Studies in the genus *Fritillaria* L. (Liliaceae)

Peter D. Day, BSc.

October 2017

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

Supervisors:

Prof. Andrew R. Leitch

Dr Ilia J. Leitch

Queen Mary
University of London

Statement of Originality

I, *Peter Donal Day*, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without the prior written consent of the author.

Signature:

Date: October 31st 2017

Details of collaboration and publications.

Chapter 2 is published in Day *et al*. (2014):

Day PD, Berger M, Hill L, Fay MF, Leitch AR, Leitch IJ, Kelly LJ (2014).

Evolutionary relationships in the medicinally important genus *Fritillaria* L.

(Liliaceae). *Molecular Phylogenetics and Evolution*, **80**: 11-19.

Martyn Rix and Takato Natsui provided plant material; Jeremy Broome, Richard Kernick and Kit Strange cultivated and supplied plant material from the RBG Kew Living Collection; Jaumé Pellicer helped with figures; Yasmin Milner and Matthew O Reilly produced some of the plastid sequences used in this study; Donald Fraser helped with the LCNG screen; László Csiba and Edith Kapinos provided education about DNA extractions; Sven Buerki gave guidance on using software program PAUPRat; Christine

Leon provided guidance on medicinal usage of *Fritillaria*; Haris Saslis-Lagoudakis indulged in useful discussions about bioprospecting. All authors contributed to editing and commenting on the original manuscript.

Abstract

Genus *Fritillaria* consists of approximately 140 species of bulbous herbaceous perennials with a bicontinental distribution (Eurasia, North America) across the temperate zone of the Northern Hemisphere. The genus includes species of horticultural commercial value and those used in Traditional Chinese Medicine (TCM) preparations. Since the mid-16th century species within genus *Fritillaria* have attracted much attention as ornamental plants in the Middle East and Europe. The genus has received much attention by geneticists in recent years because it includes species with an exceptionally large range of genome sizes, including a subgenus containing some of the largest diploid genome sizes so far recorded in plants.

This thesis first presents the most comprehensive analysis of the phylogenetic relationships between the species to date, published in 2014 in *Molecular Phylogenetics and Evolution*. The work reveals two distinctive clades, one containing a small number of species from N. America and N. E. Asia and the other, with the majority of species, found in Europe, N. Africa, Middle East, China and Japan. This phylogenetic distribution of *Fritillaria* species indicates that two independent shifts towards giant genomes have taken place during evolution of the genus.

One particularly charismatic species occurring across Eurasia is the Snake's-head Fritillary (*F. meleagris*) with both horticultural and conservation importance.

The thesis presents new genetic markers to analyse the relationships between populations of *F. meleagris* across Eurasia. Three distinct populations were found, in Northern Europe/Scandinavia, Eastern Europe and the Russian Federation. All the English populations are almost certainly derived from populations originally occurring in Northern Europe and likely introduced into England through multiple introductions. However, its origin in the British Isles remains a mystery, although recent research within this project has thrown new light on the history of its first record and description in France and its early presence in Britain.

Contents

List of Figures

List of Tables

Acknowledgement

I will start by thanking Andrew and Ilia Leitch for the understanding they have shown during my attempts at tackling the tasks which lay ahead, and gently pushing me in the right direction in such an encouraging way.

Next to be thanked must be Laura Kelly for the immense amount of patience she showed whilst introducing me to the intricate delights of the lab, how to make the most of my time while incarcerated there and teaching me about the dark art of phylogenetic analysis.

My thanks also go to Dion Devey, Edith Kapinos, László Csiba and Robyn Cowan for preventing a novice in the lab making a fool of himself in front of students many decades his junior, and also with so much more experience.

Special mention must go to Laura Martinez-Suz and Jaumé Pellicer for not only helping me understand how to get the most of my experimentations in the lab, but also for their immense kindness towards me throughout my stay at the Jodrell.

I would like to thank Mike Fay for providing me with such good contacts in universities across Europe, through which I was so efficiently able to start the process of collecting *Fritillaria meleagris* leaf samples for my research.

Likewise, Kevin Walker from BSBI must be thanked for enabling me to get started so quickly with collecting leaf samples in England, by providing me with valuable background on the locations and history of the most ancient *F. meleagris* populations across the country.

Without the selfless help, uncharged for, by the collectors of leaf samples across Europe, I would not have had a population genetics project to study and develop. My great thanks go to Ivan Schanzer in Moscow, Ivan Sostaric in Belgrade, Cosmin Sicora in Jibou, Albert Corporaal in Hasselt, Gerd Vogg in Würzburg, Dylan Tatti in Neuchâtel, Bernd Netz in Haseldorf, Anders Backlund in Uppsala, Elzbieta Kuta in Krakow, Valery Malecot in Angers and Michael Popchenko in Timiriazev.

During my research into the early history of *Fritillaria* in Europe, both Florike Egmond and her husband Peter Mason in Rome were of great help in enabling me to determine how to recover to great effect the huge wealth of information about European natural history in the 16th century.

Last but certainly not least, my thanks to Laurence Hill for his great enthusiasm and skill in everything *Fritillaria*, and who is indefatigable in his quest for ever more knowledge about his favourite subject, and so willing to share it with me.

Dedicated to my long-suffering wife Kay for her unswerving support throughout these past seven years and without whom I could not have achieved my goal.

List of Abbreviations

AAMC	Assigned Allele Membership Coefficient
AP	ALKALINE PHYTASE
BCE	Before the Common Era
BLAST	Basic Local Alignment Tool
BP	Bootstrap percentage
BSA	Bovine Serum Albumin
CE	Common Era
CTAB	Cetyltrimethylammonium bromide
dNTP	Deoxynucleotide triphosphate
Gb	Gigabase pairs
HTS	High Throughput Sequencing
HWE	Hardy-Weinberg Equilibrium
IAM	Infinite Allele Model
ITS	Internal transcribed spacer
IUCN	International Union for the Conservation of Nature
LCNG	Low-copy nuclear genes
LD	Linkage disequilibrium
MS	MALATE SYNTHASE
MCMC	Markov Chain Monte Carlo
MG	Mesotrophic Grassland
nrITS	Nuclear ribosomal internal transcribed spacer
PCoA	Principal coordinates analysis
PCR	Polymerase Chain Reaction
pg	Picograms
PP	Posterior Probability
RBG	Royal Botanic Gardens, Kew
SMM	Stepwise Model of Mutation
TCM	Traditional Chinese Medicine
T_m	Primer melting temperature
TPM	Two-Phase Model

Chapter 1

General introduction

1.1 Genus *Fritillaria* (Liliaceae)

The genus *Fritillaria* consists of approximately 140 species of bulbous herbaceous perennials with a bicontinental distribution (Eurasia, North America) across the temperate zone of the Northern Hemisphere (Rix, 2001; Hill, 2013; WCSP, 2014). The genus has been shown to be broadly divided into two strongly supported clades and at least eight subgenera comprising: *Fritillaria*, *Japonica*, *Rhinopetalum*, *Petilium*, *Korolkovia*, *Theresia*, *Davidii* and *Liliorhiza* (Rix, 2001; Ronstead *et al*., 2005; Day *et al*., 2014). One clade contains species belonging to subgenus *Liliorhiza* which are predominantly found on the western side of North America. The second clade comprises species from the remaining subgenera and are distributed across Eurasia (the phylogenetic analysis of the genus by Day *et al*. (2014) is presented in Chapter 2 of this thesis).

The genus includes species of horticultural importance regarding their commercial value, as well as those used in Traditional Chinese Medicine (TCM) preparations. hree hybrids are known in the wild (Hill, 2011): *F. eastwoodiae* (*F. affinis* × *F. micrantha*) (Macfarlane, 1978), *F. gentneri* (*F. affinis* × *F. recurva*) (Meyers *et al*., 2006). and *F. rhodocanakis* (*F. graeca* × *F. rhodocanakis*) (Zaharof, 1987).

1.2 Why study genus *Fritillaria*?

Species of *Fritillaria* are characterised by an exceptionally large range of genome sizes in the diploid species (1C-value[1]: *c*. 29.7 pg to 100.1 pg (*c*. 29 Gb[2] to *c*. 98 Gb)) (Figure 1.1) and includes some of the largest genomes recorded in plants when

[1] C-value (or genome size) is the total amount of DNA in the unreplicated haploid or gametic nucleus of an organism (Greilhuber *et al*., 2005).
[2] 1 pg = 978 Mb (Doležel *et al*., 2003).

compared to the measured genome sizes of 7,542 angiosperm species (Bennett and Leitch, 2012; I.J. Leitch, pers. comm.). Across angiosperms as a whole, genome sizes vary *c*. 2400-fold, with the range skewed towards small genomes, resulting in modal and mean sizes of 1C-values being *c*. 0.6 pg and 1C = *c*. 5.6 pg respectively (Figure 1.1). The phylogenetic distribution of *Fritillaria* species with large genomes indicates that two independent shifts towards giant genomes have taken place during the evolution of the genus (Day *et al.*, 2014; Kelly *et al*., 2015)), thereby enabling these separate instances of genomic expansion to be investigated through comparison of the genomic processes involved. These exceptional features of *Fritillaria* provide excellent opportunities to study the mechanisms underpinning genome size evolution, as large-scale comparative analyses of plant genome sizes have shown that plants with large genomes are, for example, at greater risk of extinction, less likely to be highly specious, less adaptable to living in polluted or nutrient-poor soils, and are less able to tolerate extreme environmental conditions (Knight and Ackerley, 2002; Vinogradov, 2003; Knight *et al*., 2005; Temsch *et al*., 2010; Greilhuber and Leitch, 2013; Puttick *et al.*, 2015). The further study of genus *Fritillaria* will enhance understanding of the mechanisms, processes and evolutionary dynamics involved.

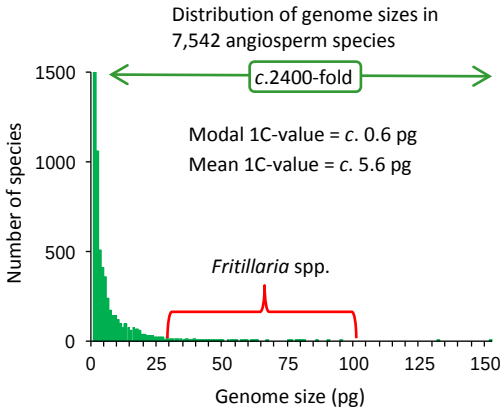

Figure 1.1 - Data from the Plant DNA C-values database (release 6.0, Dec 2012 and J. Pellicer, pers. comm.).

1.3 Traditional Chinese Medicine (TCM)

For more than 2,000 years species within the genus *Fritillaria* have been recognised for the medicinal properties and commercial value of their bulbs within the Traditional Chinese Medicine (TCM) market for antitussive, antihypertensive and expectorant remedies (*e.g*. sore throat, cough, asthma, bronchitis) (Atta-ur-Rahman *et al*., 1994; Shang and Liu,1994; Pharmacopoeia of the People's Republic of China, 2010; Zhang *et al*., 2010). Since accession of the Communist regime in 1949, the herbal tradition has flourished in China and today is recognised as a valid medical system alongside the far more expensive Western conventional medicine. With the discovery of pharmacologically-active isosteroidal and steroidal alkaloid compounds within the

bulbs of *Fritillaria* species, sourced mainly from *F. cirrhosa*, *F. przewalskii*, *F. unibracteata*, *F. delavayi*, *F.taipaiensis*, and *F. thunbergii* (Leon and Lin, 2017), these 'official' species remain one of the most important sources for TCM. However, greatly increased use of TCM in recent decades has led to over-harvesting from wild populations, leading to the majority of the species now threatened due to unsustainable collecting practices (Leon and Lin, 2017). Identification through further phylogenetic studies of the closest relatives of the 'official' species could point to other species that might be analysed for their medicinal properties. These could possibly be used in place of the 'official' species, thereby reducing pressure on the endangered species currently so intensively harvested.

1.4 History of *Fritillaria* species in Europe

The invention of the mechanised printing press by Johannes Gutenberg *c*. 1439 (Britannica, 2017) enabled the proliferation of *materia medica* – descriptions and images of medicinal plants – to be brought to the attention of the wider medical profession and public for the first time. It was during the early 16th century that the broad publication of herbals, with their woodcuts of herbs, engendered an interest in plants, not just for their practical utility in medical and culinary pursuits, but also for their aesthetic value (Arber, 1938). It was during the middle of the 16th century that travellers from Europe, particularly to the Levant (loosely defined as Syria, Jordan, Iraq, Lebanon and Israel) discovered the attractive blooms cultivated within gardens, particularly in Turkey and Iraq, and returned with bulbs for either a commercial opportunity to be realised, for scientific study or to be presented as gifts. The first European encounters with *Fritillaria* species were the product of this discovery (Stone, 2015).

1.5 Why study *Fritillaria meleagris*?

Snake's-head Fritillary (*F. meleagris*) has almost mythical iconic status in the wild flora of the British Isles as a plant surrounded by legend. The combination of its attractiveness, rarity, accompanying folklore, doubt about its position as either a British native or alien, archaeophyte or neophyte and, if the latter, how it arrived in Britain, and its many different names, have been a great source of debate. The result has been probably the greatest number of papers and articles generated about any individual

plant species in the British flora during the past half-century (Turrill, 1951; Grigson, 1955; Zhang, 1983; Oswald, 1992; King and Wells, 1993; Mabey, 1996; Kent, 2001; Pearman, 2007; Griffiths, 2015). There is rather less interest in *F. meleagris* in continental Eurasia than in Britain. Other than for reasons of conservation, this is probably because the plant is assumed to be native in every country where it is present and therefore not a subject for enquiry. However, throughout the region it is treated as a plant assigned varying levels of conservation importance depending on the country, each providing an official designation indicating its vulnerability within the local environment using the criteria set out by the International Union for Conservation of Nature (Zych and Stpiczyńska, 2012; IUCN, 2017).

For such an iconic species as *F. meleagris*, not only because of its conservation importance to the British Isles and elsewhere, but also its substantial horticultural trade, it is perhaps surprising that nothing is known of its population structure across Eurasia. This deficiency is addressed in this thesis.

1.6 Species characteristics

Fritillaria meleagris is a spring-flowering bulbous perennial communal species (Rodwell, 1992; Blamey *et al*., 2003; Stace, 2010) and is the type species of genus *Fritillaria* L., the type material of which is held in the Linnaean Herbarium at the Linnean Society of London (https://www.linnean.org).

Individuals of *F. meleagris* typically have distinctive reddish-purple tessellated, square-shouldered, drooping bell-shaped flowers 3 – 5cm long, with a perianth whorl of six tepals (Figure 1.2). Nevertheless, a cream-white or a sulphurous yellow coloured flower morph is sometimes observed with no tessellation, but often with pale-green colouration near the nectaries (Figure 1.3) (occurring in 0 - *c*. 7% of individuals within a population) (King and Wells, 1993; pers. obs.; statistics from every population leaf sample collector, pers. comm.; I. Tatarenko, pers. obs.). Potentially, this flower colour polymorphism arises through an inability or partial

Figure 1.2 - *F. meleagris*, red form. P. Day

Figure 1.3 - *F. meleagris*, colour morph. P. Day

ability to synthesise anthocyanin pigmentation. The common occurrence of the cream white/yellow flower morph, indicates a low, but relatively stable allele frequency for that flower colour, perhaps maintained by environmental heterogeneity (Richards, 1986; Bell *et al.*, 2000; Warren and Mackenzie, 2001), or perhaps for the reddish-purple colour, heterozygote advantage associated with selection against homozygous less relatively fit white morph allelic combinations.

Flowering stems are reported to grow in height to 20 cm (50 cm, pers. obs.), with sometimes twin or, rarely, triple-flowered heads (pers. obs.), supporting up to 6 (-8) leaves (Zhang, 1983; I. Tatarenko, pers. obs.; pers. obs.). A mature capsule contains between 50 - 100 seeds (Turrill, 1951; Gibbons, 1990; Oswald, 1992; pers. obs.). From a genomic perspective, the chromosome number reported for the species is 2n = 24, while its genome size is reported as and 1C-value *c*. 47.3 pg (*c*. 46.3 Gb) (Bennett and Leitch, 2011).

In Britain, pollination is usually by solitary bee species belonging to the genus *Bombus* (bumblebee - 89%), *Apis mellifera* (honeybee - 6%), *Andrena* spp. (mining bee - 5%) (R. Comont and I. Tatarenko, pers. comm.). Both *Bombus* spp. and *Apis* spp. dominate pollinator activity for *F. meleagris* across Eurasia (Zych and Stpiczyńska, 2012; R. Comont, pers. obs.). The flowers are hermaphrodite and can self-fertilise (Trist, 1981). The plant reproduces by both seed set and bulb division (King and Wells, 1993).

Plant bulbs contain the poisonous alkaloid imperialine ($C_{27}H_{43}NO_3$), which can induce vomiting and possible cardiac arrest, but otherwise have no medicinal value (Forsyth, 1954; Grieve, 1984). Parkinson (1629) made reference the plant's lack of medicinal value when stating 'I have not found or heard by any others of any property peculiar in this plant [*Fritillaria meleagris*], to be applied either inwardly or outwardly for any disease: the chief or only use thereof is to be an ornament for the gardens of the curious lovers of these delights…'. This assessment, emanating probably from a much earlier period than Parkinson, is probably why the plant was not included in any herbals until the late 16th century, when publications began to include ornamentals as well as plants with a medicinal efficacy (pers. obs.).

1.7 Habitats

Across its range, *F. meleagris* is usually found in periodically wet (from autumn, winter and sometimes early spring surface flooding), low-lying, circum-neutral unimproved eutrophic alluvial hay meadows and pastures (usually floodplains) and occasionally in (humid) woods (Turrill, 1951; Rix, 1968; Gillam,1993; Stace, 2010), thriving where annual hay-making is followed by aftermath grazing (Wells, 1994). Within Britain, the plant is found in species-rich National Vegetation Classification (NVC) type MG4 (*Alopecurus pratensis* (meadow foxtail) – *Sanguisorba officinalis* (great burnet)) mesotrophic grassland plant communities. (Turrill, 1951; Simpson, 1982; Rodwell, 1992; Jefferson and Pinches, 2011; JNCC, 2017; RBGKew, 2017a).

The species favours deep friable neutral to calcareous fine-textured clay-rich or silty alluvial loam soils with low to moderate concentrations of macronutrients in the range 5-15 mg/kg dry soil (Gowing *et al.*, 2002); and a pH ranging from about 5.8 to 7.0 maintained by fertile silt supplying nitrogen (N) and phosphorus (P) in particulate form, delivered to the land through overland flooding, which also acts as a frost inhibitor. This means that water rarely stands for long periods, as it drains freely and is drawn down when local water course levels fall (Wells and King, 1975; Price, 1983; King and Wells, 1993; Rothero *et al*., 2011).

1.8 Life cycle of *F. meleagris*

The species grows well in meadows farmed in the traditional Lammas Land regime. Bare patches of ground left by autumn/winter flooding in locations where water tables are high (Horton and Jefferson, 2006) create areas where *F. meleagris* seedlings can germinate. The above-ground part of the plant's annual cycle generally takes place between March and May, and as the plants cannot tolerate domestic animal grazing during the growing season, the sward is allowed to grow until hay cropping takes place in June/July, just after seed ripening and dispersal has completed (Corporaal *et al.*, 1993). The re-growth ('aftermath') is livestock-grazed until mid-February (traditionally, Lammas extends from 12th August – 12th February), following which the land is freed of livestock to allow the sward to grow again as hay, which also coincides with the plant's growing season (Gibbons, 1990; FMP, 2016).

The plant flowers from early to late spring across its Eurasian natural range, with seeds ripening from June to July (FMP, 2016), with bulbs in a dormant condition and without roots until August to September (FMP, 2016). The bulbs start to grow again in

autumn, followed by another growth of roots in early spring. Axillary buds to lower scale leaves form daughter bulbs, resulting in vegetative reproduction (Turrill, 1951). The seed-to-flower-to-seed reproductive cycle of an adult takes at least 5 (-6) years (Zhang, 1983; FMP, 2016). However, population sizes can appear to vary dramatically year on year, depending on the natural cycle of the environmental variables of temperature, inundation and drought (Zhang, 1983), reflecting years where bulbs may remain dormant for an annual cycle and/or erratic seedling recruitment. There is likely to be substantial vegetative reproduction, as well as recruitment from seed (Sell and Murrell, 1996). *Fritillaria meleagris* flowers are known to be grazed to destruction by pigeons, pheasants and rabbits (Trist, 1981; Price, 1983).

1.9 Outlook on this thesis

This thesis has a number of aims and objectives:

(1) To understand patterns of species divergence across the *Fritillaria* genus hitherto unresolved with regard to infrageneric relationships, with implications for genome size evolution and medicinal use of *Fritillaria* species. This to be addressed by creating a phylogenetic reconstruction of relationships encompassing most of the currently recognised species diversity in the genus (Chapter 2).

(2) To gain insights into the history of *Fritillaria* in Europe to determine its origin (s) and to attempt resolution of the disputed origin of the British and European *F. meleagris* populations, by exploring the ancient literature and determining the earliest recorded references and descriptions (Chapter 3).

(3) To comprehend the extent of genetic diversity between Eurasian populations of *F. meleagris*, by developing population genetic markers that can be used to distinguish between different populations of *F. meleagris* (Chapter 4) and apply them to determine the substructure of *F. meleagris* populations across Eurasia (Chapters 5).

(4) To understand the conservation implication of these data, by synthesising the findings to provide evidence-based advice of conservation importance relating to *F. meleagris* across Eurasia and to suggest how the work in the thesis can be taken forward (Chapter 6).

Chapter 2

Evolutionary relationships in the medicinally important genus *Fritillaria* L.

2.1 Summary

Fritillaria (Liliaceae) is a genus of approximately 140 species of bulbous perennial plants that includes taxa of both horticultural and medicinal importance. As well as being commercially valuable, *Fritillaria* species have attracted attention because of their exceptionally large genome sizes, with all values recorded to date in excess of 30 Gb. Despite such interest in the genus, phylogenetic relationships between the majority of species have remained untested. This chapter presents the first phylogenetic reconstruction of evolutionary relationships that encompasses most of the currently recognised species diversity in the genus. Three regions of the plastid genome were sequenced in 117 individuals of *Fritillaria*, representing 92 species (*c*. 66% of the genus) and in representatives of nine other genera of Liliaceae. Eleven low-copy nuclear gene regions were also screened in selected species for their potential utility. Phylogenetic analysis of a combined plastid dataset using maximum parsimony and Bayesian inference provided support for the monophyly of the majority of currently recognised subgenera. However, subgenus *Fritillaria*, which is by far the largest of the subgenera and includes the most important species used in traditional Chinese medicine, is found to be polyphyletic. Moreover, several taxa that were represented by multiple individuals show evidence of species non-monophyly. The Japanese endemic subgenus *Japonica*, which contains the species with the largest recorded genome size for any diploid plant, is resolved as sister to the predominantly Middle Eastern and

Central Asian subgenus *Rhinopetalum*. Whilst relationships between most of the major *Fritillaria* lineages can now be resolved, the results also highlight the need for data from additional independently evolving loci; an endeavour that may be particularly challenging in light of the huge nuclear genomes found in these plants.

2.2 Introduction

The genus *Fritillaria* (Liliaceae) comprises approximately 140 species of bulbous perennials (Rix, 2001; Hill, 2013; WCSP, 2014), which are distributed in the temperate zone of the Northern Hemisphere and currently divided into eight subgenera (Rix, 2001). Members of *Fritillaria* are characterised by their extremely large genomes sizes. All species examined to date have genomes in excess of 1C = 30 Gb (Bennett and Leitch, 2012), which is more than 190 times larger than the genome of the model plant *Arabidopsis thaliana* (Bennett *et al*., 2003). Moreover, these immense genomes have arisen primarily in the absence of recent whole genome duplication. Polyploidy is rare in *Fritillaria*, with triploid individuals occasionally reported for some species (*e.g*. Leitch *et al*., 2007; Ambrožová *et al*., 2011) and scarce records of higher ploidies (*e.g*. DeWoody and Hipkins, 2012). The largest genomes for diploid individuals of *Fritillaria* are found among the Japanese endemic species in subgenus *Japonica* (Leitch *et al*., 2007), which reach over 85 Gb in size (Ambrožová *et al*., 2011). Consequently, *Fritillaria* has become the focus of genomic studies aimed at elucidating the processes that govern genome size diversity and evolution (Ambrožová *et al*., 2011; Kelly and Leitch, 2011).

In addition to their extraordinarily large genomes, *Fritillaria* species have attracted much attention because of their commercial value, partly as ornamental plants, but principally as a source of material for use in Traditional Chinese Medicine (TCM). The medicinal use of *Fritillaria* bulbs (‘‘beimu’’) extends back more than 2,000 years, and today they are one of the most widely used TCMs (Shang and Liu, 1994; Lin *et al*., 2001; Chen, 2005; Zhang *et al*., 2010). In China, the production of medicinal preparations containing *F. cirrhosa* is an industry with an estimated value of US $400 million per year (Chen, 2005). Although some bulbs are cultivated, most are still collected directly from the wild (Chen, 2005). The use of *Fritillaria* extracts in TCM is also well established in the Himalayas (India, Nepal and Pakistan), Japan, Korea and Southeast Asia. Other countries (*e.g*. Burma, Turkey) are also involved in supplying the

increasing demand (Chen and Hsia, 1977; Atta-ur-Rahman *et al*., 1994; Li *et al*., 2006; Zhang *et al*., 2010). Products containing beimu are used for antitussive, expectorant and antihypertensive purposes, and extractions from bulbs of several species of *Fritillaria* have been found to contain bioactive isosteroidal and steroidal alkaloid compounds (Yan *et al*., 1999; Lin *et al*., 2001; Li *et al*., 2006; Xiao *et al*., 2007; Konchar *et al*., 2011). The Pharmacopoeia of the People's Republic of China (2010), Flora of China (Xinqi and Mordak, 2000) and the Chinese herbal medicine materia medica (Bensky *et al*., 2004) together record 16 *Fritillaria* species that are used as source material for beimu (the large number of synonyms for some Chinese species, see Xinqi and Mordak (2000), complicates the comparison of lists of medicinal species from different sources and has led to larger numbers of ''species'' being recorded as medicinally useful in some publications, *e.g*. Hao *et al*. (2013)).

