	11.04	55.13

Supplementary table 7

Trait	Correlated trait	Prediction accuracy	SD
Titratable acidity	Fruit diameter	0.525609	0.061981
Fruit diameter	Titratable acidity	0.491001	0.050817
Bitter pit freq.	Bitter pit grade	0.343545	0.081623
Bitter pit grade	Bitter pit freq.	0.390105	0.060097
Bitter pit freq.	Trunk diameter	0.332353	0.076501
Trunk diameter	Bitter pit freq.	0.533829	0.066422
Bitter pit grade	Trunk diameter	0.386973	0.055504
Trunk diameter	Bitter pit grade	0.533283	0.066374
Trunk diameter	Ground color	0.535564	0.066691
Ground color	Trunk diameter	0.671722	0.056565
Ground color	Trunk increment	0.672008	0.056245
Trunk increment	Ground color	0.53526	0.062527
Bitter pit freq.	Red over color	0.337609	0.075143
Red over color	Bitter pit freq.	0.739011	0.031821
Bitter pit grade	Red over color	0.389884	0.05501
Red over color	Bitter pit grade	0.738576	0.032294
Fruit diameter	Fruit length	0.491894	0.051
Fruit length	Fruit diameter	0.411085	0.05557
Fruit diameter	Maximum fruit size	0.492099	0.049461
Maximum fruit size	Fruit diameter	0.435251	0.053689
Fruit diameter	Fruit volume	0.491481	0.050284
Fruit volume	Fruit diameter	0.460148	0.055604
Trunk diameter	Fruit firmness	0.536825	0.06784
Fruit firmness	Trunk diameter	0.552411	0.075133
Trunk increment	Fruit firmness	0.53682	0.062322
Fruit firmness	Trunk increment	0.55196	0.075937
Floral emergence	End of flowering	0.605237	0.074641
End of flowering	Floral emergence	0.472487	0.041722
Floral emergence	Full flowering	0.603539	0.075082
Full flowering	Floral emergence	0.458237	0.06057
End of flowering	Full flowering	0.439066	0.049184
Full flowering	End of flowering	0.422582	0.051252
Flowering intensity	Number of fruits	0.588871	0.049092
Number of fruits	Flowering intensity	0.668805	0.035668
Flowering intensity	Weight of fruits	0.589196	0.048346
Weight of fruits	Flowering intensity	0.675166	0.041838
Ground color	Number of fruits	0.670055	0.056591
Number of fruits	Ground color	0.666412	0.034769
Fruit firmness	Number of fruits	0.549294	0.075508
Number of fruits	Fruit firmness	0.66864	0.037001
Number of fruits	Soluble solids content	0.668528	0.035223
Soluble solids content	Number of fruits	0.500022	0.055645
Trunk diameter	Number of fruits	0.537088	0.067398
Number of fruits	Trunk diameter	0.669305	0.036841
Trunk increment	Number of fruits	0.533509	0.063722
Number of fruits	Trunk increment	0.668003	0.036164
Ground color	Weight of fruits	0.670575	0.056243
Weight of fruits	Ground color	0.674439	0.042163
Number of fruits	Weight of fruits	0.668209	0.036285
Weight of fruits	Number of fruits	0.67557	0.041911
Fruit diameter	Weight of fruits	0.512553	0.056492
Single fruit weight	Fruit diameter	0.58288	0.065395
Fruit length	Weight of fruits	0.438015	0.060637
Single fruit weight	Fruit length	0.583546	0.066655
Maximum fruit size	Weight of fruits	0.453398	0.059075
Single fruit weight	Maximum fruit size	0.57955	0.063441
Fruit volume	Weight of fruits	0.486046	0.058002
Single fruit weight	Fruit volume	0.581213	0.064039
Weight of fruits	Soluble solids content	0.67447	0.04052

Trait	Correlated trait	Prediction accuracy	SD
Soluble solids content	Weight of fruits	0.498122	0.055217
Trunk diameter	Weight of fruits	0.536319	0.067851
Weight of fruits	Trunk diameter	0.675411	0.042045
Trunk increment	Weight of fruits	0.532901	0.065035
Weight of fruits	Trunk increment	0.674626	0.042019
Fruit firmness	Harvest date	0.552141	0.075369
Harvest date	Fruit firmness	0.776023	0.035246
Number of fruits	Harvest date	0.669665	0.037186
Harvest date	Number of fruits	0.775292	0.036274
Weight of fruits	Harvest date	0.677783	0.043729
Harvest date	Weight of fruits	0.774724	0.036695
Harvest date	Russet freq. - cheek	0.77657	0.035295
Russet freq. - cheek	Harvest date	0.343838	0.065478
Soluble solids content	Harvest date	0.503562	0.058481
Harvest date	Soluble solids content	0.775057	0.035879
Fruit length	Maximum fruit size	0.412183	0.051979
Maximum fruit size	Fruit length	0.432486	0.055152
Fruit length	Fruit volume	0.411675	0.055184
Fruit volume	Fruit length	0.459601	0.057359
Maximum fruit size	Fruit volume	0.436763	0.054666
Fruit volume	Maximum fruit size	0.461765	0.053586
Russet cover	Russet freq. - overall	0.295361	0.096852
Russet freq. - overall	Russet cover	0.4287	0.07093
Russet freq. - cheek	Russet cover	0.338839	0.06655
Russet cover	Russet freq. - cheek	0.294106	0.098278
Russet cover	Russet freq. - eye	0.295147	0.098683
Russet freq. - eye	Russet cover	0.292576	0.102492
Russet cover	Russet freq. - stalk	0.295785	0.097143
Russet freq. - stalk	Russet cover	0.414987	0.081708
Soluble solids content	Russet cover	0.497664	0.05743
Russet cover	Soluble solids content	0.296771	0.097719
Russet cover	Yellow color	0.295692	0.100138
Yellow color	Russet cover	0.259309	0.08321
Russet freq. - cheek	Russet freq. - overall	0.346794	0.064629
Russet freq. - overall	Russet freq. - cheek	0.432276	0.069192
Russet freq. - overall	Russet freq. - eye	0.429705	0.072509
Russet freq. - eye	Russet freq. - overall	0.292636	0.10376
Russet freq. - overall	Russet freq. - stalk	0.432285	0.068133
Russet freq. - stalk	Russet freq. - overall	0.417721	0.079562
Soluble solids content	Russet freq. - overall	0.499176	0.058125
Russet freq. - overall	Soluble solids content	0.427504	0.067446
Russet freq. - cheek	Russet freq. - eye	0.343302	0.067164
Russet freq. - eye	Russet freq. - cheek	0.292107	0.106512
Soluble solids content	Russet freq. - cheek	0.499806	0.05585
Russet freq. - cheek	Soluble solids content	0.343724	0.067528
Russet freq. - cheek	Yellow color	0.343653	0.072044
Yellow color	Russet freq. - cheek	0.268037	0.076973
Russet freq. - eye	Yellow color	0.289059	0.107344
Yellow color	Russet freq. - eye	0.265415	0.080825
Russet freq. - cheek	Russet freq. - stalk	0.346151	0.063823
Russet freq. - stalk	Russet freq. - cheek	0.417821	0.078543
Russet freq. - eye	Russet freq. - stalk	0.290973	0.104238
Russet freq. - stalk	Russet freq. - eye	0.415371	0.082373
Fruit firmness	Soluble solids content	0.548033	0.07541
Soluble solids content	Fruit firmness	0.499765	0.055904
Trunk diameter	Soluble solids content	0.532411	0.06926
Soluble solids content	Trunk diameter	0.500993	0.055529
Trunk increment	Soluble solids content	0.53249	0.062752
Soluble solids content	Trunk increment	0.500413	0.057317
Trunk diameter	Trunk increment	0.533268	0.066238
Trunk increment	Trunk diameter	0.534556	0.060188

4 General discussion

Closing in on genomics-assisted breeding in apple implied a search for efficient strategies of genomic prediction and GWAS. The development of genomic prediction strategies remained an active research field even 20 years after the first report of genomic selection[1,2]. Aspects of genomic analyses such as phenotyping and genotyping efficiency, genomic prediction accuracy and genotype by environment interaction still pose a challenge to the establishment of genomic selection and are ultimately discussed here. This chapter also discusses the encountered challenges related to genome-wide association studies (GWAS) and describes imminent applications of the apple REFPOP.