Current understanding of evolutionary relationships within *Fritillaria* is based primarily on the findings of Rønsted *et al*. (2005) who conducted a molecular phylogenetic analysis of 37 species (*c*. 26% of the genus), including representatives of all subgenera, and recovered strong support for two major clades; one comprising species from the mainly North American subgenus *Liliorhiza* and the other made up of species from the seven remaining subgenera. The analysis was carried out employing two regions of the plastid genome (the intron of *rpl*16 gene and the less conserved *mat*K gene and its flanking *trn*K intron sequence (Patwardhan *et al*., 2014)) and the nuclear ribosomal internal transcribed spacers (ITS1 and ITS2).

However, despite the recovery of some well resolved and strongly supported groupings by Rønsted *et al*. (2005), several subgenera were not well represented (*e.g.* only *c*. 15% of the largest subgenus, *Fritillaria*, was sampled) and the placement of certain lineages that are of key importance for understanding genome size evolution, such as subgenus *Japonica*, remained unclear. Therefore, an improved estimation of evolutionary relationships within *Fritillaria* is needed to provide a phylogenetic framework for the interpretation of results from genomic analyses. Moreover, a better understanding of the relationships within the genus could have important implications for the medicinal use of *Fritillaria* species. Utilisation of certain medicinal species has increased in recent years, leading to over-harvesting of wild populations (Zhang *et al*., 2010; Hao *et al*., 2013). Identifying the closest relatives of the *Fritillaria* species used in TCM could point to additional species that might be analysed for their potential medicinal value, which may in turn reduce pressure on those species that are currently being collected

intensively. Rønsted *et al*. (2005) included a single Chinese medicinal species in their study (*F. pallidiflora*) and intriguingly found that it did not group with the other members of subgenus *Fritillaria*, where it is currently classified (Rix, 2001), but was instead resolved in a distinct clade that included species from subgenera *Korolkovia*, *Petilium* and *Theresia*. Although more recent phylogenetic studies of *Fritillaria* have been undertaken, these have not advanced understanding of relationships within the genus as a whole as they have either been based on small geographically restricted samples of species (*e.g*. Türktaş *et al*., 2012) or have not included estimates of support for the relationships recovered (*e.g*. Hao *et al*., 2013). By employing the *rbc*L gene as an additional and more conserved plastid sequence to those previously sampled by Rønsted *et al*. (2005) (Patwardhan *et al*., 2014), it expected that resolution of deeper phylogenetic relationships will be possible.

The aim of this study is to resolve outstanding questions regarding relationships between the major taxonomic groups of *Fritillaria*, in order to provide an evolutionary framework for the interpretation of genomic data and give a better understanding of the phylogenetic distribution of species used in TCM. To achieve this, we have sampled the majority (*c*. 66%) of the approximately 140 currently accepted species of *Fritillaria* and included a minimum of *c*. 60% of species from each of the eight subgenera.

2.3 Materials and methods

2.3.1 Taxon sampling

Samples of 117 individuals representing 92 *Fritillaria* species, encompassing all subgenera and sections *sensu* Rix (2001), were included in the study (approximate percentage of species represented from each subgenus is as follows: *Davidii*, *Japonica*, *Korolkovia*, *Petilium*, *Theresia* 100%; *Rhinopetalum* 80%; *Fritillaria*, *Liliorhiza* 60%). Where possible, intraspecific samples were included for species with recognised subspecies or varieties and for those with wide geographic distributions. Thirteen species of the sister genus *Lilium*, and one species of *Nomocharis*, were sampled. Several phylogenetic studies have indicated that *Nomocharis* nests within *Lilium*, and it has been recognised that *Lilium* should be formally revised to incorporate *Nomocharis* Gao *et al*., 2012a). Therefore, henceforth we use ‘‘*Lilium*’’ to refer to *Lilium* and *Nomocharis*. Nine species from eight additional genera of Liliaceae *sensu* Fay and

Chase (2000) were used as outgroups. For a list of taxa included and details of voucher specimens see Supplementary Table S1 and Figure S1.

2.3.2 DNA isolation

DNA extractions were sourced from the DNA databank at RBG, Kew (http://data.kew.org/dnabank/homepage.html), with additional plant material taken from the RBG Kew Living Collections and from the personal collection of L. Hill (see www.fritillariaicones.com). Total genomic DNA was extracted from fresh or silica-dried leaf tissue using a 2 x CTAB method modified from Doyle and Doyle (1987). DNA samples were precipitated in chilled ethanol (-20 °C) for *c*. 24 h and purified by caesium chloride/ethidium bromide equilibrium density gradients (1.55 g/ml). Purified DNA was dialysed in 1 x Tris-EDTA buffer and stored at -80 °C.

2.3.3 Amplification and sequencing of plastid regions

Three regions of the plastid genome were sampled: (i) a *c*.1.6 kb portion of the *mat*K gene; (ii) *c*.1.4 kb portion of the *rbc*L gene; (iii) a *c*.1.4 kb portion of the *rpl*16 gene (spanning most of the intron and part of the 3' flanking exon). PCR amplification of target regions was carried out in reactions with a final volume of 25 µl, containing 1 µl genomic DNA, 0.5 µl (0.4 µg/µl) BSA (Promega), 22.5 µl ReddyMix™ PCR Master Mix (ABGene, Epsom, Surrey, UK) at 2.5 mM $MgCl_2$ concentration, 50 ng (50 ng/µl) of each primer, using a GeneAmp® PCR system 9700 (Applied Biosystems, Warrington, Cheshire, UK). Primer sequences and amplification conditions used for each of the plastid regions are listed in Supplementary Table S2. Amplified products were cleaned using a NucleoSpin® Extract II kit (Macherey–Nagel, Bethlehem, PA, USA) following the manufacturer's protocol. Cycle sequencing reactions were carried out with a Big Dye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Warrington, Cheshire, UK) using a Techne TC-412 thermal cycler (Bibby Scientific, Stone, Staffordshire, UK). Cycle sequencing products were analysed on a 3730 DNA Analyzer (Applied Biosystems, Warrington, Cheshire, UK).

2.3.4 Amplification and sequencing of low-copy nuclear gene (LCNG) regions

The nuclear ribosomal internal transcribed spacer (nrITS) was used previously for inferring phylogenetic relationships in *Fritillaria* (Rønsted *et al.*, 2005). However, sequencing of additional species revealed the widespread occurrence of polymorphic positions in electropherograms, with cloning of a sample of PCR products

demonstrating, in some cases, multiple nrITS sequence types that do not cluster by species (data not shown). Intragenomic heterogeneity in nrITS sequences has also been noted for *Lilium* (Muratović *et al*., 2010a). Therefore, to search for alternative regions of the nuclear genome for use in our phylogenetic studies we conducted a screen of primers from selected low-copy nuclear genes (LCNGs). LCNGs with primers shown to work across a wide taxonomic range, or in groups closely related to *Fritillaria*, were identified from the literature (Supplementary Table S3). In addition, sequences of commonly used LCNGs from taxa closely related to *Fritillaria* were identified from GenBank and new primers were designed from these sequences (Supplementary Table S3). In total, 11 regions from 10 LCNGs were screened in six species selected to represent major species groups within *Fritillaria* and one species of *Lilium* (see footnote in Supplementary Table S1). PCRs were conducted using the amplification conditions recommended in the original publications containing the primer sequences or, in the case of newly designed primers, using conditions optimised to the length of the target region and primer melting temperatures following standard PCR guidelines (Supplementary Table S3). Where single banded amplification products of suitable strength were obtained, PCR products were cleaned and sequenced as described for the plastid regions. Sequences were searched against the nr/nt database in GenBank using BLASTn (Altschul *et al*., 1997) with the default settings.

2.4 Sequence alignment and dataset optimisation

Contigs were assembled and edited using Sequencher™ version 4.5 (Gene Codes, Ann Arbor, MI, USA). The three plastid datasets were aligned separately in PAUP* v4.0b10 (Swofford, 2003); alignments for *mat*K and *rpl*16 included some sequences previously published by Rønsted *et al*. (2005) (see Supplementary Table S4 for details of sequences included in the alignment for each region and for GenBank accession numbers). Two inversions in *mat*K sequences were identified and alignments adjusted to separate the blocks of inverted and non-inverted bases. Indels within coding regions were positioned to maintain the correct reading frame in MacClade v4.08 OS X (Maddison and Maddison, 2005). Optimisation of the *rpl*16 alignment was carried out using webPRANK-F (Löytynoja and Goldman, 2010); the webPRANK-F alignment was compared with the alignment constructed manually, and several sections optimised in the latter as a result.

2.5 Phylogenetic analyses

Analysis of separate plastid regions yielded congruent tree topologies (result not shown), therefore the final analyses were based on a combined plastid dataset. Preliminary maximum parsimony (MP) analysis of the combined dataset of 3,901 characters/ 140 terminals was conducted in PAUP* v4.0b10 (Swofford, 2003) with the following parameter settings: Analysis 1 – heuristic search; equal character weighting; DELTRAN character-state optimisation; gaps treated as missing data; 1,000 random addition sequence replicates; tree bisection-reconnection (TBR) branchswapping, with 10 trees held at each step and multiple trees saved; Analysis 2 – with the same settings as for Analysis 1, but with 500,000 random addition sequence replicates and limiting the number of trees saved per replicate to 10. These analyses failed to run to completion due to shortage of computer RAM, each having saved >1.4 million trees. An alternative parsimony analysis was undertaken in PAUP* v4.0b10 (Swofford, 2003) using the parsimony ratchet method implemented with PAUPRat (Nixon, 1999; Sikes and Lewis, 2001).

In contrast to standard heuristic methods, the parsimony ratchet method samples many tree islands with fewer trees from each island, and therefore has the potential to find further distinct islands of most parsimonious trees. Consequently, this method may be particularly useful for analysis of large datasets. Initial optimisation conducted with 10 searches each of 200 and 400 iterations in combination with character perturbations of 10%, 15% and 25%, showed no variation in strict consensus tree topologies. For the final analysis, 10 independent parsimony ratchet heuristic searches each of 200 iterations with 15% character perturbation per iteration were performed cumulatively using the same settings as above, but with a single random addition sequence replicate and one tree saved per iteration. All saved trees were imported into PAUP* v4.0b10 (Swofford, 2003) and filtered to retain only the shortest trees, which were used to construct a single strict consensus tree. Bootstrap analysis (Felsenstein, 1985) was used to assess node support, with 1,000 heuristic pseudoreplicates conducted in PAUP* v4.0b10, using the same settings as above for ‘Analysis 1’ with a single random addition sequence replicate, but holding 10 trees at each step. Relative levels of homoplasy were assessed for each of the plastid regions using the consistency and retention indices in PAUP* v4.0b10 (Swofford, 2003).

Model selection and phylogenetic analyses using Bayesian inference were performed using MrModeltest v2.3 (Nylander, 2004) and the parallel (MPI) version of MrBayes v3.1.2 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003; Altekar *et al.*, 2004) through the University of Oslo Bioportal (www.bioportal.uio.no) as described in Kelly *et al*. (2013), with the exception that two independent MrBayes runs of 10 million generations were performed rather than four. Post analysis diagnostics were performed using Tracer v1.5 (Rambaut and Drummond, 2009) and AWTY (Nylander *et al*., 2008) as described in Kelly *et al.* (2013). Trees corresponding to the first one million generations of each run were discarded as the burn-in period. A 95% majority rule consensus tree was constructed in PAUP* v4.0b10 by combining post burn-in trees from the two runs.

2.6 Results

2.6.1 Combined plastid analyses

From the 3,901 characters included in the analysis of the combined plastid dataset, 411 (10.5%) were autapomorphic and 412 (10.6%) were synapomorphic (Table 2.1). Results from the Bayesian and maximum parsimony analyses are largely congruent (Figure 2.1, Supplementary Figure S2). There is one group that receives ≥ 70% bootstrap support, but < 0.95 posterior probability (PP; Supplementary Figure S2). Seventeen clades have PP ≥ 0.95 and a bootstrap percentage (BP) < 70 (Figure 2.1), but all are present in the consensus tree from the parsimony analysis (Supplementary Figure S2). The 95% majority rule consensus of Bayesian analysis post burn-in trees is shown in Figure 2.1; evidence for evolutionary relationships of *Fritillaria* is inferred from this consensus tree.

Table 2.1 - Details of datasets used for phylogenetic analyses.

Region	Number of taxa	Number of characters	Autapomorphic characters (%)	Synapomorphic characters (%)
*mat*K	140	1,480	213 (14.4)	213 (14.4)
*rbc*L	135	1,098	46 (4.2)	68 (6.2)
*rpl*16	140	1,323	152 (11.5)	131 (9.9)
Combined plastid	140	3,901	411 (10.5)	412 (10.6)

Fritillaria and *Lilium* together form a strongly supported monophyletic group (PP1/BP99), but *Fritillaria* itself is not recovered as monophyletic (Figure 2.1). Two strongly supported clades of *Fritillaria* species are recovered, one containing only species from subgenus *Liliorhiza* (PP1/BP98) and the other containing species from the remaining seven subgenera proposed by Rix (2001) (PP1/BP92); these two clades form a polytomy with *Lilium* (Figure 2.1). In the parsimony consensus tree *Lilium* is nested within *Fritillaria*, but with only 53% bootstrap support (Supplementary Figure S2). The largest subgenus, *Fritillaria*, which contains *c*.100 species (Rix, 2001; > 70% of the genus), is polyphyletic. The majority of species sampled from this subgenus (54) form a strongly supported clade (PP1/BP99; *F.* subgenus *Fritillaria* A, see Figure 2.1) comprising taxa that occur mainly in Europe, the Middle East and North Africa, but including some species (*F. meleagroides* and *F. usuriensis*) with ranges extending into China (Figure 2.2, Supplementary Figure S3). The remaining six species sampled from subgenus *Fritillaria*, comprising taxa occurring in China and Central Asia (with some that extend into North and South Asia; Figure 2.2, Supplementary Figure S3), form a separate well-supported clade (PP1/BP87; *F.* subgenus *Fritillaria* B). This clade groups with three small subgenera, *Petilium*, *Korolkovia* and *Theresia*, which occur in the Middle East and Central Asia (PP1/BP75; Figure 2.1, Figure 2.2). Excluding subgenus *Fritillaria*, all subgenera including multiple species are recovered as monophyletic (Figure 2.1). All eight species of subgenus *Japonica* form a strongly supported clade (PP1/BP97) with subgenus *Rhinopetalum* well supported as sister (PP0.96/BP70; Figure 2.1).

Of the 19 *Fritillaria* species represented by multiple individuals in this study, eight are monophyletic, three are unresolved and eight are non-monophyletic on the basis of data from the plastid genome (Figure 2.1). Of the eight species with evidence for non-monophyly, four are resolved as polyphyletic (*F. amana*, *F. assyriaca* subsp. *assyriaca*, *F. crassifolia* and *F. thessala*), three as paraphyletic (*F. involucrata*, *F. muraiana* and *F. obliqua*) and one (*F. amabilis*) as uncertain (due to lack of resolution; this species could be polyphyletic or paraphyletic if fully resolved) in the plastid tree (Figure 2.1).

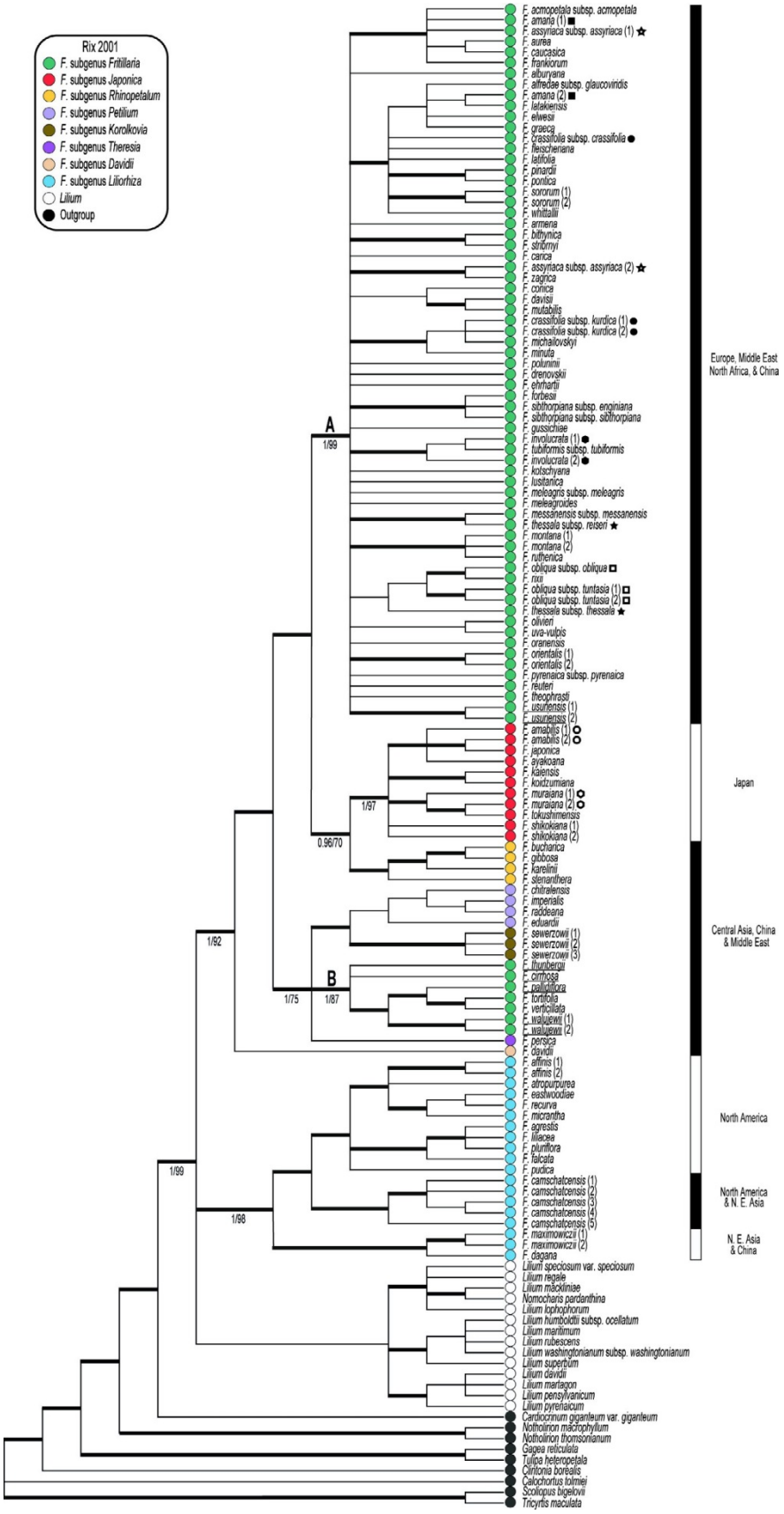
Rix 2001
F. subgenus Fritillaria
F. subgenus Japonica
F. subgenus Rhinopetalum
F. subgenus Petilium
F. subgenus Korolkovia
F. subgenus Theresia
F. subgenus Davidii
F. subgenus Liliorhiza
Lilium
Outgroup
F. acmopetala subsp. acmopetala
F. amana (1) ■
F. assyriaca subsp. assyriaca (1) ★
F. aurea
F. caucasica
F. frankiorum
F. alburyana
F. alfredae subsp. glaucoviridis
F. amana (2) ■
F. latakiensis
F. elwesii
F. graeca
F. crassifolia subsp. crassifolia ●
F. fleischeriana
F. latifolia
F. pinardii
F. pontica
F. sororum (1)
F. sororum (2)
F. whittallii
F. armena
F. bithynica
F. striblnyi
F. carica
F. assyriaca subsp. assyriaca (2) ★
F. zagrica
F. conica
F. davisii
F. mutabilis
F. crassifolia subsp. kurdica (1) ●
F. crassifolia subsp. kurdica (2) ●
F. michailovskyi
F. minuta
F. polunnii
F. drenovskii
F. ehrhartii
F. forbesii
F. sibthorpiana subsp. enginiana
F. sibthorpiana subsp. sibthorpiana
F. gussichiae
F. involucrata (1) ●
F. tubiformis subsp. tubiformis
F. involucrata (2) ●
F. kotschyana
F. lusitanica
F. meleagris subsp. meleagris
F. meleagroides
F. messanensis subsp. messanensis
F. thessala subsp. reiseri ★
F. montana (1)
F. montana (2)
F. ruthenica
F. obliqua subsp. obliqua ◘
F. nuri
F. obliqua subsp. tuntasia (1) ◘
F. obliqua subsp. tuntasia (2) ◘
F. thessala subsp. thessala ★
F. olivieri
F. uva-vulpis
F. oranensis
F. orientalis (1)
F. orientalis (2)
F. pyrenaica subsp. pyrenaica
F. reuteri
F. theophrasti
F. usuriensis (1)
F. usuriensis (2)
F. amabilis (1) ○○
F. amabilis (2) ○○
F. japonica
F. ayakoana
F. kiaensis
F. koidzumiana
F. muraiana (1) ○○
F. muraiana (2) ○○
F. tokushimensis
F. shikokiana (1)
F. shikokiana (2)
F. bucharica
F. gibbosa
F. karelinii
F. stenanthera
F. chitralensis
F. imperialis
F. raddeana
F. eduardii
F. sewerzowii (1)
F. sewerzowii (2)
F. sewerzowii (3)
F. thunbergii
F. cirrhosa
F. pallidiflora
F. tortifolia
F. verticillata
F. walujewii (1)
F. walujewii (2)
F. persica
F. davidii
F. affinis (1)
F. affinis (2)
F. atropurpurea
F. eastwoodiae
F. recurva
F. micrantha
F. agrestis
F. liliacea
F. pluriflora
F. falcata
F. pudica
F. camschatcensis (1)
F. camschatcensis (2)
F. camschatcensis (3)
F. camschatcensis (4)
F. camschatcensis (5)
F. maximowiczii (1)
F. maximowiczii (2)
F. dagana
Lilium speciosum var. speciosum
Lilium regale
Lilium mackliniae
Nomocharis pardanthina
Lilium lophophorum
Lilium humboldtii subsp. ocellatum
Lilium maritimum
Lilium rubescens
Lilium washingtonianum subsp. washingtonianum
Lilium superbum
Lilium davidii
Lilium martagon
Lilium pensylvanicum
Lilium pyrenaicum
Cardiocrinum giganteum var. giganteum
Notholirion macrophyllum
Notholirion thomsonianum
Gagea reticulata
Tulipa heteropetala
Clintonia borealis
Calochortus tolmiei
Scoliopus bigelovii
Thicyrtis maculata
A
1/99
B
1/87
1/97
0.96/70
1/92
1/75
1/99
1/98
Europe, Middle East
North Africa, & China
Japan
Central Asia, China
& Middle East
North America
North America
& N. E. Asia
N. E. Asia
& China

Figure 2.1 - Ninety-five percent majority rule consensus of 18,000 post burn-in trees from the Bayesian analysis of the combined plastid dataset. Thick internal branches indicate nodes with posterior probabilities (PP) of ≥ 0.95 and bootstrap percentages (BP) of ≥ 70 in the parsimony analysis; thin internal branches indicate nodes with PP ≥ 0.95 but < 70 BP. Actual support values are shown below branches (PP/BP) for key clades discussed in the text. Different circle types indicate taxonomic groups (*Fritillaria* subgenera, *sensu* Rix, 2001, *Lilium* and outgroup; see top left-hand corner for key); bold uppercase letters (A and B) mark the two clades recovered for subgenus *Fritillaria*. Names of species used in traditional Chinese medicine are underlined. Symbols following taxon names indicate individuals from species resolved as non-monophyletic. Alternating black and white bars denote broad geographic distributions for species of *Fritillaria* (note that the colours do not indicate groups with similar distributions).

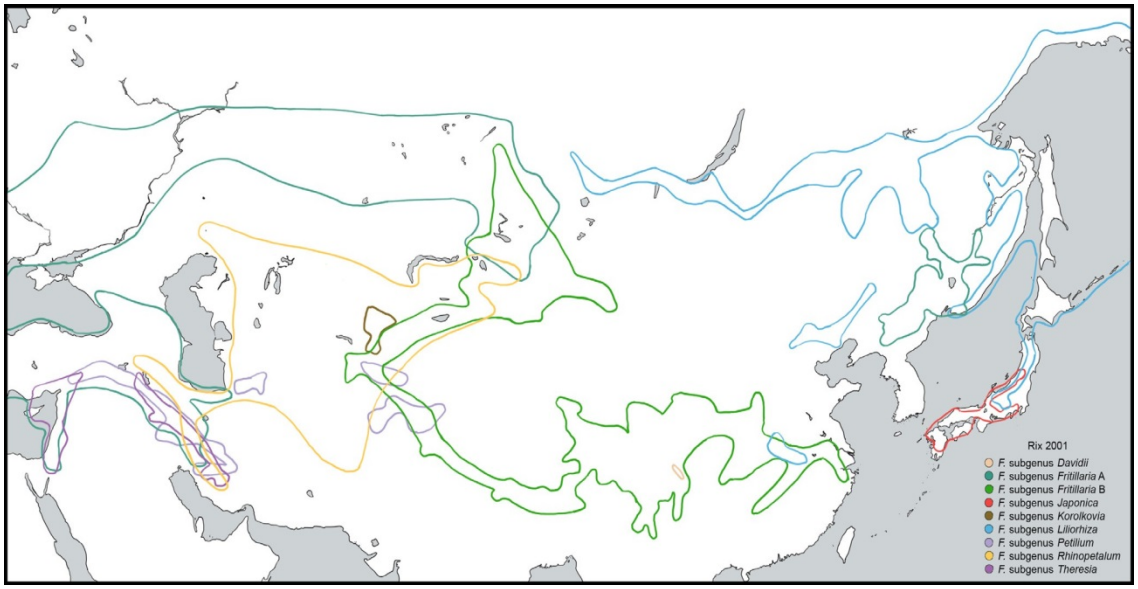

Figure 2.2 - Distribution of *Fritillaria* in Asia. A different line colour is used for each of the subgenera (see bottom right-hand corner for key); the two subgenus *Fritillaria* clades (A and B) are shown separately. Distributions are based on location data compiled from a comprehensive literature review (including journal articles, floras and monographs), from data extracted from specimens held in over 60 herbaria, from the field notes of LH and from personal communications to LH from numerous biologists and horticulturalists with knowledge of natural *Fritillaria* populations. Information from all species of *Fritillaria* was included for the purposes of drawing the distribution map, rather than from just those species sampled in the current study. For *F.* subgenus *Fritillaria*, species belonging to series 7–11 (*sensu* Rix, 2001) were included within the clade B distribution as all members of these series sampled to date (excepting *F. usuriensis*) have been shown to belong to *F.* subgenus *Fritillaria* clade B. Species from all other *Fritillaria* series, plus *F. usuriensis*, were included within the clade A distribution.

2.6.2 LCNG screen

None of the 11 pairs of LCNG primers amplified the target region across all seven species (see Supplementary Table S1). Most primer pairs generated products in some

species, but in the majority of cases these included multiple bands of different lengths and/or strong smears indicative of non-specific amplification (Supplementary Table S5). The most successful primers were those for the ALKALINE PHYTASE (AP) and MALATE SYNTHASE (MS) genes. For AP, single banded products were produced for six species (with no amplification in *F. armena*; Supplementary Table S5). Sequencing of the AP PCR products confirmed that they come from the target region (on the basis of top BLASTn hits to sequences in GenBank; E-value for best hit = 0.0), but also revealed polymorphisms in some species, suggesting the possible presence of multiple alleles or paralogues. For MS, five species yielded a primary band of *c*. 900 bp with a weaker secondary band at *c*. 600 bp. Products from two species (*F. eduardii* and *F. stenanthera*) were sufficiently strong to allow sequencing; BLASTn searches against GenBank revealed that the sequence from the larger amplification product matched the target region (E-value for best hit = $7e^{-62}$). Electropherograms from MS included many polymorphic positions, indicating the presence of multiple sequence variants of the target region and/or interference from the smaller amplification product.

2.7 Discussion

2.7.1 Relationships between *Fritillaria* subgenera

In common with the findings of Rønsted *et al.* (2005), phylogenetic trees inferred from plastid sequence data do not resolve *Fritillaria* as monophyletic. Instead, two strongly supported clades of *Fritillaria* species are recovered (one corresponding to subgenus *Liliorhiza* and the other containing species from the remaining seven subgenera), which are unresolved with respect to each other and to *Lilium* in the Bayesian 95% majority rule consensus tree (Figure 2.1). As also found by Rønsted *et al.* (2005), parsimony analysis of the combined plastid dataset recovered a paraphyletic *Fritillaria*, with the subgenus *Liliorhiza* clade sister to *Lilium* and the remaining *Fritillaria* species (Supplementary Figure S2), but with very low bootstrap support (53%). On the basis of a combined analysis of plastid and nrITS data, Rønsted *et al.* (2005) found strong support for the monophyly of *Fritillaria* and the sister relationship between *Fritillaria* and *Lilium*. However, they also reported evidence for incongruence between the plastid and ITS data, indicated by a decrease in support and resolution for some groups compared with that seen in trees from separate analyses of plastid and nuclear data (Rønsted *et al.*, 2005). Rønsted *et al.* (2005) stated that there was no obvious

explanation for the observed incongruence, such as hybridisation, and suggested that sampling of additional taxa and gene regions (both plastid and nuclear) could help to clarify the underlying cause.

Although the sequencing of an additional region of the plastid genome and inclusion of more than 50 additional species in the current study did not materially change the result based on plastid data compared with that obtained by Rønsted *et al.* (2005), sequencing of the nrITS region in additional species revealed a potential cause for the previously observed incongruence. For many of the species sampled in the current study we found evidence for multiple divergent ITS copies (see section 2.3.4). These findings are indicative of incomplete concerted evolution, which could lead to the comparison of paralogous sequences and confound the inference of species relationships (Álvarez and Wendel, 2003). Incomplete concerted evolution of nrDNA has been reported in an insect species with a very large genome, *Podisma pedestris*, and it is suggested that the inefficiency of concerted evolution in this species could be a direct consequence of its large genome size (Keller *et al*., 2006). Analysis of karyotype evolution in *Lilium* species has demonstrated they can have large numbers of loci for the 18S subunit of ribosomal DNA (P6), localised on several different chromosomes (Muratović *et al*., 2010b). Such a distribution could also promote the accumulation of intragenomic ITS variation as homogenisation may occur more readily between repeats on the same chromosome than between those on separate chromosomes (Parkin and Butlin, 2004).

As an alternative to the use of the nrITS region, we attempted to develop low-copy nuclear gene regions. However, none of the regions examined could be amplified from all taxa tested (see section 2.6.2), which may be due, at least in part, to the extremely large genomes of *Fritillaria* and *Lilium* species. Genome sizes (1C-values) exceeding 85 Gb have been reported in *Fritillaria* (Leitch *et al*., 2007; Ambrožová *et al*., 2011), and up to *c*. 48 Gb in *Lilium* (Bennett and Leitch, 2012). The chances of primers annealing to a region that has a single copy within such a vast genome is likely to be much lower than for repetitive regions such as the nrITS, which can have thousands of copies per genome (Álvarez and Wendel, 2003). Indeed, it has been suggested that a very large genome size contributed to difficulties in developing low-copy nuclear gene regions in the Sandhills lily, *Lilium pyrophilum* (Douglas *et al*., 2011). Efforts to develop LCNG regions for *Fritillaria* are ongoing (Kelly *et al*., unpublished results), and future phylogenetic analyses of such loci will be important in helping to resolve the outstanding question of the relationship between subgenus *Liliorhiza* and the remaining

subgenera of *Fritillaria*. For *Lilium*, sampling of the plastid genome for further species (14 species are included in the current work, representing *c*.12% of the genus and encompassing the seven *Lilium* sections *sensu* Comber (1949) plus *Nomocharis*), as well as the addition of nuclear data, might further clarify the *Fritillaria*/*Lilium* relationship. A number of phylogenetic studies focused on *Lilium* have been published recently (*e.g.* İkinci, 2011; Gao *et al*., 2012a,b, 2013). Unfortunately, these fail to throw any further light on the relationship between the two main *Fritillaria* clades and *Lilium* as they either include only a single *Fritillaria* species (İkinci, 2011), do not represent both of the major *Fritillaria* clades (Gao *et al*., 2012a, 2013) or infer relationships only from uncloned ITS sequences (Gao *et al.*, 2012b) which should be treated with caution in light of the issues discussed above.

Although the results are based solely on the analysis of plastid sequences and await confirmation with data from independently evolving regions of the nuclear genome, they do provide further evidence for relationships between the subgenera of *Fritillaria* compared with those of the earlier study by Rønsted *et al.* (2005). The much-increased species-level sampling of *Fritillaria* provides strong support for the polyphyly of subgenus *Fritillaria* (see below, section 2.7.2). The placement of subgenus *Japonica* is also of particular interest, as this group contains the largest recorded genome for any diploid plant species (Ambrožová *et al*., 2011). In previous phylogenetic analyses, subgenus *Japonica* was represented by a single species and, on the basis of a combined plastid dataset, was unresolved within a clade containing members of subgenus *Fritillaria* and *Rhinopetalum* (Rønsted *et al*., 2005). All eight currently recognised species of subgenus *Japonica* (Hill, 2011) have now been sampled and this, together with increased sampling of subgenus *Fritillaria* and *Rhinopetalum*, has provided strong support for the monophyly of this group of endemic Japanese species. Moreover, the strongly supported sister group relationship between subgenera *Japonica* and *Rhinopetalum* has important implications for understanding genome evolution within *Fritillaria*. These two subgenera have strikingly divergent genome sizes, with 1C values in the Japanese species almost double those found in members of subgenus *Rhinopetalum* (Leitch *et al*., 2007; Ambrožová *et al.*, 2011). Such a dramatic shift in the amount of nuclear DNA between very closely related lineages may indicate differences in the underlying mechanisms that control the amplification and removal of repetitive DNA. The availability of an expanded phylogenetic hypothesis of species

relationships in *Fritillaria* will allow the direction and rate of genome size change to be inferred and provides the evolutionary context for ongoing comparative genomic studies (Kelly and Leitch, 2011; Kelly *et al*., 2012; Kelly *et al*., unpublished results).

2.7.2 Polyphyly of subgenus *Fritillaria* and phylogenetic distribution of species used in Traditional Chinese Medicine (TCM)

We find clear evidence that subgenus *Fritillaria*, as currently circumscribed, is polyphyletic. Species of subgenus *Fritillaria* are split into two distinct groups, revealing a small but well supported clade of mainly Chinese and Central Asian species (subgenus *Fritillaria* B), not closely related to the members of subgenus *Fritillaria* found within the large predominantly European, Middle Eastern and North African clade (subgenus *Fritillaria* A; Figure 2.1). *Fritillaria* subgenus *Fritillaria* B groups with subgenera *Korolkovia*, *Petilium* and *Theresia*, all of which are distributed in the Middle East and Central Asia, in a strongly supported clade (Figure 2.1). The potential polyphyly of subgenus *Fritillaria* was highlighted by Rønsted *et al*. (2005), who found that a single species from this group (*F. pallidiflora*) was resolved within the *Korolkovia/Petilium/Theresia* clade. It was suggested that *F. pallidiflora* might either be named as a new subgenus or included in an expanded subgenus *Petilium* (Rønsted *et al*., 2005). Our new results demonstrate that at least six species currently classified within subgenus *Fritillaria* by Rix (2001) are actually more closely related to members of subgenera *Korolkovia*, *Petilium* and *Theresia*. Sequencing of the *c*. 50 *Fritillaria* species yet to be analysed (most of which have been classified within subgenus *Fritillaria* (Rix, 2001)) may uncover further members of this novel clade.

The conflict between our results and the current subgeneric classification of *Fritillaria* indicates that the morphological character states used to group species into subgenus *Fritillaria* (Rix, 2001) are likely to have evolved independently in different lineages. Convergent or parallel evolution of phenotypic traits may be a common cause of incongruence between morphology-based classifications and the results of molecular phylogenetic analyses. For example, in *Echinopsis* (Cactaceae) there has been extensive convergent evolution of floral traits, which relate to repeated switches in pollinators (Schlumpberger and Renner, 2012). In *Rheum* (Polygonaceae), specialised vegetative characters have evolved in parallel in species that inhabit arid environments (Sun *et al*., 2010). The use of such evolutionarily labile characters in taxonomic classifications can

lead species to be subdivided in a way that does not accurately reflect their evolutionary relationships (Schlumpberger and Renner, 2012).

The polyphyly of subgenus *Fritillaria* is of particular interest in terms of its implications for our understanding of evolutionary relationships between the 16 species used in TCM, as 15 of them belong to this subgenus (Rix, 2001) with the remaining species (*F. anhuiensis*) placed in subgenus *Liliorhiza* (Rix, 2001). In the current work, five TCM species (all subgenus *Fritillaria sensu* Rix (2001)) were analysed and shown to be split between the two subgenus *Fritillaria* clades recovered in the plastid tree; *F. usuriensis* is placed in subgenus *Fritillaria* A, whereas *F. cirrhosa*, *F. pallidiflora*, *F. thunbergii* and *F. walujewii* fall within subgenus *Fritillaria* B (Figure 2.1). Over-harvesting, habitat fragmentation, over-grazing and an expanding international herbal market (Zhang *et al.*, 2010; Konchar *et al.*, 2011) have placed great pressure on wild *Fritillaria* populations, and there is a need for conservation of critically endangered species, such as those in western China (Hao *et al.*, 2013). For example, *F. cirrhosa*, which is from southwest China and the eastern Himalayas of Bhutan and Nepal (Zhang *et al.*, 2010) and commonly used in TCM for treating coughs, is one of the most heavily collected Chinese alpine medicinal plants and is facing extinction in the wild (Buntaine *et al.*, 2007; Zhang *et al.*, 2010). The discovery that some of the most important species used in TCM are closely related to widely cultivated members of subgenus *Petilium* and *Theresia*, raises the possibility of substituting more commonly commercially available bulbs for the rare species that are currently collected directly from the wild. Recent large-scale analyses have demonstrated that species used for traditional medicines in independent cultures are significantly clustered in phylogenetic trees (Saslis-Lagoudakis *et al.*, 2012). These findings support the view that traditional knowledge can be effective in identifying plants with bioactive compounds, but also highlight the fact that clades within phylogenetic trees that are rich in species used in traditional medicines are potential targets for bioprospecting for novel medicinal species (Saslis-Lagoudakis *et al.*, 2012). The results demonstrate that a number of species important for TCM are clustered within the *Korolkovia*/*Petilium*/*Theresia* clade; future testing of species such as *F. imperialis* and *F. eduardii*, which are being commercially bred (Wietsma and van den Berg, 2012), could establish whether these bulbs contain the same bioactive compounds that have been identified from some of the Chinese species. The ability to substitute commercial varieties suitable for cultivation

for native Chinese species has the potential to reduce the burden on wild populations by alleviating collecting pressure on traditional bulb sources.

2.7.3 Evidence for species non-monophyly

Several species represented by multiple individuals in this study are non-monophyletic on the basis of data from the plastid genome. A variety of different processes can result in species appearing as paraphyletic or polyphyletic in gene trees, such as hybridisation/introgression, incomplete lineage sorting and inaccurate circumscription of species limits (Funk and Omland, 2003). All these factors may have contributed to the patterns of species non-monophyly observed for *Fritillaria*. For example, there are several reports of natural interspecific hybrids (reviewed by Hill (2011)), and a number of *Fritillaria* species are morphologically heterogeneous (*e.g. F. montana* (Bartolucci *et al.*, 2009) and *F. michailovskyi* (Türktaş *et al.*, 2012)) which may promote the splitting of different variants into separate species (Funk and Omland, 2003). Some cases of species non-monophyly in *Fritillaria* relate to the placement of sequences from different subspecies into separate strongly supported clades. The most obvious examples are in *F. crassifolia* and *F. thessala*, both of which are polyphyletic on the basis of the plastid sequence data (Figure 2.1). For *F. crassifolia*, *F. crassifolia* subsp. *kurdica* was previously recognised as a separate species (see Rix (2000)), raising the possibility that the polyphyly of *F. crassifolia* in the plastid tree could reflect a currently overly broad circumscription of the species. Nevertheless, additional data (*i.e.* from regions of the nuclear genome; see section 2.7.1) are needed to enable evidence for currently unrecognised species to be distinguished from that for processes such as interspecific gene flow. Moreover, only a single individual has been sequenced for the majority of *Fritillaria* species and further testing of species-boundaries will require more extensive intraspecific sampling in addition to sequencing further gene regions.

2.8 Conclusions

Our results provide support for the monophyly of all subgenera recognised by Rix (2001), with the exception of the largest subgenus, *Fritillaria*, which is polyphyletic. The division of subgenus *Fritillaria* into two discrete clades also splits the members of this subgenus that are used in TCM, with the implication that some of the most important medicinal species, such as *F. cirrhosa*, are inferred as being closely related to species that are in widespread cultivation for ornamental purposes. The Japanese

endemic subgenus *Japonica*, which comprises species with extremely large genome sizes, is resolved as sister to subgenus *Rhinopetalum*, a group containing species with genomes approximately half the size of those found in the Japanese species (Leitch *et al*., 2007; Ambrožová *et al.*, 2011). Our findings also highlight the need for further data, from the nuclear genome, in order to resolve outstanding questions including the causes of species non-monophyly. Data from multiple independently inherited loci are needed in order to distinguish evidence for additional, currently unrecognised, species from that for alternative scenarios. However, such data may be particularly hard to obtain in species with genomes that are as extraordinarily large as those in *Fritillaria*.

Chapter 3

The introduction of *Fritillaria* to Europe

3.1 Introduction

By the middle of the 16th century European knowledge of the known world was gradually expanding to encompass the Americas, Africa, India and Asia. The early travellers were usually the explorers, followed by the diplomats, colonists and then the merchants. Until trade facilitated the importation of goods, the notion of 'exotic' plants and the commercial opportunities arising from discovery and onward sale of these 'novelties' were unknown before the early 16th century. However, because the plant trade at that time was a very minor component of normal commerce, acquisition of plants relied upon existing trade routes, and the merchants that imported them brought them either as gifts or for sale (Tigner, 2008; Cumo, 2013). For more than 1,500 years prior to the 16th century, plants of genus *Fritillaria* were known for their properties and economic importance as Traditional Chinese Medicine (TCM) (Chen, 2005; Zhang *et al.,* 2010) (Chapter 2). Species of the genus established outside China with a Eurasian distribution were yet to be described in Europe and the Middle East.

Initial research undertaken for this thesis revealed that no comprehensive account existed of how species of *Fritillaria*, and particularly *F. meleagris*, were first introduced or recorded in Europe. In an attempt to create such an account, fragments of evidence have been assembled through personal correspondence and searching the records, images, archives and correspondence in herbaria, universities, libraries, and on websites dedicated to the early herbalists, botanists and apothecaries.

3.2 Materials and methods

3.2.1 Herbals

In order to understand the origins of interest in plants in Europe, an initial search was undertaken of published works dedicated to the study, history and evolution of printed herbals[3] across Europe and the Middle East. By researching the many publications available (Table 3.1), it was anticipated that a bibliography could be assembled of the most significant and many minor herbal works of the period.

It was the invention of the mechanical printing press *c*. 1439 by Johannes Gutenberg in Strasbourg (Britannica, 2017) that enabled the wider distribution of herbals for use mainly by apothecaries and the public in the fields of medicinal botany and agriculture (Arber, 1986; Blunt and Raphael, 1979). For purposes of current research this meant that now many of these ancient publications were likely to be available either from libraries or electronically on the internet. It was predicted that this approach would (i) provide an efficient method of individually studying their contents in chronological order to reveal the earliest descriptions and images of *Fritillaria* species, and (ii) would inform the direction of further research

3.2.2 Institutions, libraries, online databases and archives

Information gleaned from the publications describing the history and evolution of herbals (Table 3.1) enabled targeted internet and library searches of the resources most likely to hold source material of the most significant herbals (Table 3.2). Online search provided access to library catalogues and contact details of staff with responsibility for supporting researchers of archive material; libraries where archive material was stored were contacted to arrange personal access. For institutions housing associated archives about the most important herbals, personal contact was made with staff to enquire about enabling access to any material not accessible online, but which might be relevant to the research.

[3] described by Singer (1927) as ‘collections of descriptions of plants put together for medical purposes’.

Table 3.1 – Literature studied whilst searching for the earliest published European and Middle Eastern herbals.

Author (s)	Year	Title
Arber A	1986	*Herbals: their origin and evolution*. Cambridge, UK.
Barlow HM	1913	Old English Herbals, 1525 -1640. *Proceedings of the Royal Society of Medicine*, **6**: 108-149.
Blunt WJ, Raphael S	1979	*The illustrated herbal*. London, UK.
Egmond F	2008	Apothecaries as experts and brokers in the sixteenth-century network of Carolus Clusius. In: Dupré S, Kusukawa S (eds) *History of Universities* vol. XXIII/2. Oxford, UK. pp 59-91.
Greene EL	1983	*Landmarks of botanical history*. California, USA.
Henrey B	1975	*British botanical and horticultural literature before 1800*. Oxford, UK.
Marcus MF	1944	The Herbal as art. *Bulletin of the Medical Libraries Association*, **32**: 376-384.
Nissen C	1958	*Herbals of five centuries*. Zurich, Switzerland.
Pächt O	1950	Early Italian nature studies and the early calendar landscape. *Journal of the Warburg and Courtauld Institutes*, **13**: 13-47.
Rohde ES	1974	*The old English herbals*. London, UK.
Singer C	1923	Herbals. *Edinburgh Review*, **237**: 95-112.
Singer C	1927	The herbal in antiquity and its transmission to later ages. *Journal of Hellenic Studies*, **47**: 1-52.
Singer C	1928	*From magic to science: essays on the scientific twilight*. New York, USA.
Sprague TA	1938	The evolution of the herbal. *South Eastern Naturalist*, **43**: 33-39.
Tomasi LT	2001	The study of the natural sciences and botanical and biological illustration in Tuscany under the Medicis from the sixteenth to the eighteenth centuries. *Archives of Natural History*, **28**: 179-193.

Table 3.2 – Major institutions, libraries, herbaria and website resources accessed whilst searching for information about early European and Middle Eastern herbals.

Institution	Access acquired or internet link
Sherardian Library of Plant Taxonomy, Oxford.	Personal visits
Radcliffe Science Library, Oxford.	Personal visit
British Library, London.	Personal visits
Royal Botanic Gardens Library, Art and Archives, Kew.	Personal visits
National Archives, Kew.	http://discovery.nationalarchives.gov.uk
Felix Platter Herbarium, Basel.	https://platter.burgerbib.ch/blattern
Conrad Gessner Herbarium, Erlangen-Nürnberg, Nuremberg.	https://ub.fau.de
Ulisse Aldrovando Archive, Bologna.	http://aldrovandi.dfc.unibo.it
Petrus Cadé Herbarium, Utrecht.	http://www.nationaalherbarium.nl/Cade/index
Caspar Bauhin Herbarium, Basel.	https://herbarium.unibas.ch/index.php
Pietro Antonio Michiel herbarium archive, Venice.	https://marciana.venezia.sbn.it
Nicolas Rasse correspondence, London.	http://blog.wellcomelibrary.org/2015/07/a-parisian-surgical-dynasty-and-their-books
Clusius correspondence, Leiden.	http://clusiuscorrespondence.huygens.knaw.nl/editions/search
Biblioteca Digital, Real Jardín Botánico de Madrid, Madrid.	http://www.rjb.csic.es/jardinbotanico/jardin
Botanicus, Missouri Botanical Garden, St Louis.	http://www.botanicus.org
NYBG Index Herbariorum, New York.	http://sweetgum.nybg.org/science/ih
Österreischische Nationalbibliothek, Vienna.	https://www.onb.ac.at
Linda Hall Library Digital Collections, Kansas City.	http://lhldigital.lindahall.org
Biodiversity Heritage Library.	https://www.biodiversitylibrary.org
Dfg-Viewer internet archive library, Düsseldorf.	http://dfg-viewer.de
Internet Archive, San Francisco.	https://archive.org
Lancashire and Cheshire Entomological Society.	http://www.record-lrc.co.uk

3.2.3 Academics, researchers and authors

Communication with a number of academics, researchers and authors (Table 3.3) enabled an efficient approach to the subject to be achieved. This manifested itself particularly in relation to contacts to enable database access and transcribing and translation support during the research.

Table 3.3 – Personal communications with academics, researchers and authors.

Person	Contribution/institution
Dauwalder, Lea	Felix Platter (1536 -1614) archive in Zurich.
Egmond, Dr Florike	Institutions holding 16th century herbal archive material.
Funk, Dr Holger	Translation of the Penny letter and watercolour annotation.
Kohlmann, Sigrid	Archives at University Library Erlangen-Nürnberg (FAU).
Kusukawa, Sachiko	'Naturalia' and 16th century botanical art.
Lienhard, Luc	Felix Platter archive in Zurich.
Lugato, Elizabetta	Archives at Biblioteca Nazionale Marciana (BNM) in Venice.
Mason, Dr Peter	Translation of the Penny letter and watercolour annotation.
Steinmann, Martin	Translation of the Penny letter and watercolour annotation.

3.2.4 Searching for *Fritillaria* from the late 15th century

It is generally accepted that the earliest forms of printed matter to which the term 'herbal' is commonly applied appeared in the late 15th century (Arber, 1986). These were essentially medical recipe books and came from Italy – the *Herbarium of Apulei Platonici* of 1484 in Latin, and from Germany – the *Latin Herbarius*, also from 1484, printed in Mainz (Arber, 1986). The first herbal printed in English was an untitled work by Bartholomeus Anglicus in 1495 which contained the first botanical illustration in an English book (Rohde, 1974). The early 16th century witnessed the printing of many herbals (Rohde, 1974), emanating mainly from Italy and Germany and usually without illustration or containing only a small number of botanical woodcuts usually of inferior quality (pers. obs.). However, many included indexes of plant names, usually in Latin, Greek or German (pers. obs.). In consequence, the presence of these could now facilitate searches to be conducted for the presence of '*Fritillaria'* (and '*Flos*', '*Lilionarcissus*', '*Lilium*', '*Meleagris*' and '*Narcissus*' by which it was also known) within European and Middle Eastern herbals from the late 15th century onwards.

3.3 Results

3.3.1 Searching for *Fritillaria*

Online access provided the facility to search efficiently many herbals unavailable elsewhere. The process enabled a broad understanding of the content of these early herbals to be acquired. All the generally recognised major herbals and many minor ones (Nissen, 1958; Singer, 1927; Arber, 1986; Rohde, 1974; Blunt and Raphael, 1979) (Table 3.4) were researched either online or, where the publication was not available, studied personally at one of the libraries or archives listed in Table 3.2. No evidence of Middle Eastern herbals or their history could be found for the period under study.

After an exhaustive search, it is believed there is no European herbal published between 1484 and 1578 and available online, at a library or archive that has not personally been consulted.

A history of human interest in plants both wild and cultivated was also acquired as an important by-product of these searches.

Table 3.4 – Major herbals studied in determining first mention of *Fritillaria* species.

Author	Pub. Date	Title
Banckes, Richard.	1525 - 1555	*Herbal* – many editions. London.
Brunfels, Otto von.	1532	*Herbarum*. Strasbourg.
Turner, William.	1538 - 1568	*The names of herbes….*, several editions. London.
Gessner, Conrad.	1539	*Historia plantarum et vires…..* Paris.
Fuchs, Leonhart.	1542	*De historia stirpium…..* Basel.
Bock, Hieronymus.	1546	*Kreuter Bűch*. Strasbourg.
Dodoens, Rembert.	1554 - 1557	*Kruydeboeck* (*Histoire des plantes*). Antwerp/Anvers.
Anguillara, Luigi.	1561	*Semplici dell' eccellente…..* Florence.
Cordus, Valerius.	1561	*In hoc volumine continentur…..* Strasbourg.
Mattioli, Pierandrea.	1563	*Neuw Kreüterbuch*. Rome.
Dodonaeus, Rembert.	1568	*Florum, et coronariarum…..* Antwerp.
De l'Obel, Matthias and Pena, Petrus.	1570 - 1571	*Stirpium adversaria nova…..* London.
Lyte, Henry.	1578	*A niewe herbal (or history of plantes)*. London.