4.1 A call for new phenotyping approaches

Manual phenotyping can restrain application of genomic selection due to the high labor costs[3]. To increase automation of trait measurements, a sorting machine (GREEFA) as an industrial solution using high-resolution optical systems and built-in software was deployed to measure nine traits related to productivity, fruit size and color. The sorting machine contributed measurements with no signs of decreased precision for eight traits, but low genomic prediction performance, pointing to a potential phenotyping problem, was found for yellow color. This trait showed a moderate phenotypic correlation of 0.54 with russet cover. These results suggest that the sorting machine might have interpreted some russet skin as yellow color. Moreover, restrictions on the access to data imposed by the sorting machine producer hindered automated acquisition of additional traits of interest such as the extent of russet. Instead, russet cover was estimated visually as a percentage of fruit surface covered by the russet, but measurement difficulties were experienced for this trait. While scoring the apple REFPOP genotypes, russet was observed around the stalk or eye cavity, scattered in small patches on fruit cheek, around lenticels or caused by frost or powdery mildew. These various forms of russet, often present in very low extent, made the phenotyping laborious and difficult, which probably showed in the obtained low genomic prediction accuracy. In a former study of apple russeting, scoring of russet cover from digital images appeared a more objective and precise method compared with visual scoring[4]. For traits such as yellow color or russet cover, establishment of automated recognition from images of apples taken with optical devices would be desirable to increase phenotyping precision and throughput in the future. An optical device for post-harvest phenotyping of fruits is in development as a result of a collaboration between ETH Zurich and Agroscope (Broggini G. and Patocchi A., personal communication). Additionally, near-infrared spectroscopy has also been identified as a reliable non-destructive phenotyping method when estimating traits such as soluble solids content, fruit acidity, firmness or water core in apple[5-7].

For high-throughput phenotyping of plants in fields or trees in orchards, drones with various sensors can provide remotely-sensed phenotypes characterizing large populations. From these remote sensing data, characteristics of the plants such as indices evaluating health of vegetation can be obtained[8]. Spectral data have been measured to phenotype bread wheat GNDVI (green normalized difference vegetation index estimating photosynthetic activity in the plant canopy) and sorghum plant height to finally perform genomic prediction with these phenotypes[9,10]. In apple, traits extracted from remote sensing data (box-counting fractal dimension, tree row volume, and alpha hull volumes)

were shown to be highly correlated (r>0.7) with trunk cross sectional area, tree height, total fruit yield, and tree leaf area[11,12]. Therefore, similar high-throughput measurements in apple may help accurately estimate and predict tree vigor, yield as well as symptoms of diseases and stress in the future. The high-throughput remotely-sensed assessment of fruit phenotypes in orchards is yet unavailable to apple breeding.

4.2 Towards cost-effective genotyping

Due to the known rapid decay of linkage disequilibrium in apple[13,14], high-density genotyping was desirable for the apple REFPOP to ensure linkage of markers with genomic regions associated with traits of interest. However, prices for high-density genotyping remained considerable for most apple breeding programs. Establishment of accurate high-density marker imputation of the dataset obtained with the Illumina Infinium® 20K SNP genotyping array[15] to the resolution of the Affymetrix Axiom® Apple 480K SNP genotyping array[16] in this thesis allowed for a cost-effective genotyping of many individuals. GWAS using these data indeed showed that high marker density was required for uncovering marker-trait associations in a complex trait such as floral emergence (see 2.3.5, Figure 5d). Nevertheless, marker density as low as 10K SNPs did not decrease genomic prediction accuracy when compared with the accuracy obtained using the full marker resolution of 303K SNPs (see 2.3.5, Figure 5e).

Aiming for genomic selection in apple, the necessary ~10K SNPs could be obtained with genotyping-by-sequencing (GBS). Although the rates of missing data can be lower using SNP arrays than applying GBS, the costs for genotyping with a SNP array remained substantially higher than for GBS while data produced by both platforms led to comparable genomic prediction model performance in wheat[17]. In apple, the GBS has been recently applied and produced datasets of 6–98K SNPs that led to predictions of low to high accuracy depending on the studied trait[18,19], even though a comparison between SNP arrays and GBS was not yet conducted in apple. Protocols used for GBS in the past studies[20-23] could be used as a basis for the establishment of GBS in additional germplasm. A comparison of prediction performance between SNP array and GBS datasets as well as evaluation of their overlap and estimation of missing data imputation accuracy would be desirable when looking for a solution to decrease genotyping costs in apple.

As an alternative to genomic selection based on SNP arrays and GBS, high-throughput phenotyping platforms can produce data for the so-called phenomic selection[24]. Instead of a kinship matrix constructed from genomic data to be applied in G-BLUP[25] such as it was done in this thesis, the kinship matrix can be based on phenotypic data from, e.g., near-infrared spectroscopy[24]. Phenomic selection built upon near-infrared spectroscopy could be used instead of genomic selection, which is based on more expensive genetic markers.

4.3 Ways to improve genomic prediction accuracy

Low to high genomic prediction accuracies were found for different traits in this thesis using the baseline model G-BLUP. The lowest prediction accuracies, i.e., below 0.3, were obtained for yellow color and russet cover. As discussed earlier in this chapter, their low prediction performance might have resulted from an insufficiently precise phenotyping. The highest prediction accuracies, i.e., larger than 0.7, were found for harvest date and

red over color. This observation was related to the relatively simple (oligogenic) genetic architecture of these traits, which were associated with a low number of large-effect loci and displayed a very high heritability (see 3.3.5, Figure 6). The number of causal loci as well as trait heritability, but also the size of the reference population affect precision of genomic prediction, as it was estimated in the first article following the equation by Elsen[26] (see 2.3.5, Figure 5fg). This simulation showed that the size of the apple REFPOP (534 genotypes) was sufficient to make precise genomic predictions for highly and moderately heritable traits underlined by up to ~100 causal loci. For traits of higher complexity, the simulation indicated that a reference population of approximately twice the size of the apple REFPOP (1,000 genotypes) may be needed to obtain more precise genomic predictions. The prediction accuracy can increase with an increasing reference population size until a plateau is reached, as it was shown for bread wheat[27]. In that study, the rate at which prediction accuracy increased was slower beyond approximately 2,000 genotypes[27]. A simulation study in maize found that fewer genotypes, i.e., 1,000, (and also fewer markers, i.e., 200–500) were needed for precise predictions for highly related reference and predicted populations than would be required for less related germplasm[28]. When genomic predictions should be performed in a breeding program using the apple REFPOP, the population might be enriched with specific breeding material to ensure relatedness between the reference and predicted populations. Since a plateau of prediction accuracy was not yet estimated for the apple REFPOP, these additional genotypes might also contribute to an increase in prediction accuracy alone due to the enlargement of the training population, the extent of the improvement depending on architecture of the predicted traits.

Most genomic prediction models only consider additive marker effects, although the estimation of non-additive effects (dominance and epistasis) may contribute to increased prediction accuracy[29]. A substantial contribution of the non-additive effects to phenotypic variance of several fruit quality traits measured from breeding material was observed in apple, but the genomic prediction accuracy was almost identical for models with or without the non-additive effects[30]. Estimation of the non-additive effects for the apple REFPOP dataset may further elucidate the role of these effects in a more diverse apple germplasm and for a broader spectrum of traits than shown before.

4.4 The puzzle of the genotype by environment interaction

Plant breeders generally agreed about the importance of genotype by environment interactions (G×E), although they have seemed divided over the solutions to this problem[31]. Since the establishment of genomic selection[1], various statistical approaches for genomic predictions accounting for G×E have become available[32,33]. One of these methodologies, the marker by environment interaction model (G-BLUP.E.G×E) described by López-Cruz et al.[34], was applied in this thesis. The G-BLUP.E.G×E model seemed to capture G×E in traits with a high proportion of their variance explained by the effect of G×E (~20%) such as flowering intensity and soluble solids content, but it did not succeed in capturing G×E for weight of fruits and number of fruits showing a similar G×E. Furthermore, the G-BLUP.E.G×E was not able to outperform the baseline model G-BLUP for any of the studied traits when phenotypes of unknown genotypes were predicted (see 3.3.4, Figure 5). In addition to the already applied G×E model, other models incorporating

G×E could be tested with the apple REFPOP dataset in an attempt to improve genomic prediction accuracy.

Extending the well-known G-BLUP model[25], the G-BLUP.E.G×E of López-Cruz et al.[34] introduced the main marker effects as well as the interaction effects of markers with environments using high-dimensional random variance-covariance structures. Two models similar to G-BLUP.E.G×E model but including nonlinear Gaussian kernels were proposed by Cuevas et al.[35]. For a maize dataset, the accuracy of the models with nonlinear Gaussian kernels was 5–6% higher than that of G-BLUP.E.G×E[35]. One weakness of the G×E models reported by López-Cruz et al.[34] and Cuevas et al.[35] was that they assumed positive correlations between environments[36]. As shown in Chapter 3 (see 3.3.4), this assumption resulted in low prediction accuracy for traits with phenotypic correlations between environments that were negative or close to zero. In the following work, Cuevas et al.[37] overcame this limitation proposing the multi-environment G×E model with linear and nonlinear kernels. Estimated for the same maize dataset, the accuracy of the latter model was on average 3% higher than the accuracy obtained with the model of Cuevas et al.[35]. The model of Cuevas et al.[37] should be tested with the apple REFPOP dataset in a comparison of G×E models.

The G×E models can also integrate environmental covariates such as measurements of soil or weather characteristics[33]. As the environmental covariates can be numerous and highly correlated, Heslot et al.[33] used a crop model to reduce the high dimension of daily weather data to fewer covariates estimated per growth stage of winter wheat before fitting the G×E model. When weather data were included in the model, more accurate predictions of genotype performance in unobserved environments were obtained (11% increase in accuracy)[33]. Millet et al.[38] applied a stage-wise environmental variable selection procedure to identify a final set of three environmental indices affecting maize yield, which were then used to characterize environments in a G×E model. Predictions made with similar G×E models including environmental covariates may be able to support breeding decisions in apple when selecting cultivars for specific environments in the future.

4.5 Challenges of the genome-wide association studies

When a diverse population like the apple REFPOP is used for GWAS, some individuals can be more closely related than others leading to subpopulations. The higher frequency of alleles in a subpopulation may lead to spurious marker-trait associations. Because it is not easy to avoid presence of subpopulations, a fixed effect of population structure (e.g., principal components) and kinship matrix as random effects have been widely adopted in GWAS algorithms for control of false positives[39]. Relying on these principles, the MLMM method[40] was applied in Chapter 2 and allowed to combine diverse accessions with highly related progeny. The more recent algorithm BLINK[41] used in Chapter 3 builds upon MLMM[40], which addresses the population structure, and SUPER[42], which contributes to increased computational efficiency and power to detect associations. The Chapters 2 and 3 assumed that correction of the effect of population structure was appropriate in MLMM and BLINK. However, avoiding false associations and understanding differences between results of various algorithms were recently identified as ongoing challenges in GWAS[39].

In Chapter 2 and 3, MLMM and BLINK were used with phenotypic data from one and three years of phenotyping, respectively. When MLMM was applied to harvest date, four significantly associated SNPs were found. Expanding the phenotypic dataset to three years of phenotyping and deploying BLINK, all four loci found for harvest date were rediscovered and located closely to the associations found by MLMM. The R^2 obtained for these new associations was on average 0.02 higher than for the associations of MLMM and one year of phenotyping. Three of these four loci were classified as major (R^2>0.01). Additionally, two minor loci were discovered using BLINK in the across-location GWAS and further 30 unique major and minor associations were found in the location-specific GWAS. For floral emergence, none of the p-values for the loci identified in the first article reached the Bonferroni-corrected $-\log_{10}$-transformed significance threshold when MLMM was applied with the full set of SNP markers, but these associations were significant in GWAS with MLMM based on various marker subsets (see 2.3.5, Figure 5). In the subsequent analysis with BLINK after three years of phenotyping, one significant association was found in the across-location GWAS and 26 additional unique associations were discovered in the location-specific GWAS. The results suggest a better performance of BLINK after three years of phenotyping, but the contribution of either the GWAS method or additional phenotyping to this outcome remain confounded. A direct comparison of GWAS models using the same dataset was able to disentangle these effects and showed that the statistical power to discover significant associations was higher in BLINK than MLMM (see Appendix).

To clarify the role and minimize false associations resulting from population structure in the GWAS output, the M-blot designed by Spindel et al.[43] plotted for GWAS of all location-year (and later also management) combinations could be used to identify associations, which are less likely an artifact of population structure. Applying this approach to location-year combinations available for harvest date and floral emergence, all known associations of large and low effect rediscovered in Chapter 2 and 3 using across-location GWAS were found again (associations on chromosomes 3, 10 and 16 for harvest date and 9 and 11 for floral emergence, see Appendix). The known associations of large and low effect were also obtained when using the M-blot to visualize GWAS performed for random genotype subsets of different size and structure (see Appendix). The novel association found in Chapter 3 for harvest date on chromosome 2 ($R^2 = 0.06$) was rediscovered using both additional GWAS analyses (see Appendix). However, neither the GWAS for location-year combinations nor GWAS for random genotype subsets was able to rediscover the novel loci of very low effect ($R^2 < 0.05$) found in Chapter 3 for harvest date on chromosome 17 and floral emergence on chromosome 10. This may be a result of very low $-log_{10}(p)$ values for these associations and the shown lower detection power of smaller genotype sets than that of the whole apple REFPOP (see Appendix). Nevertheless, the additional GWAS analyses added evidence that the known large and low-effect loci associated with harvest date and floral emergence, and one novel association with harvest date on chromosome 2, are not an artifact of population structure.