3.3.2 First descriptions and depictions of *Fritillaria*

Early herbals were dedicated to the description and depiction of *materia medica*, the study of medicinal plants. It was only during the 16th century that the study of botany became of interest, which then widened the subject matter of herbals to all plants, whether medicinally relevant or not (Egmond, 2008). However, it was not until the mid-16th century that the great herbals arrived containing many woodcuts and descriptions of plants, from which identification of wild and cultivated plants could be reasonably achieved (pers. obs.).

The first reference found for *Fritillaria* was that of *Crown Imperial* (now *Fritillaria imperialis*), originally known in Europe by its Persian appellation *Tusac* (Beckmann, 1823). It first appeared in Italy in 1553 as *Hiacintho di Homero* (named after the Greek Homer) (Saccardo, 1909) and described by Anguillara (1561). Linnaeus (1753) describes the species as originating in Persia (modern day Iran). This was one of the earliest bulbs in cultivation to be recorded in Europe, and was most probably brought from Constantinople (modern day Istanbul) by Ogier Ghiselin de Busbecq (1522-1592). He was the Austrian imperial ambassador of Emperor Ferdinand I in Vienna to the court of Sultan Suleiman I in Turkey between 1555-1562 (Forster and Daniell, 1884; Egmond, 2013). Ogier Ghiselin de Busbecq had a great interest in plants and travelled much between the two countries during his term of office (Stone, 2015).

Having returned to Vienna from Constantinople in 1562, de Busbecq is recorded as having received specimens of *Crown Imperial* from Constantinople in 1572, amongst

many other rare and novel plants, and forwarding them to Carolus Clusius (Clusius or Charles de L'Écluse (1526-1609)). He was a Flemish botanist and Director of the Imperial Medicinal Garden of Emperor Maximilian II in Vienna from 1573-1576 (Forster and Daniell, 1884; Park and Daston, 2003; Egmond, 2010). In Europe, the plant was first recorded in bloom in Vienna in April 1576 by Rembert Dodoens (1517-1585) from Southern Netherlands (Dodoens, 1583). Dodoens was Physician to Maximilian II, with whom Clusius worked alongside in Vienna (Dodoens, 1583; Willes, 2011). In Europe *Crown Imperial* became known by the scientific name *Lilium persicum* (Dodoens, 1583). However, on recognising that there was already another species of the same name already to be found in European Gardens (see below), was renamed *Corona imperialis* by Alfonso Pancio (1530-1610) (Parkinson, 1629). From about 1550-1574 Pancio was the private physician to Duke Alfonso d'Este of Ferrara in Italy (Egmond, 2010). From Vienna these horticultural novelties were sent initially to Italy, then Southern Netherlands, with subsequent redistribution to the imperial courts of Europe (Mason, 2007; Egmond, 2010). The species was renamed *F. imperialis* by Linnaeus (1753).

Linnaeus (1753) describes the bulbs of *Lilium persicum* as originating in the city of Susa in Persia. The plant was originally known as *Susam giul* (Clusius, 1601) – '*giul*' probably a corruption of the Turkish 'gül', meaning cultivated (pers. obs.). It was brought to Italy in 1576 through Constantinople, probably by the same route as that of *F. imperialis* and facilitated by the London merchant Master Nicholas Lete (?-1602). Lete worked for the Levant Company (Potter, 2006) and was 'a worthy merchant and a lover of all faire flowers from Constantinople' (Parkinson, 1629) and had business interests in Persia and Syria (Saccardo, 1909; Parkinson, 1629). The species was renamed *Lilium susianum* (Persian lily) by Clusius (1601). It was subsequently renamed *F. persica* by Linnaeus (1753).

De Toni (1910) referred to two plant watercolours he had found in the Venice herbarium archive of the nobleman Pietro Antonio Michiel (1510 – 1576). Michiel had created an experimental and botanical garden in the middle of the century on the island of Trovaso in Venice (Pavord, 2005). De Toni (1910) doubted that these images (of about 1566 – the artist was probably Domenico Dalle Greche, Michiel's principal commissioned painter (Egmond, 2017)) were of *Fritillaria* plants, but concluded they were probably originally destined for Michiel's unpublished work '*Erbario o istoria*

generale della piante' ('Herbarium' or 'General history of plants') (Fratti and Segarizzi, 1909).

On personally requesting access to the two watercolour images from the Biblioteca Nazionale Marciana (BNM) in Venice (acc. *Giallo* [yellow book] nos. 82, 83), it was ascertained they had never been published (S. Marcon, A. Moro, pers. comm.), and for reasons of conservation were not to be copied. However, payment sent for photographic images to be made obtained the appropriate permissions and resolved the situation.

Figure 3.1 - '*Hiacinto de homero*' *Giallo 82. BNM.*

Figure 3.2 - '*Lilium susianum*' *Giallo 83. BNM.*

On receipt of the images it was discovered from their archive labelling that they had been identified by BNM as *Hiacinto de homero* (*F. imperialis*) and *Lilium susianum* (*F. persica*) respectively (Figures 3.1, 3.2), and were both designated 'Reproduzione vietato' – reproduction prohibited. These are almost certainly the earliest known images of these species (pers. obs.) and presumably therefore have never before been seen outside BNM.

3.3.3 The tulip

The tulip was another of the earliest bulbs to be brought to Europe as a horticultural novelty, probably by Pierre Belon (1517-1564) the writer, trader and French explorer to the Levant between 1546 and 1549 (Pavord, 1999). The plant was first painted and named in Europe by Conrad Gessner (1516-1565) (Figure 3.3), a Zurich botanist, zoologist and doctor of medicine (Egmond, 2008). Gessner had seen a specimen in flower in April 1559 in the garden of Johann Heinrich Herwart, an Augsburg (Bavaria) magistrate, but he states that he had already been sent and seen a drawing of this same tulip by Johannes Kentmann (1518-1574), a German physician and naturalist who had been in Italy (Padua and

Figure 3.3 - '*Tulipa turcarum*' Conrad Gessner, *c.* 1559. Folio v220/FAU.

Figure 3.4 - '*Tulipa turcarum*' woodblock print. Conrad Gessner (1561).

Bologna) between 1549 and 1551 (Pavord, 1999). Gessner called it *Tulipa turcarum* ('Turkish tulip') (now thought by modern botanists to be similar to *Tulipa suaveolens* (Christenhusz *et al*., 2013)), and described it with an accompanying woodblock print in 1561 (Gessner, 1561) – thought to be the first European published illustration. As was typical of the time, the original sketch was drawn as originally seen, with the result that the printed image is seen in reverse (Figure 3.4) (Pavord, 1999).

3.3.4 *Fritillaria meleagris* in the mid to late 16th century

In correspondence of 1571 (now held at Leiden University Library (VUL)) sent from Noël Capperon (?-1572), a protestant apothecary in the French city of Orléans, to Clusius in Malines (now Mechelen) in Belgium, reference is made to Clusius' request for specimens to be sent of a plant found by Capperon in the late 1560s - which Clusius called *'Lilium or bulbus variegatum'* (Egmond, 2008; Griffiths, 2015) (Figure 3.5). In the letter, Capperon describes the plant growing wild and abundant in the water meadows along the River Loire around Orléans and Lyons, naming it *Fritillaria* (Figure 3.6), referring to the chess or draughtboard pattern on the flowers as similar to the design of a dice-box (L. *fritillus*) (Capperon, 1571). To date, this has been accepted as the first time the plant had been described, and Capperon as the first to record it. In 1574 Rembert Dodoens drew the first known image of the plant to date (Figure 3.7) and gave the epithet *meleagris* to the species, because the marks on the flowers suggested the variegation found on the feathers of the guineafowl (*Numida meleagris*) (Dodoens, 1574).

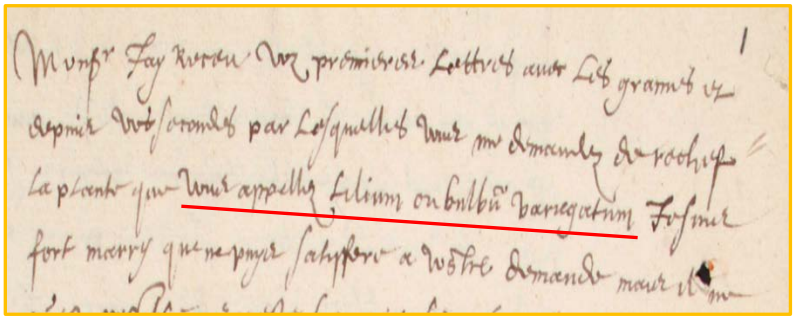

Figure 3.5 - Fragment of letter from Capperon in Orléans to Clusius in Malines, 12/12/1571, referring to Clusius' request for specimens of '*Lilium or bulbus variegatum*' – underlined in red; VUL101.

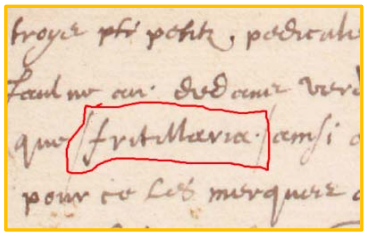

Figure 3.6 - '*Fritillaria*', named by Capperon in his 12/12/1571 letter to Clusius; VUL101.

It is often reported that Capperon first introduced the plant to England in 1572, where it was also known as ‘Caperon’s narcissus’ (Egmond, 2010). However, it is also recorded that Capperon was among some 500 Huguenots massacred by the Catholics during the St Bartholomew’s Day tide on 25th August 1572 (Dodoens, 1583; Raven, 1947), therefore leaving doubt as to the part he played in a possible English introduction. The more likely path of introduction was through any number of merchants now regularly travelling between France and Holland in the late 16th and early 17th centuries, who might have brought bulbs to England for their commercial value to the grand estates of the landed gentry of the time (Thick, 1990).

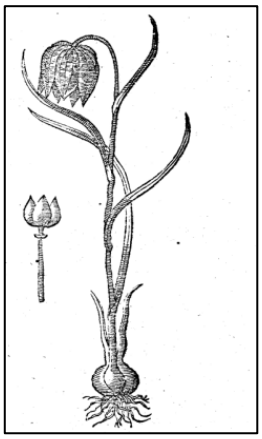

Figure 3.7 - *F. meleagris*. Dodoens (1574).

At sometime between 1567 and 1568 Thomas Penny (1530-1589), a Lancastrian physician, entomologist and botanist (Potts and Fear, 2000; Egmond, 2010), travelled to Orléans to work in association with Capperon on ‘*Phytognomices studio*’ (plant physiology) (Raven, 1947). He is known then to have travelled to Zurich to help annotate Gessner’s collection of watercolours in anticipation of the proposed publication of Gessner’s *Historia Plantarum* (Wotton *et al*., 1634; Raven, 1947; Egmond, 2017).

Among the literary estate archives of Conrad Gessner held at the Friedrich Alexander Universität Erlangen-Nürnberg (FAU), a letter has personally been discovered written in Latin and initialled by Thomas Penny (‘T.P.’) (Figure 3.8) describing a plant, ‘*Bulbus variegatus*’ found by ‘*Natalis Caperonus*’ (Noël Capperon). (A personal request made to S Kohlmann (from the Department of Manuscripts, Rare Books, Graphic Collection, University Library

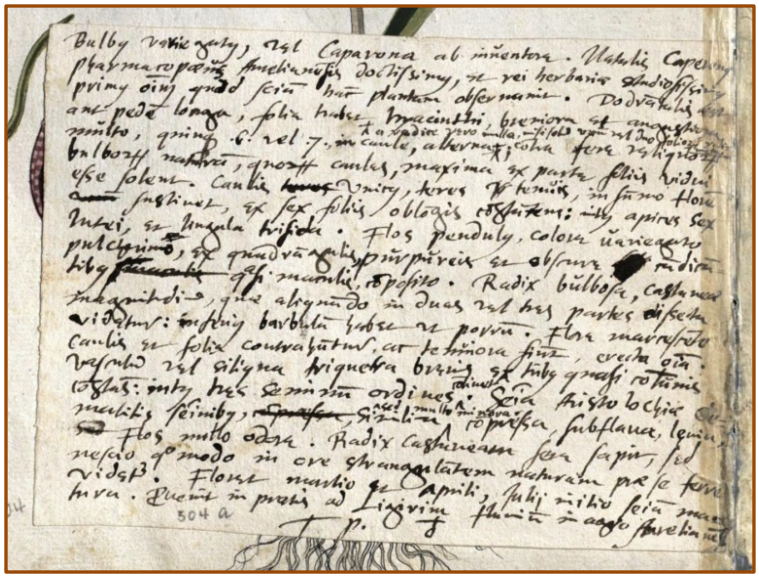

Figure 3.8 - Letter written by Thomas Penny (*c*.1568) discovered bound in with a watercolour of ‘*Narcissus variegatus*’ (*F. meleagris*). From the Conrad Gessner archive; v482/FAU.

Erlangen-Nürnberg (FAU)) for comparison of the letter with known handwriting of Thomas Penny, confirmed they were both from the same hand). The letter was noticed as a piece bound to the side of a personally requested photograph of a discovered watercolour image of '*Narcissus variegatus*' (Gessner archive: v482/FAU) (Figure 3.9) by an unknown artist (but probably by Johannes Kentmann (1518-1574) - contracted by Gessner to execute many of the watercolours he required during preparation of his publication *Historia Plantarum* (M. Steinmann, pers. comm.; Egmond, 2017)). The collection of plant images was finally published as *Opera Botanica* in two volumes long after Gessner's death (Gessner, 1751, 1771).

It is believed the letter describes, for the first time, the structure, colouration, markings, root shape, months of flowering and seed maturity of the plant now known as *F. meleagris* (pers. obs.). Figure 3.11 (A) and (B) respectively provides a Latin transcription of the Penny letter and translation into English.

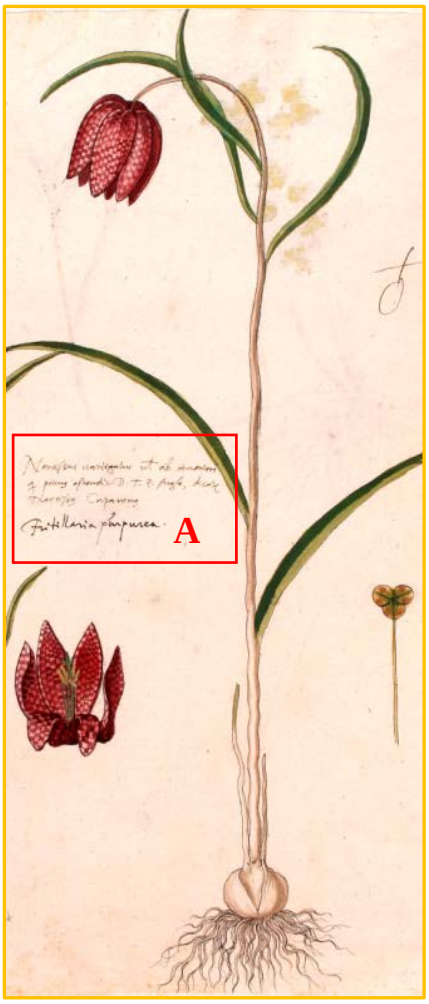

Figure 3.9 - '*Narcissus variegatus*' (*F. meleagris*) *c.*1567, by an unknown artist - from the Conrad Gessner archive. Folio v482/FAU; annotation 'A' highlighted.

An accompanying annotation was discovered on the watercolour (Figure 3.10 - labelled 'A' on Figure 3.9), the first part of which names the plant *Narcissus variegatus*, a name believed unknown hitherto (written probably by Caspar Wolf (1525-1601), but date unknown – S. Kohlmann, pers. comm.), followed by an additional name *Narcissus Caparonus* - interpreted to be an alternative name for the plant (pers. obs.). An addition further below in another and later hand (written probably by Jean Bauhin (1541-1613), date unknown - S. Kohlmann, pers. comm.), gives the name *Fritillaria pupurea* (Figure 3.10 - labelled 'A' on Figure 3.9), a name also believed unknown hitherto (however, used subsequently and erroneously as a synonym for *Calochortus purpureus* by Carl Sigismund Kunth in 1815 (von Humboldt *et al*. (1815). The annotation refers to Mr Thomas Penny 'D. T. P.'. A translation of the annotation

reads: '*Narcissus variegatus* as it is called by the discoverer [*Capperon is assumed*] who first showed it to the Englishman D. T. P.'.

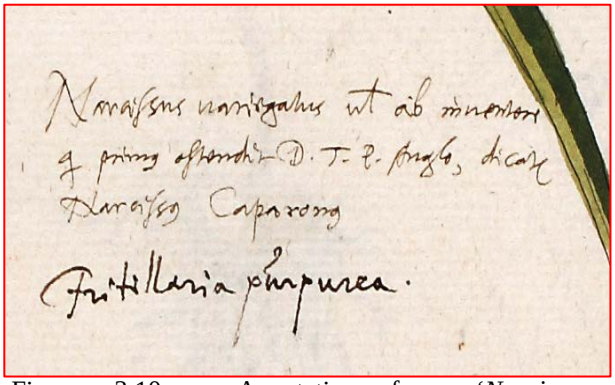

Figure 3.10 - Annotation from '*Narcissus variegatus*', labelled 'A' on Figure 3.7, from the Conrad Gessner archive. Folio v482/FAU.

The Penny letter attached to the watercolour was almost certainly written in Zurich in 1568 (pers.obs.) following Penny's return there after his stay with Capperon in Orléans between 1567 and 1568 (Raven, 1947). The letter has been confirmed as authentic and contemporaneous with the watercolour of *F. meleagris* (S. Kohlmann, pers. comm.). In consequence, both the Penny letter (Figure 3.8) and the watercolour (Figure 3.9) are now thought to be the earliest known description and depiction of *F. meleagris* respectively (pers. obs.).

(A) *"Bulbus variegatus, est Caparona ab inventore. Natalis Caperonus pharmacopoeus Aureliensis doctissimus et rei herbariae studiosissimus primus omnium quod sciam hanc plantam observavit. Dodrantalis est aut pedem longa, folia habet hyacinthi, breviora et angustiora multo, quinque, 6 vel 7, in caule alterna, a radice vero nulla, nisi solum unum vel duo foliorum r...; contra fere reliquorum bulborum naturam, quorum caules maxima ex parte foliis vidui esse solent. Caulis unicus, teres, tenuis, in summo florem sustinet, ex sex foliis oblongis constantem: intus apices sex lutei, et lingua trifida. Flos pendulus, colore variegato pulcherimo, ex quadrangulis purpureis et obscure candicantibus quasi maculis composito. Radix bulbosa, castaneae magnitudine, quae aliquando in duas vel tres partes dissecta videtur: inferius barbulum habet ut porrum. Flore marcescente caulis et folia contrahuntur, ac tenuiora fiunt, erecta omnia. Vasculum vel siliqua triquetra brevis ex tribus quasi columnis constans: intus tres seminum ordines continet. Semina Aristoloquiae Clematitis seminibus similia, sed multo minora, compressa, subflava, lenia. Flos nullo odore. Radix castaneam fere sapit, sed nescio quomodo in ore strangulantem naturam prae se ferre videtur. Floret Martio et Aprili, Iulii initio semina matura. Provenit in pratis ad Ligirim fluvium in agro Aurelianensi.*

T. P."

(B) "Variegated bulb, it is Caparona from its discoverer. Natalis Caperonus, most learned apothecary from Orléans and extremely assiduous in matters botanical, was the first of all, as far as I know, to observe this plant. Its stem is indeed nine inches long, it has leaves like the hyacinth, much shorter and narrower, five, 6 or 7, alternating on the stem, but none near the root, except one or two leaves [*word illegible*]; contrary to the nature of nearly all other bulb [plants], whose stems are mostly without foliage. There is only a single stem, smooth, thin, having the flower on top, regularly consisting of six oblong petals: within are six golden-yellow stamens, and a three-forked tongue. The flower hanging down, very beautifully coloured, composed, as it were, of quadrangular purple and dark-whitish spots. Root bulbous, of chestnut size, which sometimes seems to be divided in two or three parts: the inferior part has a little beard like the leek. When the flower is wilting, the stem and leaves contract and all upright parts become thinner. The small [seed] vessel or triangular, short pod consists, as it were, of three columns: inside it has three rows of seeds. The seeds are similar to those of *Aristolochia clematitis*, but much smaller, compressed, pale yellow, smooth. The flower is odourless. The root tastes almost like a chestnut, but I do not know why it seems to manifest in the mouth clearly a suffocating nature [effect]. The plant blossoms in March and April, at the beginning of July the seeds are ripe. It occurs in the meadows close to the River Loire near Orléans.

T[homas] P[enny]"

Figure 3.11 – Transcription from Latin **(A)** and translation to English **(B)** of the Thomas Penny letter of *c.* 1568 discovered bound to the side of the watercolour of '*Narcissus variegatus*' in the Conrad Gessner archive at University Library Erlangen-Nürnberg (FAU). Folio v482/FAU. Transcription work was undertaken by Peter Mason and Martin Steinmann; translation by Holger Funk and Peter Mason.

3.4 Discussion

The facilities afforded by the internet for searching archival material within multiple databases across multiple institutions, has enabled research such as that conducted here to proceed with great efficiency. It is only within the past two decades or so that such a wealth of historical literature has been made so accessible and freely available. A great pioneer of the progress made in digitising and presenting 'legacy' natural history material through this 'open access', is the important Biodiversity Heritage Library online facility, a consortium of 14 natural history and botanical libraries formed in 2005.

This facility, along with now several others, has revealed to researchers in many different disciplines a much broader range and depth of historical material than they were probably previously aware. In consequence, this has enabled them to study their elected subject in far greater detail and more comprehensively than was possible hitherto. However, the additional value of personal visits to appropriate libraries and archive repositories to study the private correspondence and other material of important established figures in the field of interest should not be underestimated. Much by way of important detail can only be accessed this way.

The study of all the important 16th century European herbals has revealed that *Fritillaria* species entered the public consciousness for the first time in the mid-16th century with the discovery of *Crown Imperial* (*F. imperialis*) and *Lilium persicum* (*F. persica*) by travellers to the Levant. Following the invention of the mechanical printing press by Johannes Gutenberg *c.*1439, the availability of printed herbals enabled interest in plants to move from their utilitarian qualities as *materia medica* to their more aesthetic virtues. The commercial opportunities afforded by these horticultural 'novelties' provided even greater impetus to the creation of a passion for new and exotic varieties of flora, fully appreciated by those returning with these flowers (usually bulbs) from abroad.

Without digitisation and online access to the Gessner archive at the Friedrich-Alexander-Universität Library Erlangen-Nürnberg (FAU) in Nuremburg, and the relationship built up with Sigrid Kohlmann of the Rare Manuscripts department there, the existence of the image of *Narcissus variegatus* (*F. meleagris*) and its accompanying letter from Thomas Penny would not have been discovered. The following events then

led to the discovery of probably the earliest known description and image of *F. meleagris*, and evidence of other appellations for the species at that time. To date, the year 1571 has been accepted as the first time *F. meleagris* had been described, with Nöel Capperon as the first to record it; and in 1574, Rembert Dodoens' woodcut image of *F. meleagris* in his '*Purgantium*' of 1574 (Dodoens, 1574) attributed as the earliest known image. The research described in this Chapter now provides evidence that 1568 is the date of the first record and description of *F. meleagris* and 1567 or 1568 for the first depiction *F. meleagris.* As to its first appearance in the British Isles, there is now some evidence suggesting that *F. meleagris* was first recorded here in 1596 – see Chapter 5, section 5.5.2

The author believes the content of Chapter 3 provides the most comprehensive account and chronology of the history of *Fritillaria* species in Europe found in a single document.

Chapter 4

Microsatellite development and analysis for *F. meleagris*

4.1 Introduction

Microsatellites (simple sequence repeats) are co-dominant molecular markers (loci) that have proved to be generally highly polymorphic, highly reproducible and locus-specific (Cipriani and Testolin, 2004; Jones *et al*., 2009; De Barba *et al.*, 2017), and because of these characteristics they are routinely isolated from plant material for purposes of population genetics differentiation analysis.

To understand the extent of genetic diversity in *F. meleagris* and apply that understanding to questions relating to the population structure of the species and to issues of relevance to their conservation, it was necessary to develop loci that can be used in population genetic studies. This Chapter describes the development of microsatellites using a high throughput sequencing (HTS) approach to produce a library of raw sequence fragments from a vouchered DNA sample. With development of screening and genotyping methods, these data were used to enable identification of potentially variable loci to find microsatellites most suitable for detecting variability between individuals. It is known that employment of microsatellites as markers has worked effectively in previous plant population and conservation genetics studies (Szczecińska *et al.*, 2013; Ambreen *et al.*, 2015; Bastias *et al.*, 2016; Fontúrbel *et al*., 2016). However, a comprehensive search of the literature did not reveal any previous development of microsatellite loci or associated primers for *F. meleagris*.

The approach adopted here was informed by the particularly large genome size of *F. meleagris* (1C-value *c*. 47.3 pg (*c*. 46.3 Gb)), the third largest reported thus far of any wild diploid plant species of the British Isles after *Viscum album* (mistletoe) and

Paris quadrifolia (herb-paris) (Leitch *et al.,* 2007; Zonneveld, 2010; Pellicer *et al.*, 2014). Large genome sizes make a number of HTS approaches unviable, because the technical strategies employed within systems supplied by individual manufacturers affect the cost per base sequenced, percentage genome coverage per run, read length and read accuracy of the resulting sequences (Ewing *et al*., 1998; Metzker, 2010; Morey *et al*., 2013; Reuter *et al.*, 2015). These constraints have a concomitant effect on the overall cost, elapsed time and human endeavour needed to produce viable and meaningful results.

The technology platform chosen to produce sequence fragment data was the second generation Roche 454 pyrosequencing Titanium chemistry GS-FLX Titanium XL+ system (commonly known as Roche 454) (Roche, 454 Life Sciences, Branford, CT, USA). The design principle is based on detection and quantification of DNA polymerase activity carried out using the enzyme luciferase (Morey *et al*., 2013). The Roche 454 system was found to be the most suitable for microsatellite locus development because it produced the longest sequence reads (1,000bp) at the time of the then commercially available Roche 454 products, and had a measured mean error rate of ≤ 1% (Gardner *et al*., 2011; Gilles *et al.*, 2011). Long uninterrupted fragment reads increase the probability that they will contain flanking regions of length great enough on either side of a targeted microsatellite repeat region to enable selection of the most suitable nucleotide profile for primer design.