4.6 Some future applications of the apple REFPOP

A yet unexplored second part of the apple REFPOP multi-environment trial, where two replications of one third of the apple REFPOP genotypes were planted in parallel with the primary part used in Chapters 2 and 3 (see 2.4.5), was designed for comparing different

management practices under various environmental conditions[44]. In the future, phenotypes will be measured in the second part of each orchard grown under reduced irrigation or lower pesticide input. The application of these alternative management practices has been initially postponed to a later age of the orchards to reduce the risk of mortality in young trees due to water stress or pressure of pests. The new data can be used for estimating the proportion of the phenotypic variance explained by the effects of management, genotype by management interaction (G×M) and genotype by environment by management interaction (G×E×M) in the studied germplasm as well as for fitting genomic prediction models accommodating these interactions. Genomic selection based on these data should support the selection of germplasm resistant to apple pests and diseases or water stress.

Large size of the reference population and its relatedness with the predicted population are crucial factors in establishing genomic selection[28]. To this end, the presented dataset will be extended with breeding material phenotyped and genotyped in the same way as the apple REFPOP. The additional germplasm will stem from three Swiss breeding programs to test the impact of an increased population size on genomic prediction accuracy and to ensure relatedness between the reference and predicted populations, with the ultimate goal to establish genomic selection in the three breeding programs. The potential added value of the additional genotypic and phenotypic data will be evaluated by comparing genomic prediction accuracies obtained with the original apple REFPOP and the extended datasets.

Based on the initial simulations of genomic selection, the method could substantially increase the rate of genetic gain in plants, especially if combined with reproductive techniques to shorten the generation interval[1]. A method of speed breeding was designed to accelerate plant development when shortening generation time under the controlled environment of growth chambers[45]. Simulation of speed breeding in combination with genomic selection based on wheat data from the International maize and wheat improvement center (CIMMYT) showed that this combination of treatments might substantially reduce the length of the breeding cycle and increase genetic gain more than traditional phenotypic selection or genomic selection alone[46]. In apple, a speed breeding methodology called “fast track” breeding using controlled glasshouse conditions was developed to reduce the juvenility[47]. The method succeeded in shortening the juvenile period of apple seedlings from 4–5 to approximately 2.5 years[47]. Another method inducing rapid crop cycles in apple, the “early flowering” approach based on overexpression of a silver birch *MADS4* transcription factor, was able to reduce juvenility to only several weeks allowing for one generation per year[48]. The combination of “fast track” or “early flowering” approaches with genomic selection may improve the breeding efficiency in apple but remains to be tested.

The genotypic data of the apple REFPOP as well as methodologies for phenotypic and genomic analyses established in this thesis can be used with phenotypic measurements of additional traits. An example of such application poses the screening of the diverse apple REFPOP germplasm for resistance against *Neofabrea sp.* causing the bull’s-eye rot on stored apples[49]. GWAS based on the new phenotypic data may help discover the yet unknown loci responsible for resistance against the bull’s-eye rot disease.

4.7 Conclusion

Within the framework of this thesis, high-density genomic data for diverse germplasm and phenotypes for 30 quantitative trait measures were assembled into an extensive dataset, which alone can boost future breeding research. Using the dataset for GWAS, this thesis contributed to characterization of genetic architecture of the studied traits, reporting many known and more than 200 novel loci. The generated knowledge about trait biology can assist with creating DNA tests for marker-assisted selection and fuel genetic engineering approaches. Additionally, the results of GWAS confirmed the expectation that high marker density is needed for identification of loci associated with complex traits due to rapid linkage disequilibrium decay in apple, and recent model advancements in GWAS lead to increased power to detect significant associations. It was also shown here that a various number of loci stable across and specific to individual environments were detected depending on the analyzed trait. The contributed knowledge about these technical aspects of GWAS will be valuable when establishing GWAS for unexplored apple traits.

The large-scale genomic and phenotypic dataset was used to test various genomic prediction scenarios, which is essential prior to establishing genomic selection. No previous study in apple came as close to genomic selection for as many trait-environment combinations as it was done in this thesis. Performance of the so far largest and most diverse set of genomic prediction models in apple (single-environment univariate, single-environment multivariate, multi-environment univariate, and multi-environment multivariate models) was tested here. This work revealed that marker density as low as 10,000 SNPs did not decrease genomic prediction accuracy of a single-environment model. Therefore, less expensive genotyping technologies of low or moderate density can be applied as an alternative to the 480K SNP array when genotyping germplasm for the purposes of genomic selection in the future. Furthermore, it was shown that multi-environment prediction can be particularly accurate for incomplete trials, but also useful to evaluate untested genotypes. Multi-environment prediction was able to capture G×E for some but not all traits with a high proportion of their variance explained by the effect of G×E. This result will aid future analyses with additional multi-environment models. In addition, multivariate models were established as a useful tool for multi-environment genomic prediction in apple, but especially for prediction of traits measured at one location when phenotypes of a correlated trait measured across different locations were available. The thesis also identified the need for improved phenotyping of some traits and described potential gains from the expansion of the training dataset. Overall, the estimated genomic prediction accuracies for diverse models will aid the application of genomic selection in apple breeding programs, sustaining subsequent investigations in the field of genomics-assisted breeding of apple.

The apple remains one of the most important fruit crops worldwide. To secure a continuous production of apples appealing with their taste and appearance to the consumers, quantitative traits are key when breeding new varieties. Modern genomic technologies allow to make the breeding for a multitude of traits more efficient than it was in the past, but their potential was not yet exploited in apple breeding. This thesis provides a bridge between traditional phenotypic selection and genomics-assisted

breeding in apple, ultimately contributing to increased breeding efficiency and to sustainable apple production under the changing climate and consumer demands.

References

1 Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819 (2001).

2 Voss-Fels, K. P., Cooper, M. & Hayes, B. J. Accelerating crop genetic gains with genomic selection. *Theoretical and Applied Genetics* **132**, 669-686, doi:10.1007/s00122-018-3270-8 (2019).

3 Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R. & McCouch, S. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. *Theoretical and Applied Genetics* **126**, 867-887, doi:10.1007/s00122-013-2066-0 (2013).

4 Falginella, L. *et al.* A major QTL controlling apple skin russeting maps on the linkage group 12 of 'Renetta Grigia di Torriana'. *BMC Plant Biology* **15**, 150, doi:10.1186/s12870-015-0507-4 (2015).

5 Ventura, M., de Jager, A., de Putter, H. & Roelofs, F. P. M. M. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biology and Technology* **14**, 21-27, doi:10.1016/S0925-5214(98)00030-1 (1998).

6 Guo, Z. *et al.* Quantitative detection of apple watercore and soluble solids content by near infrared transmittance spectroscopy. *Journal of Food Engineering* **279**, 109955, doi:10.1016/j.jfoodeng.2020.109955 (2020).

7 Pissard, A. *et al.* Evaluation of a handheld ultra-compact NIR spectrometer for rapid and non-destructive determination of apple fruit quality. *Postharvest Biology and Technology* **172**, 111375, doi:10.1016/j.postharvbio.2020.111375 (2021).