4.1.1 Large genome size constraints

The frequency of microsatellite regions within a genome generally increases with increasing genome size. Consequently, the development of a primer set targeted at species with exceptionally large genomes, such as those of genus *Fritillaria,* may encounter greater difficulties in obtaining successful recovery of locus amplification products than for species with smaller genomes (Garner, 2002). With increasing genome size the PCR reaction can suffer amplification failure because the amount of target DNA is reduced relative to the total complement of template DNA, which increases search inefficiency in sequence matching and impedes discovery of target sequence primer binding sites (Farrelly *et al*., 1995; Rychlik, 1995; Garner, 2002; Altshuler, 2013). In addition, with large genomes the possibility increases of primers annealing to regions complementary to a primer or partial primer sequence (*i.e*. ‘mispriming’ – which consists of small stable sequences at the 3´ end of DNA

fragments binding to similar partial primer sequences, thereby nullifying the primer's function) rather than to the flanking regions of target microsatellites (Puskás and Bottka, 1995; Garner, 2002; Andreson *et al*., 2008). As a result, this non-specific binding will further deplete the usable primer pool, which in effect exacerbates dilution of the target/non-target DNA ratio. At the other extreme, and particularly with large genomes, too little template DNA can lead to little or no product yield except for primer-dimers and other artefacts - most often observed when primers fail to bind efficiently to the target DNA and instead bind to themselves as a result of a higher than expected concentration of complementary strings of primer nucleotides.

4.2 Materials and methods

4.2.1 DNA template sample acquisition and assay

The Roche 454 system sample preparation method for a pyrosequencing run specifies the need for ≥ 4 µg of ≥ 100 ng/µl concentration DNA with a purity (light intensity absorbance ratio of 260 nm/280 nm) of ≥ 1.8. The appropriate quantity and concentration of template DNA was acquired from a *F. meleagris* specimen (accession no. 1953 51806: DNAB no. 16114) held within the DNA bank at RBG Kew (http://apps.kew.org/dnabank/homepage.html). Calculations of DNA purity and concentration were taken from the mean of three repeat measurements of the absorbance ratio measured on a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific, 2009).

4.2.2 454 sequence production and quality assessment

A Roche 454 whole genome shotgun (random) sequencing run of 1/8 PicoTiterPlate (the minimum large genome coverage to ensure discovery of a practical number of microsatellites to work with (Schoebel *et al.*, 2013)) produces a DNA base sequence fragment library and a matching sequence quality file. For the project, these outputs were combined into a file in FastQ format using a personally written Perl program language script. The resulting file associated each sequence base-call with an encoded chromatogram quality (read accuracy) value (Phred score) based on peak shape, periodicity between peaks, signal strength and background noise - signifying the confidence level of the Roche 454 sequencer in providing a correct base-call (Phred score: 0 = unusable; 40 = perfect; 20 is an accepted standard for a high quality usable

base (Ewing and Green, 1998; Ewing *et al*., 1998; NHGRI (2017)). Sequence quality control checks were undertaken using program FASTQC v 0.10.1 (Andrews, 2012; Fernandez-Silva and Toonen, 2013), which conducted a series of tests on the FastQ file Phred scores, pinpointing any raw sequence data quality problems which might affect results at a later stage of analysis. Using program LUCY2 v2.2 (Li and Chou, 2004) and the FASTQC results file as input, sequences < 100 bases in length were discarded, and regions deemed to have suspect quality (those with ≥ 2 % error probability because of peak shape or lack of signal strength) were trimmed from the 5´ and 3´ ends of all reads.

4.2.3 Microsatellite discovery and primer pair design

The modified library of good quality sequence read data was entered into program QDD v2.1 (Meglécz *et al*., 2010) to screen for perfect repeat motif microsatellite loci, with subsequent automated design of primer pairs *de novo* undertaken by the integrated subprogram PRIMER3 v1.1.4 (Rozen and Skaletsky, 2000). The multiple output reports enable the user to select microsatellites and associated primer pairs according to multiple criteria.

Initially, QDD provides an analysis pipeline to detect candidate microsatellites and eliminate duplicated sequences and large clusters of transposable (mobile) elements, because their exclusion can increase the proportion of potentially viable primer pairs to be tested by PCR at a later stage.

The subprogram PRIMER3 provides a suite of user specifiable selection criteria to target microsatellites containing a minimum of five repetitions of 2-6 base motifs. For reads that have the potential to encompass more than one target microsatellite, all possible microsatellite combinations encompassed within the read sequence are defined as the target region. However, if the microsatellite(s) detected are too close to the extremities of the read, the sequence is automatically discarded. CLUSTALW v2.1 (Larkin *et al*., 2007), a subprogram of QDD, enables creation of microsatellites from overlapping alignments of multiple sequences (contigs), where there is a minimum match between all participating sequences of ≥ 95%.

PRIMER3 enables primer pair design to be optimised for subsequent microsatellite amplification by PCR; and an assessment of expected PCR amplification efficiency of the selected primer pairs is reported as a score of increasing success potential. Finally,

an automatic BLASTn v2.2.29 search of each generated primer sequence conducted against the GenBank database (NCBI: www.ncbi.nlm.nih.gov/blast - 13/1/14) provides information on the best sequence matches found of previously developed microsatellite loci and taxonomic lineage of the respective organism.

Default execution parameter settings were maintained for QDD and PRIMER3, except for the following: minimum (target) sequence length was increased to 100 bases from 80 bases to ensure maximum ability to encompass microsatellite loci and two usable flanking regions; maximum acceptable difference between forward and reverse primer melting temperatures was reduced from 10°C to 5°C to ensure optimal conditions for primer pairs to bind simultaneously to their target flanking sequences; minimum percentage of G and C nucleotides in any primer was increased from 20% to 35% to ensure the strongest bond to the template sequence - as Gs and Cs have three bonds, whereas Ts and As have two bonds; maximum percentage of Gs and Cs in any primer was reduced from 80% to 65% to provide greater flexibility in primer design; number of consecutive Gs and Cs at the 3´ end of the primer was increased from 0 to 2 to help promote specific binding at the 3´ end due to the nucleotides' stronger bonding characteristics (L.J. Kelly, pers. comm.).

Final selection of an optimum set of primer pair/microsatellite combinations for initial PCR amplification testing was undertaken personally by evaluating various criteria provided as guidelines within QDD and PRIMER3 documentation (Meglécz and Martin, 2011). Selection was based on characteristics which best balanced a combination of primer length, primer sequence quality (low homopolymer content, GC content), primer annealing temperature, microsatellite sequence quality and estimated PCR product length. The attributes chosen were considered those most likely to embrace the best overall characteristics of primer pair/microsatellite design and provide the best opportunities for subsequent successful PCR amplification.

4.2.4 Population leaf sample collection

To provide genetic material from which to isolate DNA in advance of addressing questions of *F. meleagris* population structure, leaf samples were collected from individual wild *F. meleagris* plants from each of 26 populations distributed across Eurasia, as described in Chapter 5 (Table 5.2). One individual was selected from each population to be used in creation of a test panel for initial primer pair/microsatellite evaluation and PCR optimisation.

4.2.5 DNA isolation from a test panel of population samples

DNA was extracted from approximately 20 mg of newly crushed and ground fresh or silica-dried leaf tissue obtained from each individual in the test panel, to create a DNA library to be used in initial primer pair screening and PCR optimisation.

The procedure used a mini-extraction protocol developed at the Jodrell Laboratory at RBG, Kew, which followed a cetyltrimethylammonium bromide (CTAB) method modified from Doyle and Doyle (1987) to remove polysaccharides, polyphenols and other secondary metabolites such as alkaloids and flavonoids (Borges *et al*., 2009). This would provide the best opportunity to yield reproducible and consistent amplification products (Azmat *et al*., 1987). To provide consistency of results for PCR reactions, all freshly isolated DNA samples were diluted to a concentration of 50 ng/µl by the addition of de-ionised H_2O. The method for ascertaining the purity and concentration of the resulting individual DNA samples has been previously described in section 4.2.1.

4.2.6 M13-tailed primer method and procedure

To enable individual DNA fragments to be detected and measured by a sequence fragment analyzer prior to allele scoring and subsequent genotyping, they must each contain a fluorescent dye incorporated into the PCR product using a labelled primer. The procedure has the following stages: initial *de novo* synthesis of template DNA takes place with a locus-specific reverse primer (Eurofins MWG Operon, Ebersburg, Germany) and a two-part locus-specific forward primer incorporating a randomly generated M13 universal tail sequence (non-homologous to any known plant genome sequence) appended to the primer 5´ end (19 nucleotides: 5´-CACGACGTTGTAAAACGAC-3´ - Eurofins MWG Operon, Ebersburg, Germany). By the second round of synthesis the product undergoing amplification will have the M13 universal tail sequence incorporated into the PCR product. The third round has an M13 sequence complementary to the M13 universal tail sequence (with one of two chosen fluorochrome dyes (labels) incorporated as a label at the 5´ end (5-FAM or JOE - DS32 Matrix Standard Kit - Applied Biosystems, Warrington, Cheshire, UK) taking the place of the original forward M13 universal tail sequence due to its lower annealing temperature. The result is an amplified product containing a fluorescent dye that can be detected by the fragment analyzer.

4.2.7 Primer pair screening

A series of PCR amplifications was undertaken on a subset of the test panel of *F. meleagris* individuals to determine the viability of each primer pair/microsatellite combination selected for further development (section 4.2.3). As a practical approach to limit both time and cost of PCR testing for each locus, one individual was selected from each of the 11 Eurasian countries from which population samples were collected (see Table 5.2 in section 5.2.3). By determining which of the microsatellite loci were successfully amplified with individuals from a limited number of populations, it was reasoned that locus selection based on this criterion could be used with reasonable confidence to predict amplification success with individuals selected from all 26 populations.

The screening runs used the manufacturer's recommended primer melting temperatures (T_m) (data not shown) in the reactions, and a volume of freshly isolated DNA representing a concentration of 50 ng/µl. The method of Lavor *et al*. (2013), incorporating and modified by the work of Almeida *et al*. (2013), was used for the PCR protocol and M13-tailed primer pair procedure.

4.2.8 PCR protocol

PCR reactions were carried out using a GeneAmp PCR system 9700 thermal cycler (Applied Biosystems, Warrington, Cheshire, UK) as follows: for each individual PCR reaction, a final volume of 10 µl was assembled using 7.1 µl Fermentas PCR Master Mix (2X) at 4 mM $MgCl_2$ and 0.4 mM dNTPs, with T*aq* DNA Polymerase (Thermo Fisher Scientific, 2014), 0.5 µl 0.4% (w/v) Bovine Serum Albumin (BSA), 0.4 µl M13-tailed forward primer, 0.8 µl reverse primer, 0.2 µl M13 fluorescent labelled universal tail sequence and 1µl (50 ng/µl) template DNA. For individual loci, the following amplification conditions were employed: an initial step of DNA template denaturation at 94 °C for 3 min, followed by 40 cycles of: 20 s template denaturation at 94 °C, 40 s primer annealing to DNA strands at the manufacturer's T_m and 20 s strand extension at 72 °C. Then to incorporate the M13 universal tail sequence, 8 cycles of: 1 min denaturation at 94 °C, 1 min primer annealing at 53 °C, 1 min extension at 72 °C, followed by a final strand extension of 30 min at 72 °C. The PCR products were run out on 2% agarose gel and analysed visually to determine locus amplification success.

4.2.9 Optimisation of template DNA quantity and primer T_m for PCR

Using two of the 11 selected individuals (to limit use of time and PCR reaction cost), optimisation of DNA template quantity in conjunction with primer T_m was carried out for each of the successfully screened primer pair/microsatellite combinations and associated M13 universal tail sequences (section 4.2.7).

A fixed volume of individual primers, reagents and a variable volume of template DNA (section 4.2.8 for PCR protocol) were used for individual gradient runs using a Veriti™ Dx 96-Well Thermal Cycler (Applied Biosystems, Warrington, Cheshire, UK). Steps were carried out for -1 °C to +2 °C of the primer supplier's recommended T_m in 1 °C stages, and for volumes of DNA template from 0.5 µl - 1.0 µl in steps of 0.1 µl. The resulting individual PCR products were run out on a 2% agarose gel and analysed visually to determine which combination of T_m and DNA template volume produced the best judged strongly defined band of expected product size. Where multiple bands provided the most favourable results, the highest T_m in combination with the associated DNA template volume was chosen. The higher T_m would allow PCR synthesis under more rigorous conditions for this critical component as, in general, the higher the melting temperature the more specific will be the annealing of the primer to its target template (McPherson and Møller, 2006; Eppendorf Scientific, Inc., Hauppauge, New York, USA). A lower DNA template volume (*i.e.* lower genome copy number) could result in enhanced annealing specificity and therefore reduction in unwanted artefact peaks due to a reduced DNA 'swamping' effect from the large genome DNA content. However, this might possibly be offset by less availability of target DNA.

Ruano *et al.* (1989) and Thermo Fisher Scientific (2015) recommend that for successful PCR amplification, 1,000 copies of target DNA (*i.e.* 1 µl of *F. meleagris* DNA at 50 ng/µl concentration) are required in order to prevent primer dimer synthesis and other spurious products unnecessarily consuming primer stock. The PCR protocol used in this project initially followed that recommendation.

The primer pair T_m values and DNA template volume combinations selected as producing the strongest and most well defined bands visualised on 2% agarose gel (not shown) were selected for use in all subsequent PCR reactions (Table 4.1).

4.2.10 Genotype recovery from panel individuals

In preparation for recovery and scoring of alleles and subsequent genotype recovery, capillary electrophoresis was performed individually for each PCR product on an ABI 3730 DNA fragment analyzer (Applied Biosystems, Warrington, Cheshire, UK), using the recommended 1 µl of PCR product added to 10 µl Hi-Di Formamide (Applied Biosystems, Warrington, Cheshire, UK), with the addition of 1.2% GeneScan 500 (-35, -50, -250) ROX size standard (allelic ladder) (Applied Biosystems, Warrington, Cheshire, UK). Genotyping of recovered allele size data was carried out, with allele sizes visualised as electropherogram traces, using the program GENEMAPPER v5.0 system (Applied Biosystems, Warrington, Cheshire, UK).

4.2.11 Problems of inconsistency in allele sizing and genotype recovery

GENEMAPPER calculates allele lengths by comparison with an internal ROX size standard run alongside the PCR amplified products during capillary electrophoresis in the ABI 3730 fragment analyzer. Calculations are based on the assumption that DNA fragments of the same length migrate (flow through the analyzer capillary network) at the same rate. This is not the case in practice as they can have different mobilities based upon their repeat unit (motif) size, GC content and fluorescent label type (Wenz *et al*., 1998; Amos *et al*., 2007). As a result, fragments can vary in their calculated sizes, because the relationship between fragment lengths, migration rates and size standard can often be non-linear (Applied Biosystems, 2004; Matschiner and Salzburger, 2009). Errors in sizing can also be introduced by the imprecision of the analyzer, as can imperfect linearity between the size standard and PCR product length due to drift resulting from random events during early PCR cycles (Wagner *et al.*, 1994). Inconsistency between runs and minor differences between individual capillaries can also have a negative effect (Matschiner and Salzburger, 2009; Thermo Fisher Scientific, 2014). In consequence, allele sizes calculated (‘scored’) by GENEMAPPER as part of the program’s automated allele scoring and genotyping facility are inconsistent, and are ‘binned’ (fragments assigned to allele categories according to their lengths) with values stated to two decimal places to enable application of further analysis to obtain more accurate and consistent genotyping results.

Use of an automated system should have many advantages, not least the amount of time to be saved by use of a ‘pipe-line’ process, such as that available within GENEMAPPER, to analyse and measure the many PCR product sequences developed

for a project. However, the effects of the many issues described in this section can cause GENEMAPPER to encounter difficulties in attempting to differentiate artefacts from true allele peaks, and therefore to correctly interpret the information represented by the electropherogram traces.

4.2.12 Genotype recovery process

Successful recovery of a viable genotype library was achieved manually using GENEMAPPER facilities. For the purposes of comparative measurement, the reproducibility and precision of allele scoring are much more important for downstream processing than absolute accuracy. Therefore, a decision was taken to check each electropherogram trace produced by GENEMAPPER. Visual electropherogram trace analysis, interpretation and editing (to avoid missed or incorrect allele calls) and individual allele scoring were conducted over several months, gradually building experience of trace structure through practical assessment of allele characteristics and electropherogram peak form. Multiple re-examinations of all traces were undertaken, followed by an independent review and re-scoring of each resolved allele trace peak position at least three times to confirm and be confident of results.

4.2.13 Consistent and logical allele boundary sizing

Simple rounding of decimal allele sizes, and therefore genotype values, to the nearest whole number can lead to inconsistencies, such as the presence of even and odd values for the same genotype at the same locus for different populations. To address the issue, an opportunity was taken to use the program TANDEM v1.09 (Matschiner and Salzburger, 2009) to adjust the manually scored allele size data. The program has the facility to assess and correct the effects on allele size measurements caused by the inherent inaccuracies and inconsistencies associated with capillary electrophoresis described in section 4.2.11, and uses the information to modify genotype values to have consistent and logical boundaries accordingly. The program completes processing by binning the genotype data in the integer format required as input by the majority of downstream population genetics analysis programs.

4.2.14 Software programs used for genotype development

With the exception of GENEMAPPER, all software programs used in development of microsatellite loci and associated primers were either commonly available freeware or open access modules, downloaded and installed on a personal PC supporting the Microsoft Windows 7 operating system. The programs are all very mature in age and facilities offered, and are regularly used by many independent researchers, as witnessed by their inclusion in the substantial number of population genetics research papers and articles published (*e.g.* Waldbieser *et al.*, 2003; Pante *et al.*, 2015; Balao *et al.*, 2016; Jaffé *et al.*, 2016) However, these programs are mostly accompanied by either rudimentary or a complete lack of documentation. As the programs are made freely accessible to anyone who wishes to use them, and are generally unsupported by their authors, they usually present numerous problems whilst attempting their execution on the target operating system in question (Microsoft Windows 7 (Microsoft Corp., Redmond, Washington, USA) in this case). This is usually because they were originally written for a Linux computer host system and latterly have been converted to other operating environments. In consequence, for some programs, a considerable amount of time was spent in program investigation and problem solving through Perl language script debugging, modification and re-programming, prior to successfully achieving the desired results from their execution.

4.3 Results

4.3.1 DNA template measurements of purity and concentration

The mean of three repeat measurements of the absorbance ratio of *F. meleagris* DNA specimen (accession no. 1953 51806: DNAB no. 16114) was calculated, with the results recorded as purity of 1.92 (ratio of sample light absorbance at 260 and 280 nm wavelength from xenon flash source) and concentration of 143.2 ng/µl.

4.3.2 454 sequence production and quality assessment

The Roche 454 whole genome shotgun sequencing run yielded a DNA sequence fragment library of 23.4 Mb (megabases) of data (representing 0.04% genome coverage). The library contained 35,369 individual sequences of read lengths 29-1,119 bases, with an average length of 661 bases. The results of quality assessment tests conducted on the fragment sequence library produced an average quality score across

all reads of 27.5 (representing 0.2% error in call quality assessment), with 2,744 bases disregarded due to an assessment by the Roche 454 sequencer of insufficient certainty to make a base call. There were 897 sequence reads discarded as unsuitable due to length of read or poor quality. The remaining 34,472 sequence reads of good quality were retained for downstream analysis.

4.3.3 Microsatellite discovery and primer pair design

Analysis of the 34,472 reads by QDD and PRIMER3 resulted in selection of 7,429 potentially suitable candidate primer pair/microsatellite combinations. Exclusion of potential target microsatellites with compound repeats (combinations of micro and nanosatellites - 3-4 tandem repeats of a 2-6 base motif) and interrupted repeats (deemed unnecessarily complex and unlikely to exhibit polymorphism) reduced the number to 3,128 potential candidates. These were selected by QDD and PRIMER3 as suitable for primer design because they contained at least one microsatellite repeat unit and $\geq$ 50 bases in each flanking region. Of these potential candidates, 346 contained $\geq$ 10 repeats (this level of repeats is associated with higher discovery rates of polymorphic loci as, it is postulated, mutation rate increases with repeat number) (Thuillet *et al.*, 2002; Vigouroux *et al.*, 2002; Gusmão *et al*., 2005; Abdelkrim *et al*., 2009; Marriage *et al*., 2009; van Asch *et al.*, 2010). The number of candidates was further reduced to 103, following elimination of microsatellites that contained $\geq$ 5 homopolymer (single identical nucleotide and therefore not microsatellite) repeats within flanking regions. This number was further reduced personally to a final selection of 36 candidates by (1) selecting primer pairs with PRIMER3 calculated values for left and right hand primer stability of approximately 9 (Rychlik, 1993). Values of this magnitude offer the greatest stability for each primer's last five 3´ nucleotides during PCR (averages for the selected 36 left and right hand primers were 9.0 and 8.9 respectively); by (2) elimination of microsatellites containing repeat motif polymorphism emanating from contigs and their consensus sequences.

A BLASTn search conducted against the GenBank database for sequence matches to the 36 selected primer pairs found no matches for *F. meleagris* entered by previous workers.

4.3.4 Primer pair screening and selection

Of the 36 candidate primer pairs tested for one individual selected from each of the 11 selected populations, 19 were successful in PCR amplification and 17 failed amplification for all 11 populations - 8 had no amplification products visible, 5 were monomorphic and 4 produced only primer-dimer products (most often observed when primers fail to bind efficiently to the target DNA and instead bind to themselves as a result of a higher than expected concentration of complementary strings of primer bases (Garner, 2002)). Candidates that failed were tested in repeat PCR reactions at least twice before they were finally rejected. The 19 successful primer pairs and associated M13 universal tail sequences (Table 4.1) were selected for use in all subsequent PCR reactions.

4.3.5 Optimising template DNA quantity and primer T_m for PCR

The results of PCR amplifications carried out for each of the 19 loci to ascertain optimal values (section 4.2.9) for primer T_m are shown in Table 4.1. With regard to DNA template, a change in volume from 1 µl to an optimal value of 0.7 µl (equivalent to 35 ng/µl - 700 genome copies for *F. meleagris*) was determined. This was balanced by a 4.2% change in volume of Fermentas PCR Master Mix (2X) from 7.1 µl to 7.4 µl, with the reasonable expectation this would have no negative effect on PCR outcome.

4.3.6 Test panel alleles and genotyping

With adjustments made to the PCR protocol (section 4.3.5) for optimised primer T_m values and DNA template volume, the test panel of 11 population individuals was now increased to 26 (one individual from each population) to undergo PCR testing for each of the 19 selected microsatellite loci.

For each locus, a mean of 17 individuals (range: 13 - 20) amplified successfully (65%), with PCR reactions repeated at least twice to ensure reproducibility of results. All images from 2% agarose gel runs of successful amplifications (results not shown) displayed evidence of locus polymorphism of approximately the expected product size measured against a molecular weight marker (ladder) (EasyLadder 1 (100 bp – 2,000 bp) - Bioline Reagents Ltd., London, UK).

Scoring of alleles and associated genotypes within GENEMAPPER resulted in recovery of 211 genotypes, with 14 (54%) heterozygous individuals present across the 19 loci

and all loci polymorphic (mean 70% within populations). The genotype recovery rate of 43% (26 populations x 19 loci = 494 possible genotypes) was judged a reasonably good result with respect to the difficulties encountered in scoring of alleles due to the numerous barriers encountered in electropherogram peak and artefact recognition and measurement (57% of electropherogram traces could not be scored with full confidence and were eliminated for purposes of further analysis).

Table 4.1 - Characteristics of 19 microsatellite loci and primers developed for *Fritillaria meleagris.*

Locus	Repeat motif	Primer sequences (5´-3´)	T_m°C
L4	$(AT)_{15}$	F: AAATTTATGTTGTGAGCATCGG* R: TTCAAGAATCTTAATTGCGGC	57
L6	$(AT)_{15}$	F: CGATAGCATTATGTATGTGGATTATCG* R: AACCAAGTAGGAAACCAGAGG	56
L11	$(AT)_{13}$	F: TTAGTTTCCCAGTTTGTCGC* R: TTGACTTAGTTGGTGGCTCG	55
L12	$(AT)_{13}$	F: GGGAAACCAATACAAGATCCC* R: GCAAGTATGGCATCATCACG	55
L13	$(AT)_{13}$	F: CTAACTTGAAGTCTGGTTATGGG* R: CACTTTCTCCAAAGATAAACATGG	56
L14	$(TA)_{13}$	F: TTGTTACTTTGGGACATTAACGC* R: AGTCGAAGGGATAGAAGAGGG	58
L15	$(AT)_{13}$	F: CAGAACAAACTAACTGGTCAGGG* R: CTCGAGGATCCCAACTCTCC	59
L16	$(ATT)_{12}$	F: AAATGAGGCTTTGAGTGTTTCG* R: TCACAAGCGCACTATAAGACC	56
L21	$(AT)_{11}$	F: CTCTGACGGGATTTCCTGG* R: GTCATATTTGTGGGCTTGGC	55
L23	$(AG)_{11}$	F: CACGACGACTCATGTAAAGCC* R: AGGTGCATATTCCAGAAGGC	55
L25	$(AG)_{11}$	F: TTGCTTGATGATAATAGGACACC* R: GAAACTTGATTCCTCCCTTTATCC	57
L27	$(ATA)_{11}$	F: CAAAGTGGTCCTAACTGTAGACCC* R: GAAGTTGAGTGGGTAGTATTGGG	59
L28	$(AT)_{11}$	F: AGTCCGCCCAAATTTAAAGG* R: GGAAGGTATGCAATAATTAGGATGG	58
L32	$(AC)_{10}$	F: GATCTAGAATACCAAAGATGTTCAGC* R: TGAGTTATGATAGCTGATAGTGAACC	58
L34	$(AC)_{10}$	F: TCTGGGAAATTTGACCAACG* R: TCCATCCACAATGACCATCC	55
L35	$(AG)_{10}$	F: TTTGGTTAAGGAGGCTGTGC* R: CCCAATGTGCTTGATTCCC	55
L36	$(AG)_{10}$	F: AAAGGGAATGAAGAAGGTCC* R: CATTAAAGCTCAAATGACTACCAGC	58
L38	$(AG)_{10}$	F: CCCTCTTCTCTATCTTTATTCTTCACC* R: GAGGAGAGAGAAAGTCAACCACC	60
L39	$(GT)_{14}$	F: TCCTCTTAGGGACTTAGCAAGAA* R: TGACACCATTTGAGTCTTTTCG	55

Asterisks* mark primers (F: forward, R: reverse) M13-tailed at the 5´ end; primer melting temperature (T_m°C).