8 Myneni, R. B., Hall, F. G., Sellers, P. J. & Marshak, A. L. The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing* **33**, 481-486, doi:10.1109/TGRS.1995.8746029 (1995).

9 Watanabe, K. *et al.* High-throughput phenotyping of sorghum plant height using an unmanned aerial vehicle and its application to genomic prediction modeling. *Frontiers in Plant Science* **8**, doi:10.3389/fpls.2017.00421 (2017).

10 Juliana, P. *et al.* Integrating genomic-enabled prediction and high-throughput phenotyping in breeding for climate-resilient bread wheat. *Theoretical and Applied Genetics* **132**, 177-194, doi:10.1007/s00122-018-3206-3 (2019).

11 Zhang, C. *et al.* Non-invasive sensing techniques to phenotype multiple apple tree architectures. *Information Processing in Agriculture*, doi:10.1016/j.inpa.2021.02.001 (2021).

12 Coupel-Ledru, A. *et al.* Multi-scale high-throughput phenotyping of apple architectural and functional traits in orchard reveals genotypic variability under contrasted watering regimes. *Horticulture Research* **6**, 52, doi:10.1038/s41438-019-0137-3 (2019).

13 Migicovsky, Z. *et al.* Genome to phenome mapping in apple using historical data. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0113 (2016).

14 Urrestarazu, J. *et al.* Genome-wide association mapping of flowering and ripening periods in apple. *Frontiers in Plant Science* **8**, 1923, doi:10.3389/fpls.2017.01923 (2017).

15 Bianco, L. *et al.* Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLOS ONE* **9**, e110377, doi:10.1371/journal.pone.0110377 (2014).

16 Bianco, L. *et al.* Development and validation of the Axiom®Apple480K SNP genotyping array. *The Plant Journal* **86**, 62-74, doi:10.1111/tpj.13145 (2016).
17 Elbasyoni, I. S. *et al.* A comparison between genotyping-by-sequencing and array-based scoring of SNPs for genomic prediction accuracy in winter wheat. *Plant Science* **270**, 123-130, doi:10.1016/j.plantsci.2018.02.019 (2018).
18 Kumar, S., Hilario, E., Deng, C. H. & Molloy, C. Turbocharging introgression breeding of perennial fruit crops: A case study on apple. *Horticulture Research* **7**, 47, doi:10.1038/s41438-020-0270-z (2020).
19 McClure, K. A. *et al.* Genome-wide association studies in apple reveal loci of large effect controlling apple polyphenols. *Horticulture Research* **6**, 107, doi:10.1038/s41438-019-0190-y (2019).
20 Elshire, R. J. *et al.* A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLOS ONE* **6**, e19379, doi:10.1371/journal.pone.0019379 (2011).
21 Scaglione, D. *et al.* Single primer enrichment technology as a tool for massive genotyping: A benchmark on black poplar and maize. *Annals of Botany* **124**, 543-551, doi:10.1093/aob/mcz054 (2019).
22 Silva-Junior, O. B., Grattapaglia, D., Novaes, E. & Collevatti, R. G. Design and evaluation of a sequence capture system for genome-wide SNP genotyping in highly heterozygous plant genomes: A case study with a keystone Neotropical hardwood tree genome. *DNA Research* **25**, 535-545, doi:10.1093/dnares/dsy023 (2018).
23 Barchi, L. *et al.* Single primer enrichment technology (SPET) for high-throughput genotyping in tomato and eggplant germplasm. *Frontiers in Plant Science* **10**, doi:10.3389/fpls.2019.01005 (2019).
24 Rincent, R. *et al.* Phenomic selection is a low-cost and high-throughput method based on indirect predictions: Proof of concept on wheat and poplar. *G3 Genes|Genomes|Genetics* **8**, 3961-3972, doi:10.1534/g3.118.200760 (2018).
25 VanRaden, P. M. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414-4423, doi:10.3168/jds.2007-0980 (2008).
26 Elsen, J.-M. An analytical framework to derive the expected precision of genomic selection. *Genetics Selection Evolution* **49**, 95, doi:10.1186/s12711-017-0366-6 (2017).
27 Norman, A., Taylor, J., Edwards, J. & Kuchel, H. Optimising genomic selection in wheat: Effect of marker density, population size and population structure on prediction accuracy. *G3: Genes|Genomes|Genetics* **8**, 2889-2899, doi:10.1534/g3.118.200311 (2018).
28 Hickey, J. M. *et al.* Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. *Crop Science* **54**, 1476-1488, doi:10.2135/cropsci2013.03.0195 (2014).
29 Varona, L., Legarra, A., Toro, M. A. & Vitezica, Z. G. Non-additive effects in genomic selection. *Frontiers in Genetics* **9**, doi:10.3389/fgene.2018.00078 (2018).
30 Kumar, S. *et al.* Genome-enabled estimates of additive and nonadditive genetic variances and prediction of apple phenotypes across environments. *G3 Genes|Genomes|Genetics* **5**, 2711-2718, doi:10.1534/g3.115.021105 (2015).
31 Allard, R. W. & Bradshaw, A. D. Implications of genotype-environmental interactions in applied plant breeding. *Crop Science* **4**, doi:10.2135/cropsci1964.0011183X000400050021x (1964).

32 Jarquín, D. *et al.* A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics* **127**, 595-607, doi:10.1007/s00122-013-2243-1 (2014).

33 Heslot, N., Akdemir, D., Sorrells, M. E. & Jannink, J. L. Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theor Appl Genet* **127**, 463-480, doi:10.1007/s00122-013-2231-5 (2014).

34 Lopez-Cruz, M. *et al.* Increased prediction accuracy in wheat breeding trials using a marker × environment interaction genomic selection model. *G3: Genes|Genomes|Genetics* **5**, 569-582, doi:10.1534/g3.114.016097 (2015).

35 Cuevas, J. *et al.* Genomic prediction of genotype × environment interaction kernel regression models. *The Plant Genome* **9**, plantgenome2016.2003.0024, doi:10.3835/plantgenome2016.03.0024 (2016).

36 Crossa, J. *et al.* Genomic selection in plant breeding: Methods, models, and perspectives. *Trends in Plant Science* **22**, 961-975, doi:10.1016/j.tplants.2017.08.011 (2017).

37 Cuevas, J. *et al.* Bayesian genomic prediction with genotype × environment interaction kernel models. *G3 Genes|Genomes|Genetics* **7**, 41-53, doi:10.1534/g3.116.035584 (2017).

38 Millet, E. J. *et al.* Genomic prediction of maize yield across European environmental conditions. *Nature Genetics* **51**, 952-956, doi:10.1038/s41588-019-0414-y (2019).

39 Tibbs Cortes, L., Zhang, Z. & Yu, J. Status and prospects of genome-wide association studies in plants. *The Plant Genome* **14**, e20077, doi:10.1002/tpg2.20077 (2021).

40 Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**, 825-830, doi:10.1038/ng.2314 (2012).

41 Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**, doi:10.1093/gigascience/giy154 (2018).

42 Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. A SUPER powerful method for genome wide association study. *PLOS ONE* **9**, e107684, doi:10.1371/journal.pone.0107684 (2014).