4.3.7 Problems encountered during the genotyping process

Problems encountered during scoring of alleles were as follows: (1) Stutter peaks, or so-called allelic shadow bands. These are multiple minor PCR products originating from the same fragment, typically shorter by one or more repeats than the full-length product and therefore always precede the true allele peaks (Figure 4.1). They are mostly formed from dinucleotide repeats and are the result of PCR *Taq* DNA polymerase slipped-strand mispairing ('slippage') during strand extension. Stutter peaks are proportional in number, periodicity and peak height to the repeat length and number of repeats (much less prevalent with tri-, tetra- and pentanucleotide repeat sequences) (Edwards *et al.*, 1991; Walsh *et al.*, 1996; Applied Biosystems, 2004; van Oosterhout *et al*., 2006; De Woody *et al*., 2006); (2) The tendency of *Taq* DNA polymerase to add non-templated 'Plus A' and split peaks (caused by incomplete 'A' nucleotide addition) to the 3´ end of PCR fragments after extension (Figure 4.1) (Ginot *et al.*, 1996, Guichoux *et al.*, 2011); (3) Large allele dropout, which results from preferential amplification of the smaller allele in a heterozygote genotype, and makes peak structure appear homozygous, for which it is measured (Björklund, 2005); (4) Primer dimers, which are artefactual non-specific and tri-allelic peaks caused by mispriming of primers during PCR (Puskás and Bottka, 1995; Brownie *et al.*, 1997; Hill *et al.*, 2009); (5) Allele peaks which deviate from the expected periodicity of repeats (peak-to-peak distance differing from the underlying repeat unit) due to migration rate variability, which can therefore suffer from erroneous measurement (Guichoux *et al.*, 2011) (Figure 4.1).

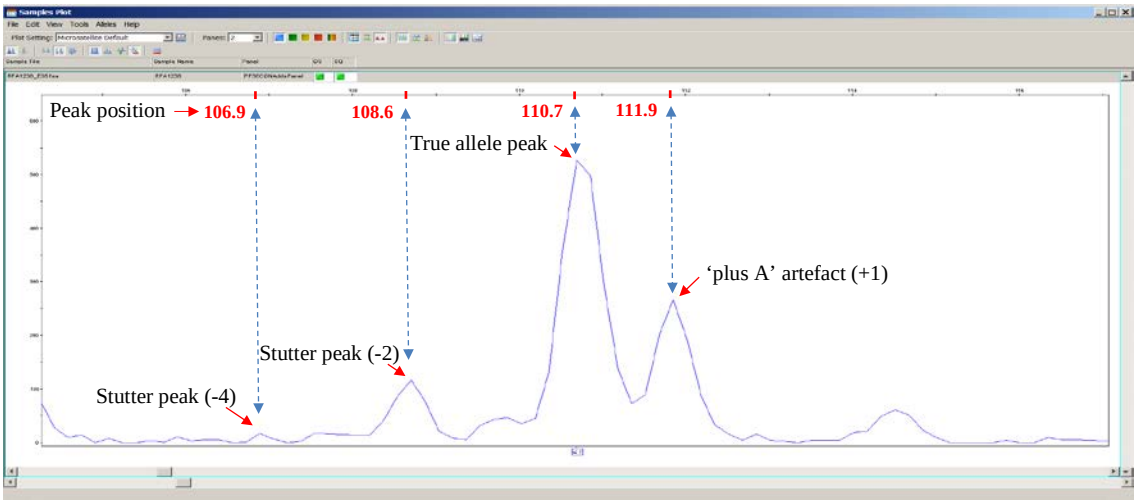

Figure 4.1 - GENEMAPPER electrophoresis pattern from L38, an $(AG)_{10}$ microsatellite locus of a homozygous individual (Table 4.1). Shown are stutter peaks at -2 and -4 nucleotide positions from the true allele peak; a 'plus A' nucleotide addition after the true allele peak; peaks showing evidence of periodicity deviation in allele peak positions.

4.4 Discussion

To explore genetic differentiation across 26 geographically isolated Eurasian populations, a panel of one individual was selected from each population for microsatellite locus recovery and genotype development from raw sequence fragments using a HTS approach. The challenge was to find suitable microsatellite loci that amplified effectively using PCR-based technologies. The approach taken here identified 19 microsatellite loci with > 50% success rate in revealing suitable loci from predicted loci. The results compare favourably with the equivalent findings of three research groups working in separate fields using the same Roche 454 HTS approach to sequence recovery and development of viable microsatellite loci using programs QDD/PRIMER3: Drechsler *et al*. (2013) analyzed the success rate of microsatellite loci development in contrast to an Illumina-based approach for three amphibian (newt) species (*Triturus cristatus*, *Calotriton asper* and *Lissotriton helveticus*) (Table 4.2); Gardner *et al*. (2011) focused on three different species in two different plant families (*Baumea juncea* (Cyperaceae), *Gahnia trifida* (Cyperaceae) and *Triglochin procera* (Juncaginaceae)) as part of an approach to provide best practice advice for ecologists (Table 4.2); whilst Rico *et al.* (2013) worked on a method to determine microsatellite characterisation of the non-model complex of whitefishes (*Coregonus* spp., specifically

Coregonus clupeaformis) with regard to its evolutionary diversification through adaptive radiation (Table 4.2).

Table 4.2 - For a standardised no. of Roche 454 HTS fragment reads: comparative statistics for no. of microsatellites with ≥ 10 perfect repeats recovered from program PRIMER3; no. of primer pairs tested for potential viability for successful downstream PCR; no. of polymorphic loci selected for further processing; percentage of primer pairs tested resulting in polymorphic loci successfully isolated.

Source of statistics	Microsatellites ≥ 10 repeats	Primer pairs (loci) tested	Polymorphic loci successfully recovered
P Day	346	36	19 (53%)
Drechsler *et al*. (2013)	204	20	11 (45%)
Gardner *et al*. (2011)*	334	12	7 (52%)
Rico *et al.* (2013)	176	9	5 (56%)

* Figures based on ≥ 8 repeats

As a result of primer T_m and DNA optimisation, it is possible that a re-test of the 17 failed primers out of the 36 originally screened might have generated additional viable loci with which to work. Cost and time resource issues prevented this approach; a decision which was also informed by the number of loci already recovered, lower numbers of which other population genetics studies have used successfully to produce informative results from limited numbers of sampled population individuals: Li *et al*. (2014) developed 10 microsatellite loci to study 44 individuals of the forest tree *Dacrycarpus imbrictus*, a vulnerable southern Chinese species (Cheng and Fu, 1978). It is found across Hainan Island and in scattered stands on the mainland (Chen *et al.*, 2004) and is used for construction and furniture (Cheng and Fu, 1978). Since 1992 it has been placed on the Red List of Endangered Plants in China (Fu and Jin, 1992). However, as a result of the Li *et al*. (2014) research, no significant differences in genetic variation and differentiation between the island trees and those on the mainland were discovered, contrary to expectation. Wogan *et al*. (2015) developed nine loci for 16 individuals of *Anniella alexanderae* (a western USA and Mexican limbless lizard) and discovered high levels of heterozygosity and polymorphism information content (PIC). These results indicated they would have high utility is assessing population genetic and demographic patterns among four hitherto unknown and recently described *Anniella* species. Nine loci were developed by van Paridon *et al*. (2016) for population genetics studies of 66 adult individuals of the cattle liver fluke *Dicrocoelium dendriticum*, newly emergent in southern Alberta, Canada. The findings are to be used towards a greater

understanding of the parasite's complex life cycle, ecology, transmission mechanisms and evolution.

The results obtained from the panel of *F. meleagris* individuals indicate that additional genotype data developed from further selected individuals from the 26 populations can be used to inform a much larger study of population structure, as outlined in Chapter 5. However, with such large numbers of individuals and alleles to be processed and analysed, and with the caveats described in section 4.2.11, it is not expected that complete success will persist in PCR amplification and subsequent recovery of alleles and genotypes for the remaining 104 individuals (26 populations x 4 individuals from each).

Chapter 5

Genetic diversity in Eurasian populations of *F. meleagris*

5.1 Introduction

An appraisal of the activities of Europe's earliest botanists, apothecaries and plant collectors (Chapter 3) now reveals that *F. meleagris* was first recorded in the wild and depicted in Northern Europe in the French city of Orléans in 1567 or 1568 and described in Zurich in 1568, not in 1571 for both, as has previously been assumed. As the present distribution range is known to be Eurasian in breadth (Figure 5.1; WCSP, 2014), the question arises as to the natural origin and population structure of *F. meleagris*. To address this question, a population genetics study was conducted from *F. meleagris* leaf samples collected across Eurasia from 26 biologically isolated populations. The study revealed distinct population structure in the data, with consequences for the conservation status of this enigmatic species.

5.2 Materials and Methods

5.2.1 Methodology of approach

In order to determine the viability of conducting a study of *F. meleagris* population structure across Eurasia, it was first necessary to develop microsatellite markers to be applied to population genetics work from a test panel of single individuals from each of 26 chosen populations (Chapter 4). The results of the tests revealed a significant level of variation amongst individuals and loci across populations (see section 4.3.6), indicating it would be informative to apply the microsatellites to a larger panel of individuals to enable a broader study of *F. meleagris* genetic differentiation.

5.2.2 Distribution and status of *F. meleagris*

Fritillaria meleagris has a wide distribution range across Eurasia and is a nationally rare and protected species in all countries where it has been recorded (IUCN, 2017). The species has been assigned variations of statutory designations of conservation status (Chapter 6) in all countries within its range (Schnittler and Günther,1999), but has not been evaluated as an IUCN red-list species (IUCN, 2017). The plant is present in Eastern, Central and Northern Europe, Denmark, Fennoscandia (Finland, Norway, Sweden), the Baltic states (Estonia, Latvia, Lithuania, Kaliningrad region) and Russian Federation (European part, southwestern Siberia (Barnaul Floristic Region) and possibly Kazakhstan) (Druce, 1897; Lozina-Lozinskaya, 1935; Simpson, 1982; Abdulina, 1999; Malyschef and Peschkova, 2001; Tomović *et al*., 2007; WCSP, 2014). However, the distribution of *F. meleagris* within this range is discontinuous often to a high degree (Turrill, 1951). The plant was introduced and now naturalised into Denmark, Fennoscandia and the Baltic states (Table 5.1), and putatively introduced (pers. obs.) and now naturalised in England (Clapham *et al*., 1989; WCSP, 2014). There are no recognised subspecies of *F. meleagris* in the wild (*F. meleagris* subsp. *burnatii* is a synonym for *F. tubaeformis* var. *burnatii* (WCSP, 2017)).

Table 5.1 - Eurasian countries within which *Fritillaria meleagris* has been recorded and status determined.

Austria, Belgium†, Bosnia and Herzegovina, Croatia, Czech Republic†, Denmark, England, Estonia, Finland, France, Germany, Hungary, Italy, Latvia, Lithuania, Kaliningrad region, Montenegro?, Netherlands, Norway, Poland, Romania, Russian Federation (European part and Altai region of S. W. Siberia, Kazakhstan), Serbia, Slovenia, Sweden, Switzerland, Ukraine

Key to species presence and putative status: native (*e.g*. Austria); doubtful ?; extinct†; 'red' countries where introduced but now naturalised (*e.g*. Denmark).

5.2.3 Determining which populations to sample

Population selection was determined by geographic location in the expectation that geographic distance might correlate with genetic isolation by distance. Historically, distance has not prevented the possibility of anthropogenically mediated *F. meleagris*

bulb migration between populations. However, with the exception of the English populations, a relative lack of interest in the recorded history of the population sites where it grows and in *F. meleagris* from a conservation point of view before the early 20th century, combined with large inter-population distance, would suggest the likelihood of anthropogenic mediation is low.

For continental Eurasia, a review of the literature revealed the locations of numerous major *F. meleagris* population sites across the region. Accurate information with regard to annual numbers of flowering plants and undisturbed population longevity is generally not available in most cases. However, anecdotal evidence from the sample collectors and references within individual country flora generally provided enough information from which to make an assessment for each population of location accessibility and value in provision of data to the overall project. Each population sampled was estimated to have been undisturbed and recorded for at least 120 years (Baumgarten, 1816; Boreau, 1857; Bouvier, 1878; Tourlet, 1908; Hayek A, 1933; Schmalhausen, 1897; Haeupler *et al.*, 1976; Floron, 2016; Sinngrund, 2016). The 17 Continental population sites selected for sampling were to be found in 10 countries (three in Romania, two in each of France, Germany, Russian Federation (European part and Siberia), and one in each of the Netherlands, Poland, Serbia, Sweden, Switzerland and Ukraine (Figure 5.1). Locations of the sampled population sites were chosen to largely represent the extent of the species' wild distribution across Eurasia (Tutin *et al.*, 1980; Hultén and Fries, 1985) (Table 5.2).

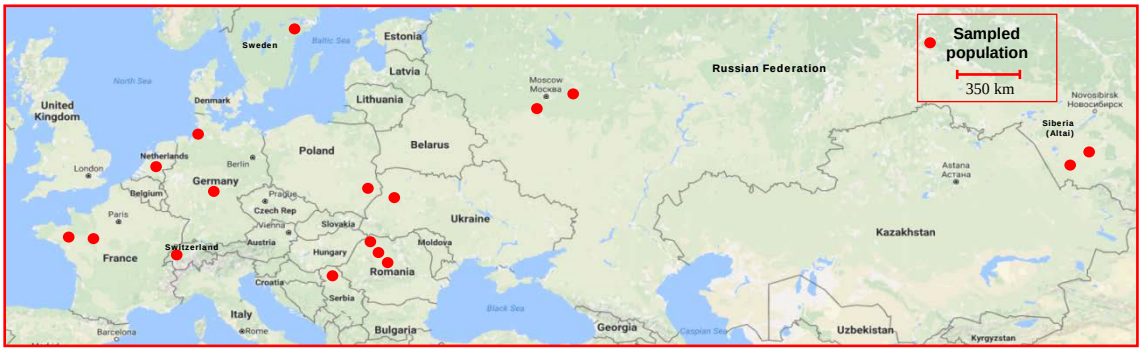

Figure 5.1 - Distribution of sampled *F. meleagris* populations in continental Eurasia. (Google Maps, 2017).

Table 5.2 - Sampled *Fritillaria meleagris* populations, host country, GPS site coordinates, collectors and affiliations (17 Continental and nine British sites (total = 26)).

	Country	Coordinates	Collector	Affiliation
Aubert Ings (Yorks)	England	53° 08' 48" N, 1° 18' 0" W	P. Day	Queen Mary Univ.
Ducklington Mead (Oxf. 1)	England	51° 46' 12" N, 1° 28' 12" W	P. Day	Queen Mary Univ.
Fox Meadow (Suffolk)	England	52° 12' 0" N, 1° 11' 60" E	P. Day	Queen Mary Univ.
Lugg Meadow (Hereford)	England	52° 04' 12" N, 2° 40' 48" W	P. Day	Queen Mary Univ.
Magdalen Meadow (Oxf. 2)	England	51° 45' 0" N, 1° 15' 0" W	P. Day	Queen Mary Univ.
Mottey Meadows (Staffs)	England	52° 43' 12" N, 2° 13' 48" W	P. Day	Queen Mary Univ.
North Meadow (Wilts)	England	51° 38' 60" N, 1° 52' 12" W	P. Day	Queen Mary Univ.
Portholme Meadow (Hunts)	England	52° 19' 12" N, 0° 10' 48" E	P. Day	Queen Mary Univ.
Stanford End (Hants/Berks)	England	51° 22' 12" N, 0° 58' 48" W	P. Day	Queen Mary Univ.
Bouchemaine - W. Loire	France (1)	47° 24' 36" N, 0° 37' 12" W	V. Malecot	Angers Univ.
Saumur - C. Loire	France (2)	47° 17' 24" N, 0° 07' 48" W	V. Malecot	Angers Univ.
Sinn Valley - Bavaria	Germany (1)	50° 04' 48" N, 9° 40' 12" E	G. Vogg	Würzburg Univ.
Elbe Marshes - Sc. Holstein	Germany (2)	53° 36' 11" N, 9° 37' 43" E	B. Netz	Elbemarschenhaus
Hasselt - Overijssel	Netherlands	52° 36' 0" N, 6° 04' 48" E	A. Corporaal	Personal contact
Kungsängen - Stockholm	Sweden	59° 49' 48" N, 17° 39' 3" E	A. Backlund	Uppsala Univ.
Les Brenets - Neuchâtel	Switzerland	47° 03' 48" N, 6° 41' 50" E	D. Tatti	Neuchâtel Univ.
Przemyśl - Subcarpathia	Poland	49° 46' 12" N, 22° 51' 0" E	E. Kuta	Jagiellonian Univ.
Poarta Salajului - Salaj	Romania (1)	47° 04' 48" N, 23° 13' 12" E	C. Sicora	Bio. Res. Cen., Jib.
Sardu - Cluj	Romania (2)	46° 53' 60" N, 23° 21' 0" E	C. Sicora	Bio. Res. Cen., Jib.
Chiesd - Salaj	Romania (3)	47° 23' 60" N, 22° 52' 12" E	C. Sicora	Bio. Res. Cen., Jib.
Obrenovac - Zabran	Serbia	44° 39' 36" N, 20° 13' 48" E	I. Sostaric	Belgrade Univ.
Novyi Kalyniv - Lviv	Ukraine	49° 34' 48" N, 23° 21' 0" E	E. Kuta	Jagiellonian Univ.
Choya - Altai - Siberia	Russ. Fed. (1)	52° 01' 48" N, 86° 35' 60" E	L. Hill	Personal contact
Cherga - Altai - Siberia	Russ. Fed. (2)	51° 31' 48" N, 85° 34' 48" E	L. Hill	Personal contact
Kremyonki - S.W. Moscow	Russ. Fed. (3)	54° 53' 60" N, 37° 06' 36" E	M.Popchenko	Timiriazev Agr. U.
Pichugino - E. Moscow	Russ. Fed. (4)	55° 22' 12" N, 41° 14' 24" E	I. Schanzer	Russ. Acad. Sci.

In Britain there are approximately 25 disjunct sites containing (putatively) naturally wild populations of *F. meleagris*. (K.J. Walker, pers. comm.; BRC, 2015). All are to be found in England, with Mottey Meadows recorded as geographically the furthest north known location (Mabey, 1996; M. Brown, pers. comm.). In early 2013, guidance was obtained from Kevin Walker, Head of Science at the then named Botanical Society of the British Isles with regard to the known long term health, flowering number history, flowering regularity and undisturbed site history of the most ancient of these populations. He also advised as to the most efficient method of obtaining the necessary site permissions to access and collect leaf samples and voucher specimens from the *F. meleagris* plants within the chosen populations. Seven sites were selected for their locations recorded as having supported undisturbed wild *F. meleagris* populations with a continuous annual presence of more than 50 flowering individuals for at least 125 years (Table 5.3). In order to provide the possibility of maximal natural biological

isolation, these sites were also selected for the separation afforded by their inter-population geographic distances. Two further locations (Portholme Meadow and Aubert Ings) with more recent first records of *F. meleagris* presence (1926 and 2000 respectively) (K.J. Walker, pers. comm.) were chosen for purposes of genetic comparison with the other sites (Table 5.3) (Figure 5.2).

In due course written permissions to access population sites and collect leaf and herbarium specimens were obtained directly in writing or email from the respective owners, site wardens and Natural England officers.

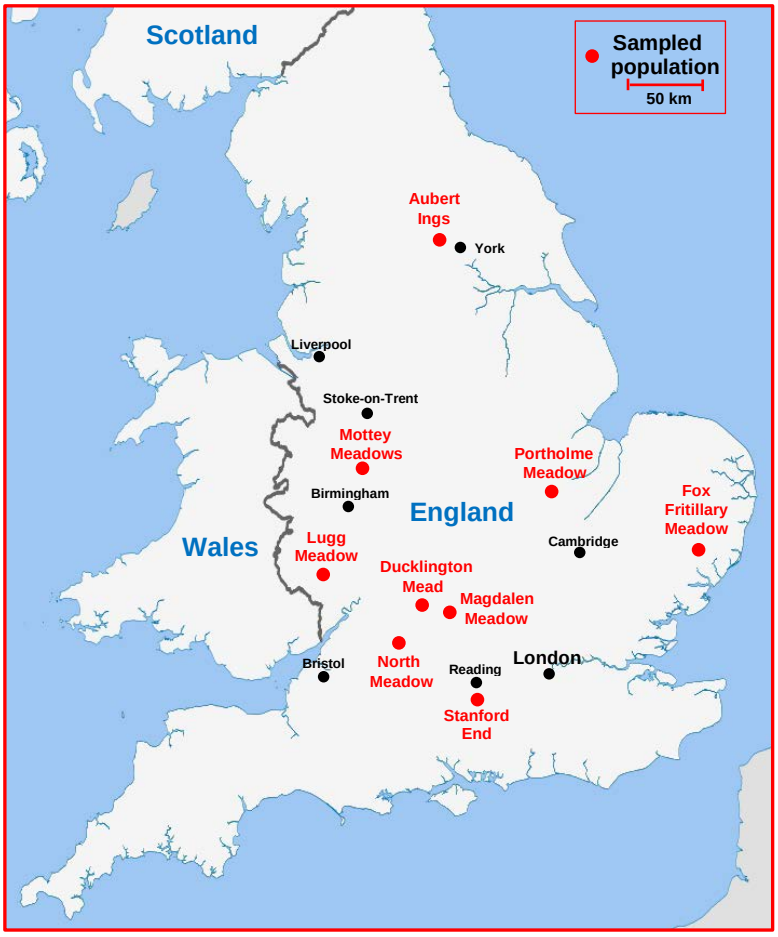

Figure 5.2 - Distribution of sampled *Fritillaria meleagris* populations in Britain. (www.d-maps.com, 2017).

Table 5.3 – Putative first site records for English *Fritillaria meleagris* populations.

Population	First site record for *F. meleagris*	Reference
Aubert Ings (Yorkshire) - SSSI/NNR	*c.* 2000	K.J. Walker, pers. comm.
Ducklington Mead (Oxfordshire 1) - SSSI	1886	Local publication (2014).
Fox Fritillary Meadow (Suffolk) - SSSI	1836	Sanford and Fisk (2010).
Lugg Meadow (Herefordshire) – SSSI	1805	Brian and Thompson (2002).
Magdalen Meadow (Oxfordshire 2) - SSSI	1785	Druce (1886).
Mottey Meadows (Staffordshire) - SSSI/NNR	*c.*1787	Mabey (1996).
North Meadow (Wiltshire) - SSI/NNR	*c.*1862	Pearman (2007).
Portholme Meadow (Huntingdonshire) - SSSI	1926	Doody (2008).
Stanford End (Hampshire/Berkshire borders) - SSSI	*c.*1802	Druce (1897).

* SSSI – Site of Special Scientific Interest; NNR – National Nature Reserve.

5.2.4 Collecting leaf sample material

Collection of leaf material and a field-pressed herbarium voucher of a stem, leaves and flowering head was undertaken personally for the British populations. For countries within continental Eurasia, collection was undertaken by the many botanists from local university, herbarium and conservation departments who were contacted personally by email and telephone to ascertain their willingness to act as collectors. As timing of collection was critical with regard to obtaining high quality material and for the project timescale as a whole, it was important to ensure individual collectors could identify *F. meleagris* plants in the field, locate their local *F. meleagris* population(s) and collect the requested specimens at the optimum time. This was ascertained in all cases.

Each collector gave most generously and freely of their time and expense to identify, preserve and return the material in an excellent state of preservation to the Jodrell Laboratory, RBG, Kew in readiness for processing (Table 5.2).

5.2.5 Sampling and sample collection strategy

Collectors were asked to provide sample material according to a sampling strategy personally devised, which followed a combination of recommendations and procedures described by Suzuki *et al*. (2004) and Hale *et al*. (2012), and was first tested personally for the sample collections in England.

At each population site, simple un-biased random sampling (Ward and Jasieniuk, 2009; Groom *et al*., 2011; Hale *et al*., 2012) was requested of one cut leaf taken mid-stem

(enough to obtain approximately 20 mg of dried leaf material) from each of 30 *F. meleagris* reproductive (flowering) individuals, each randomly chosen from each part of the populated site, maximally separated by ≥ 3 m distance to avoid problems of clonal clumping. One voucher herbarium specimen of a single *F. meleagris* stem, upper leaves and flowering head was also requested to be collected from each population and field-pressed.

Sampling took place between March and mid-May 2013 for all populations except those in Siberia (May 2015). Geographic (and altitudinal - not shown) co-ordinates were recorded by GPS from the centre of each population (Table 5.2) and a general site photograph taken (not shown). To ensure uniformity of genetic material and therefore results, leaf material was only to be collected from flowering reddish-purple form plants not the white/green variety (probably a phenotype). The samples were requested to be stored in individually labelled sealable plastic storage bags, each containing approximately 25 g silica-gel (SiO_2) preservative, approximately 600 of which were individually filled personally at RBG, Kew, of which 35 were despatched to each collector. Sampling instructions and procedures were provided for each collector, which also contained a request to store and keep the voucher specimen in the field press within layers of newspaper for two weeks - the press to be kept at room temperature to facilitate thorough desiccation. This would ensure the material would be structurally stable before returning it with the collected leaf samples to RBG, Kew for DNA extraction and subsequent analysis.

5.2.6 Completion of DNA extractions and genotype recovery

The procedures described in detail in Chapter 4 addressed processing of the test panel of 26 population samples for leaf DNA isolation and microsatellite genotype recovery and scoring. The same procedures were now extended to include four additional individuals from each of the 26 populations, bringing the total to n=5 individuals from each of the populations.

5.2.7 Validation of collected leaf and voucher samples

A DNA sequence matching procedure was conducted against vouchered specimens of

F. meleagris and *F. meleagroides* to ensure collected leaf samples had each been taken from plants of *F. meleagris* origin. Particular care was taken to eliminate the possibility that leaf samples had been collected in error from *F. meleagroides* rather than *F. meleagris* populations in the Russian Federation, Romania and Ukraine, as the two species look very similar to the relatively untrained eye (pers. obs.) and are known to have partially overlapping ranges, but residing in separate populations within these countries (Hultén and Fries, 1985) .

DNA of one individual taken from each of the 26 populations was sequenced for a 1.6 kb portion of the plastid gene *mat*K intron following the procedure described in Day *et al*. (2014). The same procedure was conducted for DNA samples of *F. meleagris* and *F. meleagroides* held within the DNA bank at RBG, Kew (http://apps.kew.org/dnabank/homepage.html) respectively (Hill, L. 864; acc. no. 34219 and RBGKew LivColl. 2005-2045; acc. no. 24371). Contigs (contiguous consensus sequences) were assembled for all sequences, then edited using the program SEQUENCHER™ version 4.5 (Gene Codes, Ann Arbor, MI, USA) and manually aligned by the program PAUP* v4.0b10 (Swofford, 2003) following the procedures described in Day *et al*. (2014), ready for comparison with the vouchered reference sequences.

5.3 Descriptive and inferential statistics

Before using genotype data for a population genetics analysis study, evaluation of the descriptive statistics most relevant to a proposed downstream analysis is necessary for assessment of data quality and its fitness for purpose. If this process is successful, inferential statistics are used to analyse a sample of population data to make predictions about a larger set of population data from which the samples have been drawn. For this project, inferences are made about the 26 populations under study based on the sample of data taken from these populations. From this position it is expected that judgements can confidently be made about whether a difference observed between populations can be depended upon or has happened by chance alone. This is achieved by the testing of statistical hypotheses.

5.3.1 Hardy-Weinberg genotypic equilibrium (HWE)

Population genetic analysis is based on a set of genotypes contained within multiple loci, and in the case of this project, multiple populations. As a standard test, each locus was tested for compliance with HWE expectations for all populations (*i.e.* the expected numbers of homo/heterozygote genotypes at each locus, given the observed allele frequencies) (Beebee and Rowe, 2008). Deviations from HWE might indicate effects caused by inbreeding, within-population fragmentation (*i.e.* sub-populations within a population), migration, mutation or natural selection.

Departure from HWE was tested by using the Markov Chain 'exact HW test' method of Guo and Thompson (1992) modified by the Metropolis-Hastings normalised likelihood algorithm (Hastings, 1970; Goldstein and Schlötterer, 1999) and implemented in program GENEPOP v4.5 (Rousset, 2008; Engels, 2009) ('exact' means a test without the imprecise parameter values utilised by, for example, the χ^2 goodness-of-fit test (Wigginton *et al*., 2005) - and therefore precise probabilities can be calculated (Rousset and Raymond, 1997) and is preferred when the sample sizes are small (Guo and Thompson, 1992). For Markov Chain (MC) parameters, the number of dememorisations was set at the recommended 100,000 (the number of chain steps or switches of allelic pairs performed before comparing alternative contingency table probabilities to that of the observed two-way contingency table – equivalent to a 'burn-in' period during which a starting point is reached corresponding to a table independent from the observed table). The recommended default number of batches of further MC *p* value estimate simulations was set to 100 and maximum steps in the Markov chain for exploration of alternative tables set to 5,000. Rather than assume these settings were appropriate for the data presented, three sets of increasingly large numbers for all three parameters were tested (x2, x5 and x10 respectively) to test for evidence of change in results.

5.3.2 Linkage disequilibrium (LD)

A standard method for establishing the presence of linkage disequilibrium between pairs of neutral loci is to test for the significant association evident when the alleles segregate together (Beebee and Rowe, 2008). Alleles at different loci are expected to be randomly associated in a large mating population at equilibrium. If they are not, then one possible

explanation is that the two loci have physical links on the same chromosome enough that recombination events have occurred between them during reproduction. The state of LD within a population can be affected by the effects of bottlenecks, admixture, inbreeding or natural selection (Beebee and Rowe, 2008; Frankham *et al*., 2010).

The presence of LD between pairs of loci was tested using a likelihood-ratio test (Lewontin and Kojima, 1960; Slatkin and Excoffier, 1996), rather than an exact test based on MC, because knowledge of which alleles appear together on the same chromosome (gametic phase) is unknown and can only be estimated (Excoffier and Lischer, 2010).

Program ARLEQUIN v3.5.2.2 (Excoffier and Lischer, 2010) compares the likelihood of the sample evaluated under the null hypothesis of 'no association between loci (linkage equilibrium)' to the likelihood of the sample when association is allowed (Slatkin and Excoffier, 1996). Under the HWE hypothesis, the significance of the observed likelihood ratio is calculated by computing the probability (null) distribution of the ratio using a permutation procedure. Recommended parameter values (Slatkin and Excoffier, 1996) were used as follows: for the number of permutations, 16,000, which guarantees to have <1% difference with the exact probability in 99% of cases (Guo and Thompson, 1992)); for the number of random initial conditions, 3, from which the Expectation-Maximisation (EM) algorithm is started for repeated estimates of sample likelihood; and the significance level ($p = 0.05$) against which the test of LD is measured.

5.3.3 Null alleles, large allele dropout and private alleles

The presence of null alleles (non-amplifying alleles in PCR) is usually caused by nucleotide point mutations or indels in one or both microsatellite flanking regions, producing disruption at primer binding (annealing) sites and therefore loss of alleles during amplification. Modifications to the 3´ end of the site where primer extension begins are thought to be particularly detrimental to PCR amplification (Kwok *et al.*, 1990; Lehmann *et al.*, 1996; Paetku and Strobeck, 1995; Dakin and Avise, 2004). The effect is to reduce accuracy in recovery of genotypes and, depending on percentage of null alleles present, potentially serious statistical errors in analyses (Bonin *et al.*, 2004; Hoffman and Amos, 2005).

Null allele presence can also be caused by large allele dropout. This is the result of preferential amplification of the smaller (lower molecular weight) allele in a heterozygote during early stages of PCR. The effect may be sufficiently extreme that the much larger peak might be scored as homozygous for the visible allele rather than as a component of a possible heterozygote pair (Callen *et al.*, 1993; Jones *et al.*, 1998). This problem can cause a reduction in observed heterozygosity, with concomitant adverse statistical effects during subsequent analyses (Wang *et al*., 2012).

The potential presence of null alleles might also be implied by significant heterozygote deficits relative to HWE, the origin of which can be caused by the Wahlund effect (reduction in the overall heterozygosity of a population as a result of subpopulation structures, *e.g.* from genotype data from multiple (sub) populations pooled for analysis), inbreeding or selection at or near a microsatellite locus (Dakin and Avise, 2004). However, these causes can be distinguished from the presence of genuine null alleles, as they are usually found to be consistent across all loci, whereas null alleles are locus-specific.

Program GENEPOP was used to identify potential occurrences of null alleles by application of the expectation–maximisation (EM) algorithm to obtain maximum-likelihood estimates of null allele frequency (Dempster *et al*., 1977; Kalinowski and Taper, 2006; Hartl and Clark, 2007). Chapuis and Estoup (2007) assessed this method for estimating null allele frequencies as the best of three standard approaches, including those of Chakraborty *et al*. (1992) and Brookfield (1996). Within the computation, sequential Bonferroni corrections are applied to correct for multiple serial comparisons (Benjamini and Hochberg, 1995).

Private alleles are those that are found only in a single population among a broader set of populations, and may be indicators of mutation (drift) rate and divergence time between populations (Szpiech and Rosenberg, 2011). The Microsoft Excel add-in program GENALEX v6.5 (Peakall and Smouse, 2012) was used to ascertain the presence and level of private alleles contained within individual populations.

Separate studies undertaken by Dakin and Avise (2004) and Guichoux *et al*. (2011) of >300 original articles published between 2001-2010 in the journals *Molecular Ecology* and *Molecular Ecology Notes*, discovered that explicit testing for the presence of null

alleles was reported in only 40% of the research, with 90% of these taking no further action - thereby casting doubt on the accuracy of the work.

5.3.4 Locus heterozygosity and polymorphism

The programs FSTAT v2.9.3.2 (Goudet, 2002) and POPULATIONS v1.2.32 (Langella, 2012) were used to determine the observed heterozygosity (H_o) and expected heterozygosity (H_e) of the 19 loci, estimated across all populations.

Program CERVUS v3.0.7 (Marshall *et al*., 1998) was used to calculate mean polymorphic information content (PIC). PIC is a measure of the relative informativeness of a locus for purposes of population genetics analysis, and depends on the number of alleles at the locus and their relative frequency. The PIC value will be approaching zero if there is no allelic variation. For a multiallelic locus a PIC value ≥ 0.5 is deemed highly informative with regard to its utility for population genetics analysis (Botstein *et al.*, 1980; Anderson *et al.*, 1993).

5.3.5 Population bottlenecks

If a population size has been subject to a bottleneck reduction, and mutation-drift equilibrium has been temporarily disrupted, the measured Hardy-Weinberg expected heterozygosity (H_e) will be significantly greater than, rather than nearly equal to, the expected equilibrium heterozygosity computed from the number of alleles in a population sample (H_{eq}) (Nei *et al.*, 1975, Watterson, 1984; Maruyama and Fuerst, 1985). This transient 'heterozygosity excess' relative to that expected in a population subject to genetic drift and constant size and number of alleles, is due to the likely loss of low frequency (rare) alleles (Nei *et al*., 1975; Watterson, 1984; Cornuet and Luikart, 1996; Luikart and Cornuet, 1998; Chiucchi and Gibbs, 2010).

The program BOTTLENECK v1.2.02 (Piry *et al.*, 1999) was used to test for evidence of past population bottleneck or expansion events. This is achieved by testing for excess allelic diversity relative to that expected under selective neutrality and mutation-drift equilibrium. BOTTLENECK simulates the coalescence process (Cornuet and Luikart, 1996) using three models: (1) for loci evolving under the stepwise model of mutation (SMM - specific to microsatellites), where mutations occur at a constant rate and gain or lose a single repeat unit to create a novel allele (Ohta and Kimura, 1973; Selkoe and

Toonen, 2006); (2) the infinite allele model of mutation (IAM), in which mutations occur at a constant rate with each creating a unique allele (Kimura and Crow, 1964);

Balloux and Lugon-Moulin, 2002); or (3) an intermediate two-phase model (TPM), by which most microsatellite loci probably evolve, with the great majority of mutations as single repeat gains or losses and the remainder as multiple base-pair repeat mutations. (Chakraborty and Jin, 1992; Di Rienzo *et al*., 1994; Luikart and Cornuet, 1998; Estoup and Cornuet, 1999). The TPM model was chosen with which to proceed.

The parameters recommended by Piry *et al.* (1999) for testing TPM mutation were selected as 1,000 iterations, with 95% single-step mutations, 5% multiple-step mutations drawn from a geometric distribution, and a variance among multiple steps of approximately 12. Using the Wilcoxon signed-rank test for $H_e > H_{eq}$ (Piry *et al.*, 1999) and standardised difference test recommended by Cornuet and Luikart (1996) for analysing approximately ≥ 20 polymorphic loci, the probability of bottleneck presence was calculated from the observed number of alleles and sample size of individuals.

5.3.6 Principal coordinates analysis (PCoA)

Principal Coordinates Analysis is a statistical method of calculating the simultaneous variation (covariance) of two or more dependent variables, and presenting the results as patterns of genetic distance relationships. For this project the dependent variables are multiple samples (within multiple populations) across multiple loci. The rationale behind PCoA is to exploit inter-relationships between the variables, which are individually redefined through standard transformation (standardisation) as linear correlations between genetic distances in a distance matrix (Fowler *et al*., 1998). This technique enables graphical representation of the principal axes as plots, the first of which explains the most variation, with each then successively explaining proportionately less of the total variation within the data. When there are distinct groupings of samples, the first three axes typically reveal the majority of genetic separation between them (Sokal and Rohlf, 1995; Peakall and Smouse, 2012).

Using program GENALEX, a broad characterisation of genetic diversity among 826 genotypes between populations was conducted using PCoA analysis to determine evidence of population group structure. The procedure uses a standardised covariance distance matrix based on a multivariate analysis algorithm published by Orloci (1978). As data input, a triangular matrix of genotype individual-by-individual pairwise

Euclidean genetic distances was created within GENALEX from an Excel table containing genotypes recovered from the 19 loci (see section 5.2.6). As Euclidian distances are used within the computation, the two techniques of PCoA and Principal Components Analysis become identical (Palmer, 2008).

5.3.7 Mantel test

The Wright (1943) hypothesis states that population pairs more distant from each other are increasingly differentiated because of 'Isolation-by-Distance' (IBD). This hypothesis was tested by using the non-parametric Mantel test to spatially analyse the degree of correlation between genetic dissimilarity and geographic distance among populations (Mantel 1967; Slatkin, 1993; Rousset, 1997; Hutchinson and Templeton, 1999).

The Mantel test is by far the most commonly used method to evaluate the effect of geographic space on genetic data (Diniz-Filho *et al*., 2013), and has been in popular use for many years in ecological and population genetics research related to IBD studies (Reusch *et al.*, 2000; MacDougall-Shackleton and MacDougall-Shackelton, 2001; Lloyd, 2003) and continues to be so (Palma-Silva *et al*., 2009; Wilke *et al*., 2014). However, there have been recent criticisms as to its statistical performance (Harmon and Glor, 2010; Legendre and Fortin, 2010; Guillot and Rousset, 2013). The review by Diniz-Filho *et al.* (2013) shows that careful application and interpretation of Mantel test results, particularly in complex multivariate relationships, can overcome some of these potential statistical problems. The Mantel test remains a simple and extremely useful tool for testing matrix correlation and analysis of IBD, primarily if the ecological or evolutionary hypotheses are expressed as pairwise distances or similarities (Legendre and Fortin, 2010) - which is the case for this project's data.

The Mantel test implemented within program GENALEX was used to assess the relationship between the individual elements of a lower triangular matrix of pairwise genetic distances of 1,652 alleles (826 genotypes), with a corresponding lower triangular matrix containing manually entered individual population geographic coordinates. Nei (1972) determined 'genetic distance' (*D)* to be the accumulated allele differences per locus; and his assumption was that if the rate of gene substitution in neutral loci per year is assumed to be constant, it is linearly related to the divergence time between populations under reproductive isolation and to geographic distance. Nei's

measure *D* (Nei, 1972) is one of the most widely used for estimating genetic distance among populations (Peakall and Smouse, 2012).

Each row of population alleles within the matrix is treated as a single unit rather than assuming each allele genetic/geographic distance pair to be independent (the $N(N\text{-}1)$ pairwise elements of an (N x N) matrix cannot be mutually independent, as there can be no more than (N - 1) independent comparisons among a set of N elements). GENALEX plots a graph of the relationship between the elements of the two matrices and applies regression analysis to estimate and plot a trend line and determine its slope and intercept. How well the data points fit the regression line is indicated by the Mantel correlation coefficient (*Z*), which is a statistical measure of how closely the two matrices correspond. To test the significance of *Z*, the cross-products of the corresponding off-diagonal elements of the two matrices are summed within GENALEX to create an 'observed' *Z*-value. Random permutation of the rows and columns of the geographic-distance matrix is then computed whilst maintaining the genetic-distance matrix as constant. This is repeated 999 times (Peakall and Smouse, 2012) with recalculation of the *Z*-value at each iteration. This establishes a probability reference distribution of *Z*-values.

The null hypothesis states there is no significant relationship between the two matrix datasets. If this is true, random permutations of the data should produce *Z*-values similar to the observed *Z*-value. However, if there is correlation between the two datasets, these *Z*-values will be more extreme (*i.e*. closer to +1 or -1). To test this, the observed *Z*-value is compared with the reference *Z*-value frequency distribution to determine the level of statistical significance (*P*). The value for *P* of the observed *Z*-value is calculated from the cumulative tail of the frequency distribution evaluated as $\frac{(n_T+1)}{(N+1)}$, where n_T is the number of replicated randomised *Z*-values that are ≥ the observed *Z*-value (Sokal and Rohlf, 1995) and *N* is the number of randomised permutations.

5.3.8 STRUCTURE analysis

Genetically structured populations frequently can be viewed as a set of discrete clusters of individuals; alleles in each population having distinctive frequencies (Rosenberg, 2004).

The program STRUCTURE v2.3.4 (Pritchard *et al*., 2000) places individuals into genetic clusters based on their allele frequencies, and is probably the most widely used population analysis tool available with which to assess patterns of genetic structure within multi-locus allele frequency data. In the 10 years following creation of STRUCTURE in the year 2000, more than 3,000 population genetics papers cited the program (Kalinowski, 2011), and to 2015, more than 11,500 citations were to be found in ISI Web of Science (Gilbert, 2016).

STRUCTURE estimates recent ancestry proportions to assign individuals (from 26 populations in this case) to different genetic groups (parental populations) within clusters (*K*), which are the uppermost level of hierarchical structure. Functionally, *K* is a parameter that can be varied by the user in multiple independent STRUCTURE runs in construction of a statistical dataset, post hoc analysis of which is used to infer the number of putative clusters matching the data presented (Schwartz and McKelvey, 2009; Kalinowski, 2011; Earl and vonHoldt, 2012; Porras-Hurtado, 2013; Gilbert, 2016).

The program is parameterised to ensure accommodation of the most commonly used population genetics ancestry and allele frequency models (Porras-Hurtado, 2013), and requires as input allelic data specifically structured as a matrix. The latter was achieved by using the Excel spreadsheet data formatting and exporting facilities offered within program GENALEX.

The results from STRUCTURE analyses can be used by a large suite of freely available supporting programs for the creation of graphical output and statistical data for use in further analysis.

STRUCTURE detects and analyses allele frequency variations in a set of genotype samples, from which subsets of the data are identified. Based on a Bayesian analysis clustering approach, Markov Chain Monte Carlo (MCMC) estimation is applied to determine the best fit for the allele frequency variation, resulting in individuals assigned to one or more of *K* population clusters. The MCMC estimation starts by randomly assigning individuals to the *K* number of clusters. Allele variant frequencies are estimated in each cluster and individuals possibly re-assigned to another cluster based on those estimates. Initially, this process is iterated many times as a burn-in process (the

permutations executed before the MCMC calculated data are recorded) in order to reach approximate stationarity (convergence), thereby enabling reliable allele frequency estimates to be obtained subsequently for each cluster and membership probabilities of individuals within clusters ascertained.

Posterior probabilities of *K* are generated using a measurement model of MCMC estimation to calculate the membership likelihood of the data for a range of *K* values. This calculation is performed separately from the burn-in process and is applied to subsets of the previously identified allele frequency estimates. Kalinowski (2011) notes that the most realistic value of *K* is achieved by obtaining the smallest value of *K* that maximises the global likelihood of the data. This is achieved by the strategy of focusing analysis on a *K* value that captures the majority of the structure present in the data without overestimating it, but which should also be a value which can be inferred to be biologically and geographically reasonable. Individuals are given an assigned allele membership coefficient (AAMC) *Q* (probability of membership) for each cluster, such that total membership for each individual sums to 1 across *K* clusters. Hence individuals are always found to be a member of one or more populations within clusters. The AAMC of an individual for a cluster represents the fraction of the genome that has ancestry in the cluster.

Parameterisation available within STRUCTURE enables various assumptions to be tested to ascertain the effects of different model approaches to ancestry and allele frequency.

Ancestry models are available to test for *no admixture* or *admixture* of individuals between clusters. The *no admixture* model assumes each individual has its origin in only one cluster, and STRUCTURE calculates the posterior probability of the individual belonging to a particular cluster. This model is also more successful at detecting subtle population structure differentiation (Pritchard *et al*., 2010).

Alternatively, the *admixture* model assumes that individuals can have admixed ancestry, such that each can inherit a fraction of their genome from ancestors in other clusters. The no *admixture* model was chosen for testing as the most appropriate, as it was assumed that the distances and geography separating each of the 26 populations (Figure 5.1) inferred biological isolation of each population from another was highly probable (For E. Europe, the populations of Romania 1 and Romania 2 are closest in proximity at

22 km. The Russian Federation populations in the Moscow and Siberia regions are separated by 260 km and 75 km respectively. Within England, the two populations with closest proximity are those in Oxfordshire, separated by 16 km; all other English populations are located at least 30 km apart).

Two different allele frequency models are available within STRUCTURE. The *independent allele frequencies* model assumes there is no correlation of allele frequencies across clusters for the same loci, and also that ancestral relationships are not expected between clusters (Rosenberg *et al.*, 2005). In consequence, the model expects allele frequencies to show some differentiation between distinct clusters. In contrast, the *correlated allele frequencies* model assumes each of the clusters underwent drift from a common ancestral cluster, with therefore different clusters having correlated frequencies due to shared ancestry (Rosenberg *et al*., 2005) and closely related clusters exhibiting similar allele frequencies (Falush *et al*., 2003). However, Porras-Hurtado *et al*. (2013) observed that STRUCTURE has the power to detect distinct cluster affinities only when they are particularly closely related, and if high levels of correlation across clusters are absent, both allele frequency models will provide the same results (Rosenberg *et al*., 2005). In consequence, the *independent allele frequencies* model was chosen as the most appropriate with which to proceed, as variation from the assumption that biological isolation exists between populations could most easily be determined, and the model can produce the most accurate clustering (Pritchard *et al.*, 2000).

As STRUCTURE runs are stochastic, they have a propensity to produce different likelihood values for replicate runs due to 'label switching' across replicates, even when the same choice of parameters and models is used (different replicates can obtain the same AAMC estimates, but with an arbitrary permutation of the way inferred clusters are labelled (Stephens, 2000; Jasra *et al.*, 2005)). To address this problem, 15 iterations were carried out for each value of *K* for the selected parameter and model pairing (*no admixture* and *independent allele frequencies*) in order to more accurately quantify the log-likelihood variation. In addition, for each iteration the STRUCTURE parameter *LOCPRIOR* was set to indicate the presence of sampling location information (population identifiers provided as input data), as this is particularly informative with regard to population structure when there are small sample sizes, but allows for a small probability that the population contains immigrants or contains immigrant ancestry (Pritchard *et al.*, 2000; Porras-Hurtado *et al*., 2013). The parameter Lambda (λ) was set

to 1.0 to ensure an even distribution of independent draws of allele frequencies from each population. All other parameters were set to their default values as advised in the STRUCTURE v2.3 user manual (Pritchard *et al*., 2010).

Using results returned by STRUCTURE, identification of the value of *K* that best fitted the data was achieved using the program STRUCTURE HARVESTER v0.6.94 (Earl and vonHoldt, 2012). If there is hierarchical structure in the studied populations, the log-likelihoods estimated by STRUCTURE do not necessarily reflect the real number of *K* clusters, as likelihood values may increase with stepwise values of *K* (Evanno *et al.*, 2005; Pritchard *et al*., 2010). The estimate of the most likely value of *K* was achieved therefore by employing a statistical method proposed by Evanno *et al*. (2005), and implemented within STRUCTURE HARVESTER, based on the point of largest second order rate of change in the likelihood of data between successive values of *K* (the *ad hoc* statistic ΔK). The results were visualised as a plot of the mean likelihood of data for each value of *K*.

Results additionally retrieved from STRUCTURE HARVESTER contained in an individual *population Q-matrix* data set for each value of *K* for further processing, consisted of the calculated AAMC and ancestry components for each of the pre-defined populations (26 in this case), determined for each iteration for each value of *K*.

The stochastic effects of multiple iterations for each value of *K* were addressed by using the *population Q-matrices* data files as input to the program CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007). The program aligns replicated STRUCTURE runs to maximise individuals' membership across clusters by calculating medians of the consolidated individual AAMC and population ancestry components. The amalgamated results are consolidated into the individual *population Q-matrix* for each *K* value for further analysis. Correlation level among different runs was assessed using the similarity coefficient *h´* produced using the *FullSearch* algorithm implemented in the program. The program DISTRUCT (Rosenberg, 2004) was used subsequently to analyse the *population Q-matrix* data to produce bar plots of population and cluster structure in segmented columns for each value of *K*. The graphical output was created from a data file retrieved from DISTRUCT, from which the Postscript file graphical interface program GSVIEW v5.0 (Lang, 2009), in conjunction with interpreter program GHOSTSCRIPT v9.21 (Deutsch, 1988), were used to visualise the results.

5.3.9 Software programs used for population structure analysis

The statements made in section 4.2.14 pertain in equal measure to the multiple software programs utilised to achieve the results described in this Chapter. With the exception of GENALEX, the lack of documentation, presence of execution problems, and sometimes even rudimentary explanation of the methods and principles employed by individual software packages, caused serious set backs at various times during attempts to achieve the desired results from the software.

5.4 Results

5.4.1 Population sizes

Data obtained, some anecdotally, from the individual collectors of leaf and herbarium voucher samples, and personally observed within the English populations, indicated the size of populations sampled ranged from approximately 250 to approximately 100,000 flowering heads, with the majority of populations supporting about 3,000 within the boundaries of each of the targeted sites. In England, the lowest number encountered was approximately 50 at Mottey Meadows in Staffordshire. However, in the previous several years more than 350 had been regularly reported annually at the site by the local Natural England Reserves Manager (M. Brown, pers. comm.).

5.4.2 Validation of collected samples to be of *F. meleagris* origin

Following the procedures described in section 5.2.7, results confirmed all leaf samples to be of *F. meleagris* species origin. All herbarium vouchers for the 9 English populations and those returned by collectors from the 17 continental Eurasian populations were personally verified visually as flowering heads and stems taken from *F. meleagris* plants.

5.4.3 Leaf sample DNA isolation and genotype recovery

From the leaf samples of 130 individuals (five from each of 26 populations) which underwent the process of DNA isolation and subsequent amplification in PCR, 15 failed to amplify for any locus and were eliminated from further analysis. Recovery of 826

genotypes was achieved from a possible maximum of 2,185 (38% - 115 samples x 19 loci) (Table 5.4).

The high proportion of genotypes remaining unrecovered during development (62%) reflects the substantial difficulties encountered in allele identification and measurement described in section 4.2.11. However, as a result of the careful and conservative approach maintained during microsatellite development, a high level of confidence is placed in the quality of genotype data successfully recovered ready for downstream analysis.

Table 5.4 - Characteristics of 19 microsatellite loci developed for *Fritillaria meleagris* and tested for a total of 115 individuals collected from 26 Eurasian populations.