43 Spindel, J. E. *et al.* Association mapping by aerial drone reveals 213 genetic associations for *Sorghum bicolor* biomass traits under drought. *BMC Genomics* **19**, 679, doi:10.1186/s12864-018-5055-5 (2018).

44 Jung, M. *et al.* The apple REFPOP—a reference population for genomics-assisted breeding in apple. *Horticulture Research* **7**, 189, doi:10.1038/s41438-020-00408-8 (2020).

45 Watson, A. *et al.* Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants* **4**, 23-29, doi:10.1038/s41477-017-0083-8 (2018).

46 Voss-Fels, K. P. *et al.* in *Applications of Genetic and Genomic Research in Cereals* (eds Thomas Miedaner & Viktor Korzun) 303-327 (Woodhead Publishing, 2019).

47 Kellerhals, M. & Bühlmann-Schütz, S. "Fast Track" Apfelzüchtung. *Schweizer Zeitschrift für Obst- und Weinbau* **156**, 24 (2020).

48 Schlathölter, I. *et al.* Generation of advanced fire blight-resistant apple (*Malus* × *domestica*) selections of the fifth generation within 7 years of applying the early flowering approach. *Planta* **247**, 1475-1488, doi:10.1007/s00425-018-2876-z (2018).

49 Gariepy, T. D. *et al. Neofabraea* species associated with bull's-eye rot and cankers of apple and pear in the Pacific Northwest. *Canadian Journal of Plant Pathology* **27**, 118-124, doi:10.1080/07060660509507202 (2005).

5 Appendix

5.1 A comparison of modelling approaches to GWAS

5.1.1 Introduction

The main developments in GWAS algorithms focused on addressing population structure, improving statistical power, and increasing computational speed[1]. In diverse populations used for GWAS, it can occur that some individuals happen to be more closely related than others. In such subpopulations, both the trait of interest and some alleles can be present at a higher frequency, leading to false associations (i.e., false positives).

A widely adopted approach to correcting for population structure is the use of a mixed linear model with a fixed effect of population structure (often principal components) and a random effect of polygenic background defined by a kinship matrix[1]. The multi-locus mixed-model (MLMM) method was developed integrating these principles[2]. The Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK)[3] is a more recent method for GWAS, which was based on a combination of the MLMM[2] to address the population structure, and SUPER[4] to increase statistical power and computational speed. A higher power to detect marker-trait associations was reported for BLINK than MLMM[1,3].

In spite of the availability of algorithms able to account for population structure, avoiding false associations remained a challenge in GWAS[1]. Spindel et al.[5] proposed a methodology to minimize false associations resulting from population structure in GWAS output when repeated phenotypic measurements of the same population are available. In *Sorghum bicolor*, the study performed GWAS for combinations of traits, treatments, timepoints, and locations, resulting in 460 GWAS[5]. For an efficient plotting of many GWAS results, the authors designed an M-blot. The M-blot should be useful for visual assessment of subpopulation artifacts by examining how many circles fall in the physical space of a peak, i.e., in a vertical line. The more circles fall in a vertical line, the less likely should the peak be an artifact of population structure.

In this study, the MLMM and BLINK were compared in terms of their power to detect significant marker-trait associations. Due to the diversity of germplasm included in the apple REFPOP and its weak population structure (see Chapter 2), the impact of population structure on MLMM and BLINK was examined here repeating GWAS for combinations of traits and environments. Finally, the effect of population size and structure on associations detected with BLINK was assessed.

5.1.2 Methods

The GWAS with MLMM[2] and BLINK[3] was applied using the R package GAPIT 3.0[6]. The genomic matrix $\boldsymbol{M}$ ($n \times m$ matrix for $n = 534$ genotypes with $m = 303{,}148$ markers) was used as input. The GWAS was performed with two principal components and the minor allele frequency threshold set to 0.05. Marker-trait associations were identified as significant for p-values falling below a Bonferroni-corrected significance threshold with $\alpha = 0.05$.

For the first analysis (Figure 1), the across-location BLUPs (see Chapter 3) were used as response. The GWAS with MLMM and BLINK was performed for every of the 30 analyzed traits. The resulting significant marker-trait associations were compared with former studies (see Chapter 3). In the second analysis focused on harvest date and floral emergence analyzed in both the Chapter 2 and 3 (Figure 2), the adjusted phenotypic values of each genotype (see Chapter 3) estimated for every environment (location-year combination) were used as response. This resulted in 15 GWAS for harvest date and 14 GWAS for floral emergence with each of the compared models (i.e., MLMM and BLINK). The third analysis (Figure 3) tested the impact of population size and structure on BLINK. The GWAS with BLINK was performed with the across-location BLUPs for harvest date and floral emergence used as response. Before fitting the model, 25 random samples of 80% genotypes each were drawn from every of the four different sets of apple REFPOP genotypes: (i) the whole apple REFPOP (each of the random samples was equal to 80% of the 534 genotypes), (ii) equal proportions of the accession and progeny group amounting to approximately 50% of the genotypes from each group (each sample equal to 80% of the 268 genotypes), (iii) the progeny group (each sample equal to 80% of the 265 genotypes) and (iv) the accession group (each sample equal to 80% of the 269 genotypes). The resampling resulted in 100 different GWAS for each analyzed trait.

5.1.3 Results and discussion

Statistical power to detect associations with MLMM and BLINK

BLINK and MLMM were run for 30 traits under the exact same settings as were used for GWAS in Chapter 3, which allowed for a direct comparison between the models (Figure 1). The associations found with BLINK showed a larger overlap with former studies and were generally more numerous than the associations found with MLMM. As expected[1,3], these results pointed to a higher power to discover significant associations using BLINK than MLMM.

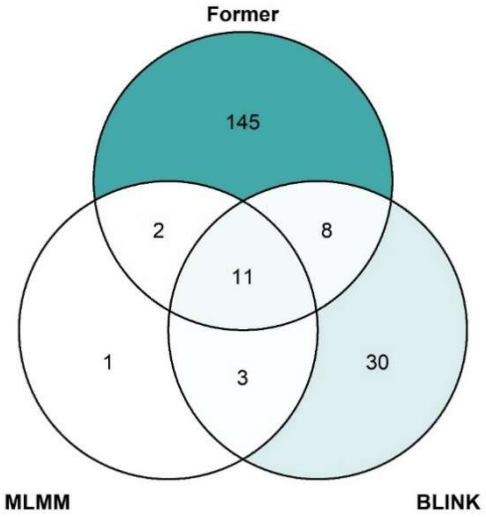


Figure 1: Venn diagram comparing former studies with the associations found using MLMM and BLINK. Color intensity reflects the number of associations per diagram area.

Impact of population structure on MLMM and BLINK

Inspired by the M-blot[5], the Figure 2 showed that several circles were falling on a few vertical lines for both harvest date and floral emergence. For harvest date, the circles aligned on at least three chromosomes (chromosome 3, 10 and 16 with known loci of large effect, see Chapter 2 and 3). For floral emergence, the circles aligned most often on two chromosomes (chromosome 9 and 11 with known loci of low effect, see Chapter 2). In addition to the known loci, other genomic regions were observed with circles repeatedly falling in a vertical line for both traits (Figure 2). These results would suggest that the known loci rediscovered in Chapter 2 and 3 and some additional loci are not an artifact of population structure. As the M-blot was established for as many as 460 GWAS in Spindel et al.[5], a weakness of this approach applied to apple REFPOP might be the low number of compared GWAS (15 for harvest date and 14 for floral emergence). Another concern may be that repeated measurements of the same population could again lead to results affected by the same population structure. Instead of testing the effect of population structure, this approach might be rather useful to test the stability of loci across environments.

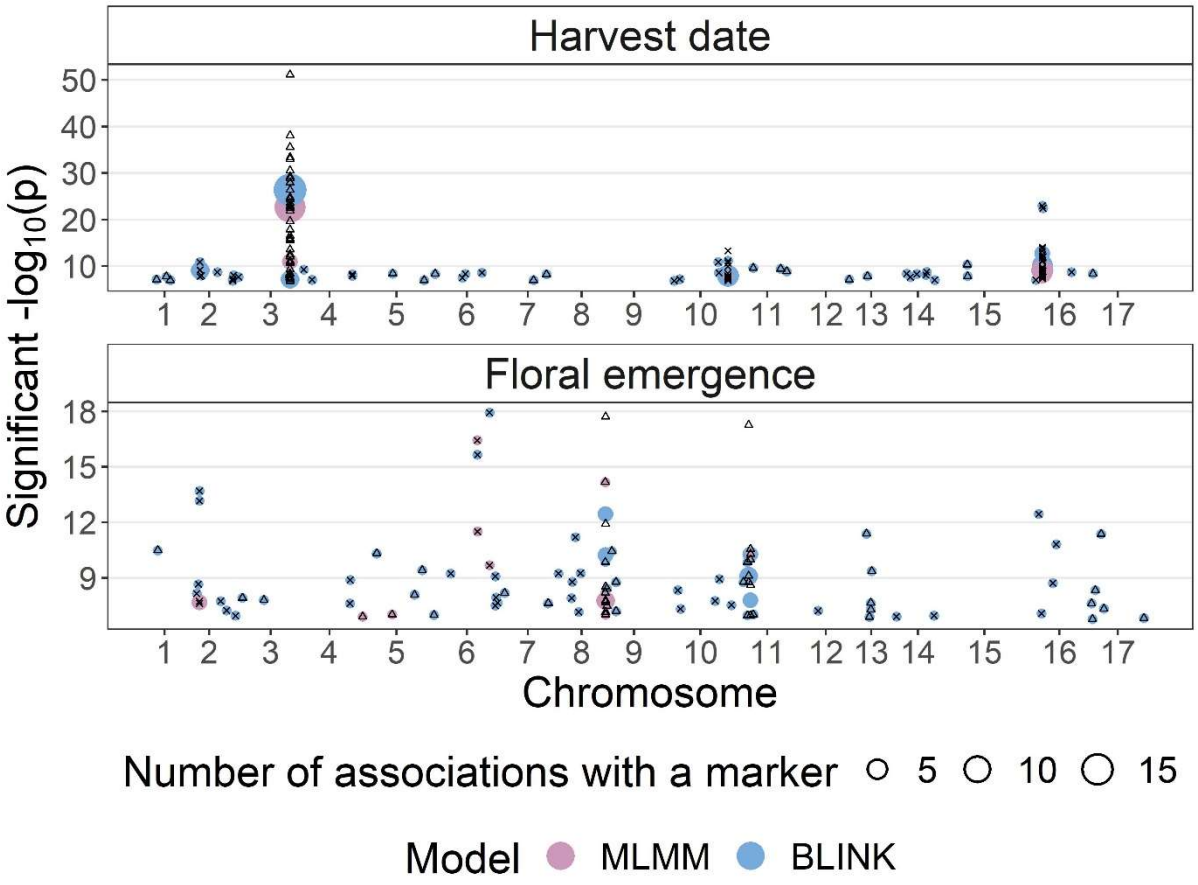


Figure 2: Manhattan plots of the significant marker-trait associations found for harvest date and floral emergence using MLMM and BLINK. Significant associations (i.e., SNPs) on chromosomes with odd numbers are shown as triangles, for chromosomes with even numbers as crosses. The size of the colored circles shows the number of associations with a marker (the larger the circle, the more often the SNP was significant in different GWAS). The position of the colored circles along y-axis gives the median of the Bonferroni-corrected $-log_{10}(p)$ across all GWAS where the SNP was significant. The circle color stands for the model used to perform GWAS. The x-axis shows physical positions of the significant associations on chromosomes, the y-axis indicates the Bonferroni-corrected significant $-log_{10}(p)$. The plot shows a combined result of GWAS for every location by year combination.

The impact of population size and structure on associations detected with BLINK

The impact of population size and structure was tested only with BLINK because this model showed a higher statistical power to detect significant associations than MLMM (Figure 1). Before fitting BLINK, 25 random samples of 80% genotypes each were drawn from every of the four different sets of apple REFPOP genotypes: (i) the whole apple REFPOP, (ii) equal proportions of the accession and progeny group amounting to approximately 50% of the genotypes from each group, (iii) the progeny group and (iv) the accession group. It is generally known that GWAS power decreases with a decreasing population size[1]. All known loci (on chromosomes 3, 10 and 16 for harvest date, on chromosomes 9 and 11 for floral emergence) were rediscovered if the full set of apple REFPOP genotypes was resampled (Figure 3a). However, the locus on chromosome 10 for harvest date and the locus on chromosome 11 for floral emergence were barely visible after the second genotype set of equal proportions of the accession and progeny group amounting to approximately 50% of the genotypes from each group was resampled (Figure 3b). Additionally, one of the novel loci of low effect ($R^2 < 0.1$) found on

chromosome 17 for harvest date (Chapter 3) was not obtained here with any of the resampled genotype sets, i.e., 20% decrease in the number of genotypes used for GWAS resulted in a loss of this association. These results confirm that the power to detect significant associations in GWAS decreases as the number of individuals in the genotype set is reduced.

Except for the novel locus on chromosome 17 where the lack of its association with harvest date resulted from a decreased sample size, all five known loci (on chromosomes 3, 10 and 16 for harvest date, on chromosomes 9 and 11 for floral emergence) and the novel loci (on chromosome 2 for harvest date, on chromosome 10 for floral emergence, Chapter 3) were rediscovered using GWAS based on resampling of the whole apple REFPOP and at least one apple REFPOP group (i.e., accession and progeny group, Figure 3). As shown in Chapter 3, all progeny was fixed for the reference allele of the SNP on chromosome 3 associated with harvest date. Therefore, this locus could not be rediscovered using GWAS of the progeny group. Although the allele frequencies were not examined for the remaining loci analyzed here, the allele frequencies of the SNP on chromosome 3 suggest that (i) in case an association could not be rediscovered in the progeny group, this was due to a lower allelic diversity of that locus in the progeny group than in the accession group, (ii) in case the association could not be rediscovered in the accession group, such association was a sole contribution of the allelic diversity present in the progeny group. Hypothetically, a locus discovered in the set of all apple REFPOP genotypes, but not rediscovered separately for the progeny and/or the accession group, could be assumed a product of population structure. Such locus would potentially be associated with the differences between the accession and the progeny groups rather than between individual genotypes. As this was not the case for any known and novel loci analyzed here, the associations should not be assumed an artifact of population structure.

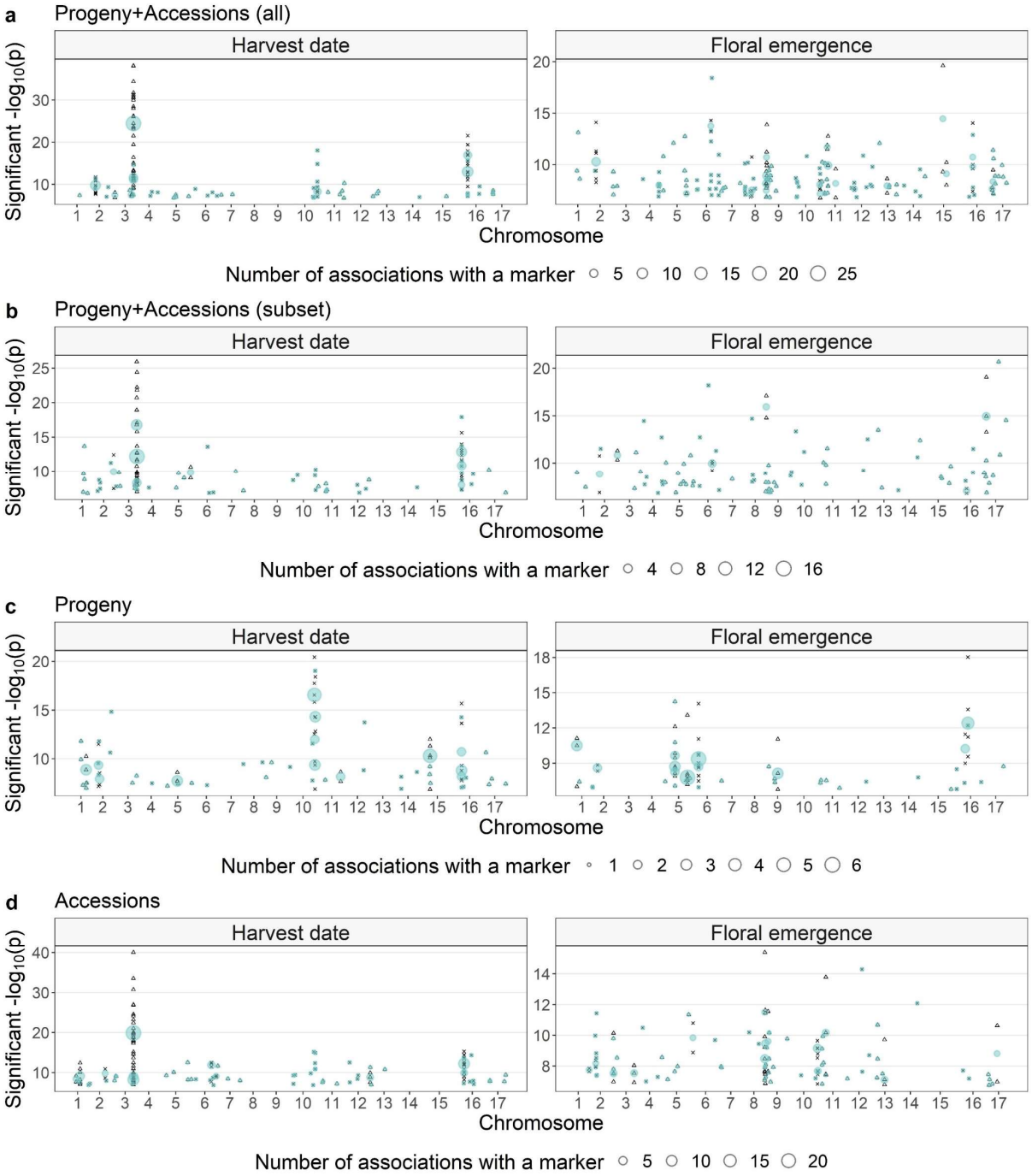


Figure 3: Manhattan plots of the significant marker-trait associations found for harvest date and floral emergence using BLINK and four sets of apple REFPOP genotypes: (a) the whole apple REFPOP, (b) equal proportions of the accession and progeny group amounting to approximately 50% of the genotypes from each group, (c) the progeny group and (d) the accession group. The GWAS was repeated 25-times for every genotype set, each of the runs based on a random sample of 80% genotypes from the set. Associations (i.e., SNPs) on chromosomes with odd numbers are shown as triangles, for chromosomes with even numbers as crosses. The size of the colored circles shows the number of associations with a marker (the larger the circle, the more often the SNP was significant in different GWAS). The position of the colored circles along y-axis gives median of the Bonferroni-corrected $-log_{10}(p)$ across GWAS where the SNP was significant. The x-axis shows physical positions of the significant associations on chromosomes, the y-axis indicates the Bonferroni-corrected significant $-log_{10}(p)$.

5.1.4 Conclusion

The results presented here confirmed that using more individuals in GWAS can improve statistical power to detect significant marker-trait associations. When comparing MLMM and BLINK, enhanced power was found for BLINK, which corresponded with BLINK being a more recent methodological development that builds upon MLMM, aiming for higher power and computational efficiency. Furthermore, the analyses with both MLMM and BLINK showed no signs of false associations with harvest date and floral emergence. The weak population structure resulting from two different kinds of germplasm in the apple REFPOP – the accessions and the progeny – was well accounted for in the performed GWAS.

References

1 Tibbs Cortes, L., Zhang, Z. & Yu, J. Status and prospects of genome-wide association studies in plants. *The Plant Genome* **14**, e20077, doi:10.1002/tpg2.20077 (2021).

2 Segura, V. *et al.* An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**, 825-830, doi:10.1038/ng.2314 (2012).

3 Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**, doi:10.1093/gigascience/giy154 (2018).

4 Wang, Q., Tian, F., Pan, Y., Buckler, E. S. & Zhang, Z. A SUPER powerful method for genome wide association study. *PLOS ONE* **9**, e107684, doi:10.1371/journal.pone.0107684 (2014).

5 Spindel, J. E. *et al.* Association mapping by aerial drone reveals 213 genetic associations for *Sorghum bicolor* biomass traits under drought. *BMC Genomics* **19**, 679, doi:10.1186/s12864-018-5055-5 (2018).

6 Tang, Y. *et al.* GAPIT version 2: An enhanced integrated tool for genomic association and prediction. *The Plant Genome* **9**, doi:10.3835/plantgenome2015.11.0120 (2016).

Acknowledgements

I would like to thank to Prof. Dr. Bruno Studer for the opportunity to conduct this thesis and his unwavering support and supervision during its course. I am deeply grateful to Dr. Andrea Patocchi for his invaluable assistance, support, and guidance at all times. Andrea, I could not have imagined having a better advisor and mentor than you. I would also like to offer my sincere thanks to my advisors Dr. Hélène Muranty, Dr. Morgane Roth and Dr. Beat Keller for their insightful feedbacks, fruitful discussions, and contributions to the thesis' methodology and datasets. I extend my gratitude to Prof. Dr. Fred van Eeuwijk for accepting the task of a co-examiner. I thank the partners from the apple REFPOP consortium for entrusting me with their data and their thoughtful comments. I would like to express my gratitude to the members of the Breeding Research Group at Agroscope, the Molecular Plant Breeding Group at ETH Zurich and to other colleagues for their repeated help, stimulating discussions and happy distractions. I am very grateful to my family for always being there for me.