Locus	No. alleles	Geno-types	Allele size range (bp)	*Ho*	*He*	HWE	PIC	Null alleles	B/neck evidence	ProbeDB acc. no.
L4	12	35	249 - 277	0.34	0.9	*	0.88	0.03	No	Pr032826426
L6	33	47	264 - 346	0.51	0.97	*	0.95	0.03	No	Pr032826437
L11	8	41	227 - 245	0.85	0.76	*	0.71	0	No	Pr032826438
L12	8	30	259 - 279	0.5	0.73	**	0.68	0.11	No	Pr032826439
L13	17	52	229 - 265	0.36	0.92	*	0.9	0.03	No	Pr032826440
L14	19	45	237 - 273	0.6	0.93	***	0.92	0.1	No	Pr032826441
L15	13	53	244 - 278	0.68	0.77	NS	0.74	0.02	No	Pr032826442
L16	10	36	141 - 171	0.97	0.89	NS	0.87	0.05	No	Pr032826443
L21	10	37	248 - 272	0.3	0.84	**	0.82	0.03	No	Pr032826444
L23	21	50	166 - 244	0.62	0.89	NS	0.88	0.03	No	Pr032826427
L25	20	60	190 - 234	0.42	0.94	*	0.93	0.04	No	Pr032826428
L27	8	39	158 - 179	0.79	0.8	NS	0.76	0.06	No	Pr032826429
L28	10	33	236 - 266	0.64	0.74	NS	0.69	0.02	No	Pr032826430
L32	13	46	237 - 261	0.61	0.91	NS	0.89	0.03	No	Pr032826431
L34	14	41	195 - 221	0.54	0.82	*	0.8	0.02	No	Pr032826432
L35	9	41	140 - 156	0.66	0.8	NS	0.76	0.09	No	Pr032826433
L36	11	33	171 - 191	0.67	0.88	NS	0.85	0.12	No	Pr032826434
L38	7	37	105 - 119	0.35	0.75	**	0.7	0.03	No	Pr032826435
L39	20	70	149 - 195	0.61	0.88	*	0.86	0.02	No	Pr032826436

Allele total; genotype total; scored allele sizes; observed heterozygosity (*Ho*); expected heterozygosity (*He*); HWE deviation: NS not significant, * significance at $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$; polymorphism information content (PIC) value; % frequency of null alleles; evidence of population bottleneck; NCBI archive primer pair sequence accession numbers.

5.4.4 Hardy-Weinberg equilibrium

Results from tests conducted with program GENEPOP indicated evidence of departure from Hardy-Weinberg equilibrium in 11 loci (Table 5.4). Testing of three sets of increasingly large MC parameter numbers (x2, x5 and x10 respectively) for comparison of results showed no discernible change in observed outcomes (see section 5.3.1).

When compared to HWE expectations, population fragmentation (subpopulation structures) can result in the creation of spatial genetic structure through loss of allelic richness (Buza *et al*., 2000; Shea and Furnier, 2002), and a significant heterozygote deficiency at a locus for the entire population - the origin of which can be caused by the Wahlund effect (reduction in the overall heterozygosity of a population caused by drift, inbreeding depression within isolated (sub) populations and possible selection at or near a microsatellite locus (Dakin and Avise, 2004)). For this project, genotype data from 115 samples was pooled for analysis from 26 populations for each of 19 loci, thereby enabling analysis of the population structure revealed (see Figure 5.3, section 5.4.9). With the exception of loci L11 and L16, heterozygote deficiency (homozygote excess) was evident within all loci (Table 5.4), with 11 loci showing significant deviation from HWE (Table 5.4), indicating probable evidence of the Wahlund effect. However, low numbers of individuals present for each population (average = 4.4) might cause statistical bias and affect results, although this is unlikely to be significant with such a high proportion (89%) of loci displaying heterozygote deficiency.

5.4.5 Linkage disequilibrium

Using a likelihood-ratio test, the program ARLEQUIN found the number of linked loci per locus to be zero for all 19 loci, indicating no evidence for linkage disequilibrium between locus pairs.

5.4.6 Alleles, private alleles and null alleles

Program GENALEX revealed a total of 263 unique alleles recovered for the 26 populations, with total allele numbers varying between 7 at locus L38 to 33 at locus L6 (mean = 14) (Table 5.4). Private alleles accounted for 69 (26.2%) of unique alleles and 8.4% of total alleles (826) recovered (private alleles are exclusive to a single population within a broader collection of populations (clusters in this case) (Table 5.5) (Spiech and Rosenberg, 2011). Private allele presence within populations was found to generally follow a geographic west to east cline of increasing numbers (Table 5.5).

Evidence of null allele presence was found by program GENEPOP for loci L12 (11%), L14 (10%) and L36 (12%) (Table 5.4). The individual percentages of null allele presence for these loci were assessed as not statistically significant with respect to the effect on the individual loci or within the overall number of loci recovered (Huang *et al*., 2016), and were therefore retained for further downstream analysis.

Table 5.5 - Private alleles by population, location and population cluster.

Population	Country	Private alleles	Cluster
Aubert Ings (Yorks)	England	0	**A**
Ducklington Mead (Oxf. 1)	England	1	**A**
Fox Meadow (Suffolk)	England	2	**A**
Lugg Meadow (Hereford)	England	0	**A**
Magdalen Meadow (Oxf. 2)	England	3	**A**
Mottey Meadows (Staffs)	England	1	**A**
North Meadow (Wilts)	England	1	**A**
Portholme Meadow (Hunts)	England	0	**A**
Stanford End (Hants/Berks)	England	1	**A**
Bouchemaine - W. Loire	France (1)	1	**A**
Saumur - C. Loire	France (2)	0	**A**
Sinn Valley - Bavaria	Germany (1)	2	**A**
Elbe Marshes - Schl. Holstein	Germany (2)	2	**A**
Hasselt - Overijssel	Netherlands	1	**A**
Kungsängen - Stockholm	Sweden	0	**A**
Les Brenets - Neuchâtel	Switzerland	2	**A**
Przemyśl - Subcarpathia	Poland	1	**B**
Poarta Salajului - Salaj	Romania (1)	3	**B**
Sardu - Cluj	Romania (2)	7	**B**
Chiesd - Salaj	Romania (3)	4	**B**
Obrenovac - Zabran	Serbia	5	**B**
Novyi Kalyniv - Lviv	Ukraine	4	**B**
Choya - Altai - Siberia	Russ. Fed. (1)	9	**C**
Cherga - Altai - Siberia	Russ. Fed. (2)	7	**C**
Kremyonki - S.W. Moscow	Russ. Fed. (3)	9	**C**
Pichugino - E. Moscow	Russ. Fed. (4)	3	**C**

Colour coding reflects population membership within clusters defined in Figure 5.3.

5.4.7 Locus heterozygosity and polymorphism

The recovery of 826 genotypes resulted in 477 (58%) heterozygous for the 19 loci, with observed heterozygosity (H_o) ranging from 0.3 at locus L21 to 0.97 at locus L16 (mean = 0.58), and expected heterozygosity (H_e) ranging from 0.23 at locus L21 to 0.6 at locus L39 (mean = 0.41) (Table 5.4). All loci displayed high levels of polymorphism with calculated PIC values ranging from 0.68 for locus L12 to 0.95 for locus L6 (mean across all loci = 0.82 (Table 5.4)). A figure $\geq$ 0.5 indicates a level of locus polymorphism highly informative for population genetics studies (see section 5.3.4).

5.4.8 Population bottlenecks

Program BOTTLENECK provided no evidence of bottlenecks having occurred in any of the 26 populations (Table 5.4). However, the relatively low number of individuals present for each population, albeit counterbalanced by a high number of polymorphic loci (19) available for the analysis, might cast some doubt on accuracy of the results.

BOTTLENECK documentation states that a minimum of four polymorphic loci are required for successful analysis, but that inaccuracies can occur with small numbers of population individuals.

5.4.9 Principal coordinates analysis

Principal coordinates 1 and 2 (17.8% and 12.6% respectively - cumulatively 30.4%) of the PCoA scatter plot (Figure 5.3) explain the majority of genetic variation found among the 826 genotypes recovered from 19 loci representing 115 *F. meleagris* individuals from 26 Eurasian populations. The third axis, principal coordinate 3 (not shown), explains a further 8.8% of genetic diversity (cumulatively 39.2%). The population groupings recovered from the two major axes show there to be three distinct centres (clusters) of genetic diversity, marked A – C in Figure 5.3: cluster A contains populations from England, N. Europe and Scandinavia; cluster B populations from E. Europe, and cluster C populations from the Russian Federation regions of Moscow and the Altai in southern Siberia.

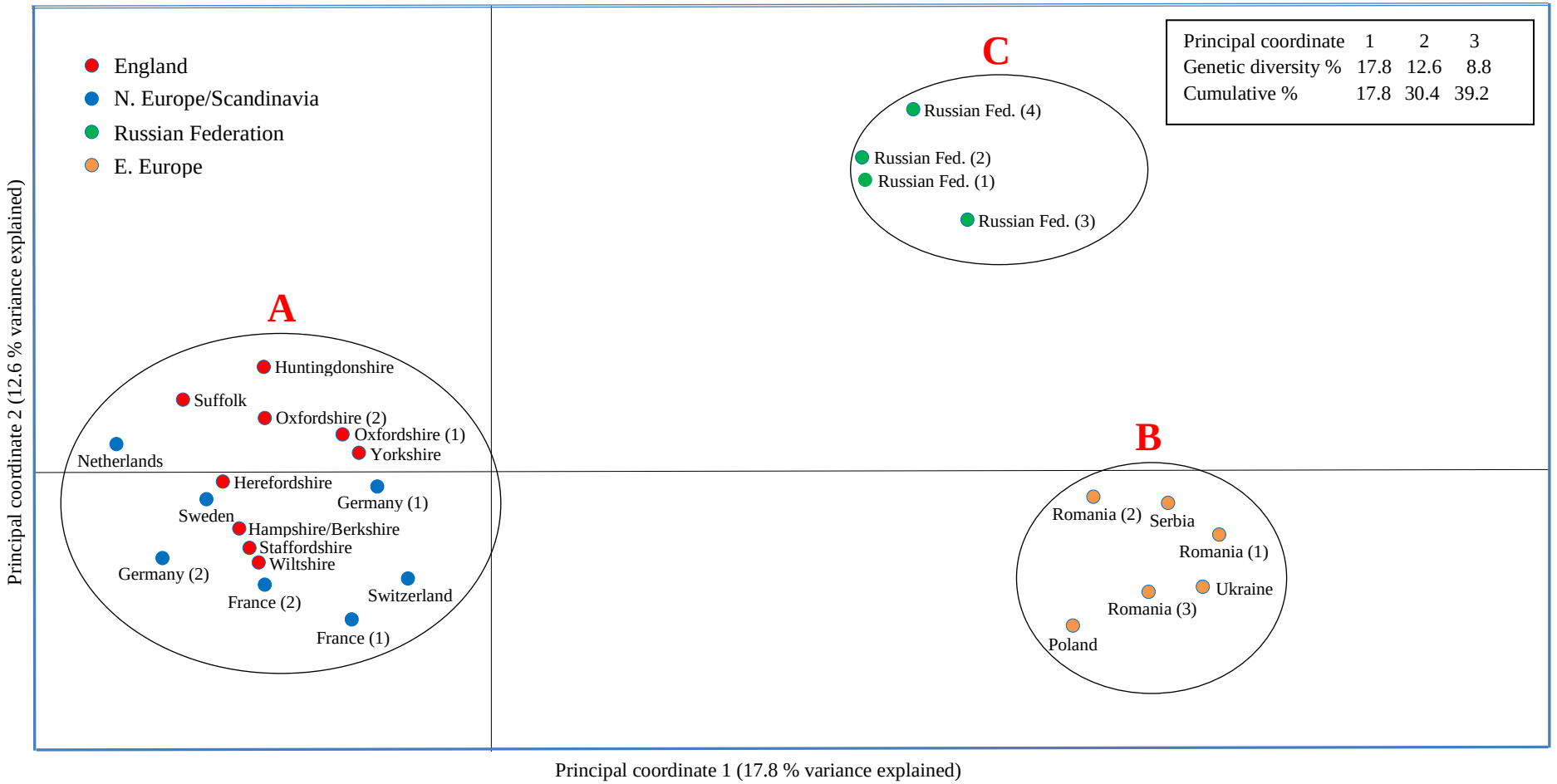

Figure 5.3 - Principal coordinates analysis scatter plot of 826 *Fritillaria meleagris* genotypes recovered from 19 loci representing 115 individuals from 26 Eurasian populations. The first eigenvector (Principal coordinate 1) explains 17.8% of variation and the second (Principal coordinate 2) explains 12.6% of variation. Population clusters are marked A - C, containing populations from England, N. Europe and Scandinavia; E. Europe and Russian Federation respectively.

5.4.10 Mantel Test

The 999 random permutations of the rows and columns of the constructed geographic-distance matrix within program GENALEX produced a frequency distribution (Figure 5.5) indicating a right-hand tail of nine *Z*-values ≥ the observed *Z* correlation coefficient value of 0.158 (Figures 5.4, 5.5). Therefore, following Sokal and Rohlf (1995), with $n_T = 9$, $N = 999$, the right-hand tail of the distribution is $\frac{(9+1)}{(999+1)}$ = *P*-value = 0.01, *i.e.* the chance of randomly obtaining a value as large as that observed is $\frac{1}{100}$, thereby revealing a strong correlation between the two matrices - shown graphically by the positive trend line slope of the relationship between pairwise genetic distance and geographic distance (Figure 5.4), and indicating evidence for genetic IBD. The probability exists that the N. Europe, E. Europe and Russian Federation populations are relics of a much larger continuous distribution (now three refugia, formed during the Last Glacial Maximum (see Figure 5.8)), which originally evolved in an intra-population stepping-stone model gene flow (population structure) pattern exhibiting IBD (Kimura and Weiss, 1964; Hedrick, 2005). However, the English and Scandinavian populations show strong affinities with N. European populations (see section 5.4.11), from where they are suspected to have been anthropogenetically introduced (see section 5.5.1), and therefore would not display the same IBD characteristics.

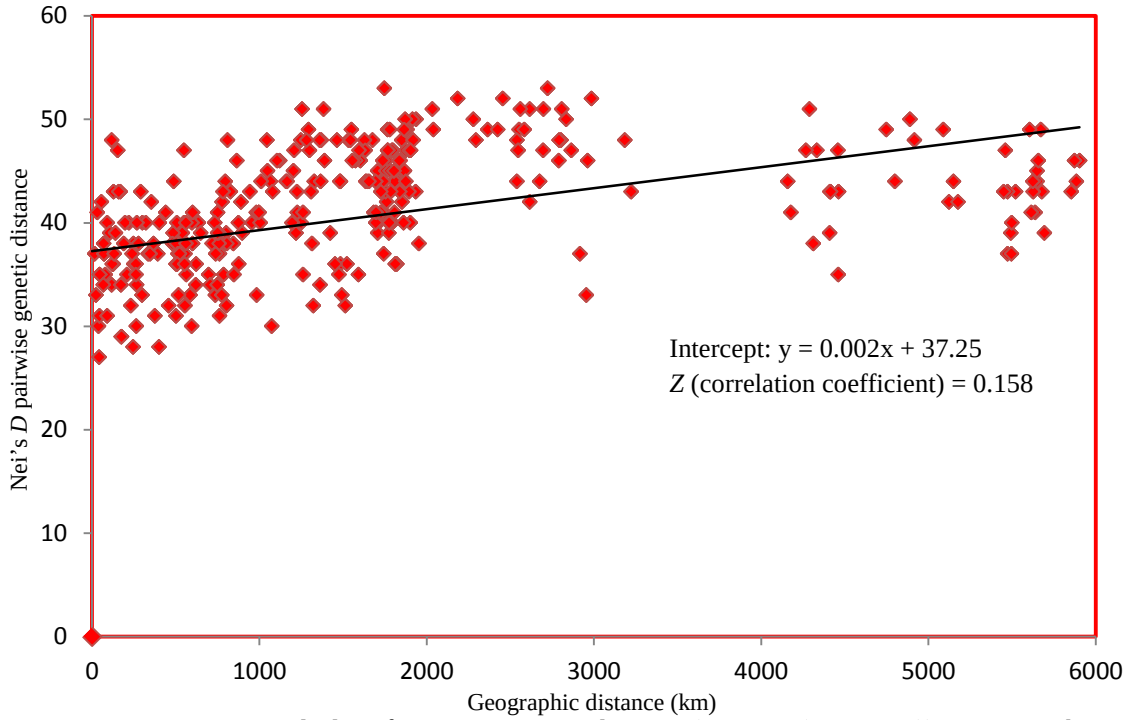

Figure 5.4 - Mantel plot of pairwise genetic distance (Nei's *D* (Nei, 1972)) vs geographic distance for 826 genotypes contained within 26 geographically separated populations; indicating trend line slope, intercept and Mantel *Z* correlation coefficient.

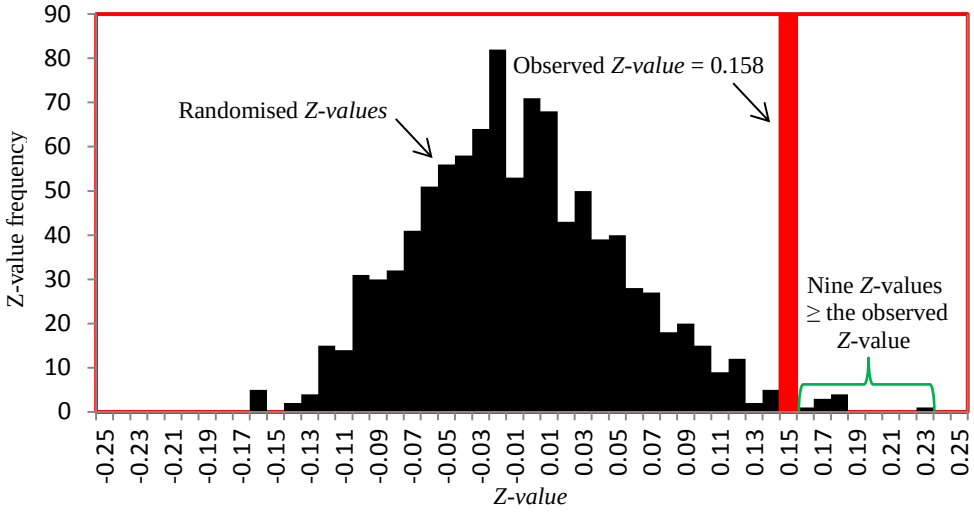

Figure 5.5 - Frequency distribution of randomised *Z-value* and observed *Z-value* correlation coefficients, to test the significance of the correlation between genetic and geographic distance for 826 genotypes contained within 26 geographically separated populations.

5.4.11 STRUCTURE analysis

For STRUCTURE to infer population and therefore cluster structure successfully, assumptions are made that loci exhibit HWE and absence of linkage disequilibrium (Pritchard *et al*., 2000). The results obtained from the analyses undertaken here and described in sections 5.3.1 and 5.3.2 indicate the project data conform to these requirements.

Given that three distinct centres of genetic diversity were recovered in PCoA analysis: N. Europe/Scandinavia, E. Europe and Russian Federation (section 5.4.9), but with only 30.4% of genetic variation accounted for by principal coordinates 1 and 2 (Figure 5.3), STRUCTURE was used to further examine and characterise genetic variation in the data, assuming the presence of both three and four population clusters. Individual analysis runs were undertaken to test a range of possible *K* values from 3 to 6 (the upper value representing the likely true number of populations (3) plus 3, as recommended by Porras-Hurtado *et al*. (2013).

From a pilot trial, a length of 30,000 for initial burn-in and 60,000 subsequent MCMC replications were found to be enough to produce approximate stationarity in computation for all models.

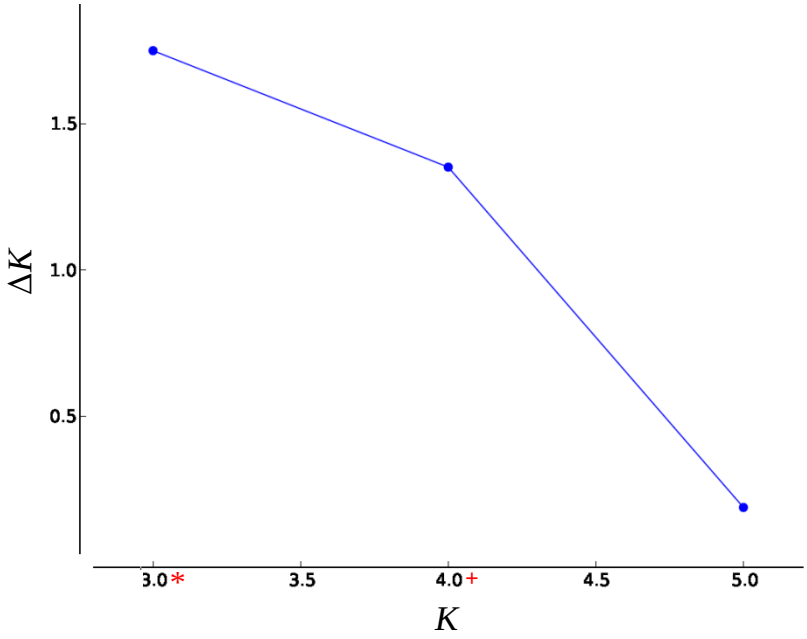

Figure 5.6 – Estimated probability of data for each value of *K*. Analysis parameters: *no admixture* model, *independent allele frequencies*, population geographic location provided - with 15 replicates for each value of *K* (2-6). The *ad hoc* statistic Δ*K* calculated by the Evanno *et al*. (2005) method. The maximum value is observed at *K* = 3*, but with evidence of a minor *K* = 4^{+} presence.

Concordant with the results from PCoA (section 5.4.9), the tested combination of *no admixture* and *independent allele frequencies* models for all values of *K* indicated a best-fit inter-population structuring for an inferred *K* = 3 (three population clusters) as the most appropriate for the presented data (Figure 5.6) - but with an inferred presence of a substructure cluster *K* = 4. The bar plot for *K* = 3 (Figure 5.7(a)) resolved mean assigned allele membership coefficient (AAMC) support for the population clusters as N. Europe/Scandinavia (0.84), E. Europe (0.88) and Russian Federation (0.97) major groups (Figure 5.3, Table 5.6). From results contained in STRUCTURE HARVESTER *population Q-matrix* output files, the correlation level of AAMC values among all 15 *FullSearch* algorithm runs for each *K* value was calculated. This provided a measure of the similarity of results between each run, producing a mean pairwise population membership similarity coefficient (*h´*) of 0.87.

Table 5.6 - Assigned allele membership coefficients (AAMC) by population, location and population cluster (K = 3).

Population	Country	AAMC	Cluster
Aubert Ings (Yorks)	England	0.95	**A**
Ducklington Mead (Oxf. 1)	England	0.99	**A**
Fox Meadow (Suffolk)	England	0.96	**A**
Lugg Meadow (Hereford)	England	0.95	**A**
Magdalen Meadow (Oxf. 2)	England	0.86	**A**
Mottey Meadows (Staffs)	England	0.86	**A**
North Meadow (Wilts)	England	0.61	**A**
Portholme Meadow (Hunts)	England	0.99	**A**
Stanford End (Hants/Berks)	England	0.99	**A**
Bouchemaine - W. Loire	France (1)	0.41	**A**
Saumur - C. Loire	France (2)	0.9	**A**
Sinn Valley - Bavaria	Germany (1)	0.6	**A**
Elbe Marshes - Schl. Holstein	Germany (2)	0.7	**A**
Hasselt - Overijssel	Netherlands	0.77	**A**
Kungsängen - Stockholm	Sweden	0.91	**A**
Les Brenets - Neuchâtel	Switzerland	0.9	**A**
	Mean	**0.84**	
Przemyśl - Subcarpathia	Poland	0.96	**B**
Poarta Salajului - Salaj	Romania (1)	0.8	**B**
Sardu - Cluj	Romania (2)	0.96	**B**
Chiesd - Salaj	Romania (3)	0.91	**B**
Obrenovac - Zabran	Serbia	0.8	**B**
Novyi Kalyniv - Lviv	Ukraine	0.82	**B**
	Mean	**0.88**	
Choya - Altai - Siberia	Russ. Fed. (1)	0.9	**C**
Cherga - Altai - Siberia	Russ. Fed. (2)	1	**C**
Kremyonki - S.W. Moscow	Russ. Fed. (3)	0.99	**C**
Pichugino - E. Moscow	Russ. Fed. (4)	1	**C**
	Mean	**0.97**	

Colour coding reflects population membership within clusters defined in Figure 5.3.

All clusters contained genetic affinities (AAMC) with the other two clusters (data not shown, but visualised in Figure 5.7(a)). The four members of the Russian Federation cluster contained the lowest AAMC from other clusters, with 0.02 from N. Europe/Scandinavia and 0.01 from E. Europe. The E. Europe cluster of six populations contained an AAMC of 0.13 from other clusters, with 0.11 from the Russian Federation and 0.2 from N. Europe. For the 16 populations within the N. Europe/Scandinavia cluster, the AAMC figure was 0.16, with 0.1 from E. Europe and 0.06 from the Russian Federation respectively. The bar plot for K = 4 (Figure 5.7(b)) indicated the presence of a minor cluster representing 0.15, 0.07 and 0.008 AAMC presence of K = 4 alleles within the N. Europe/Scandinavia, E. Europe and Russian Federation clusters respectively.

The STRUCTURE plot correlations in AAMC for the populations within the K = 3 clusters (Figure 5.7 (a), Table 5.6) infer similarity between various populations which, however, appear not so strongly supported by the PCoA results (Figure 5.3). This is likely to be due to the different approaches to analysis of alleles by GENALEX and STRUCTURE - by Euclidian genetic distance, and by frequency respectively. For N. Europe/Scandinavia (cluster **A**) there is similarity between Aubert Ings and Lugg Meadow; Ducklington Mead, Portholme Meadow and Stanford End; Fox Meadow, France (2) and Switzerland; Magdalen Meadow, Mottey Meadows and possibly Sweden; France (1), Germany (1) and North Meadow; and between Germany (2) and the Netherlands. Within E. Europe (cluster **B**), the populations of Poland and Romania (2) appear strongly correlated, as do Romania (1), Romania (3), Serbia and Ukraine. For the Russian Federation (cluster **C**), all four populations are strongly correlated, but with some evidence of a N. Europe/Scandinavia association (AAMC = 0.09) within the Siberian, Choya population.

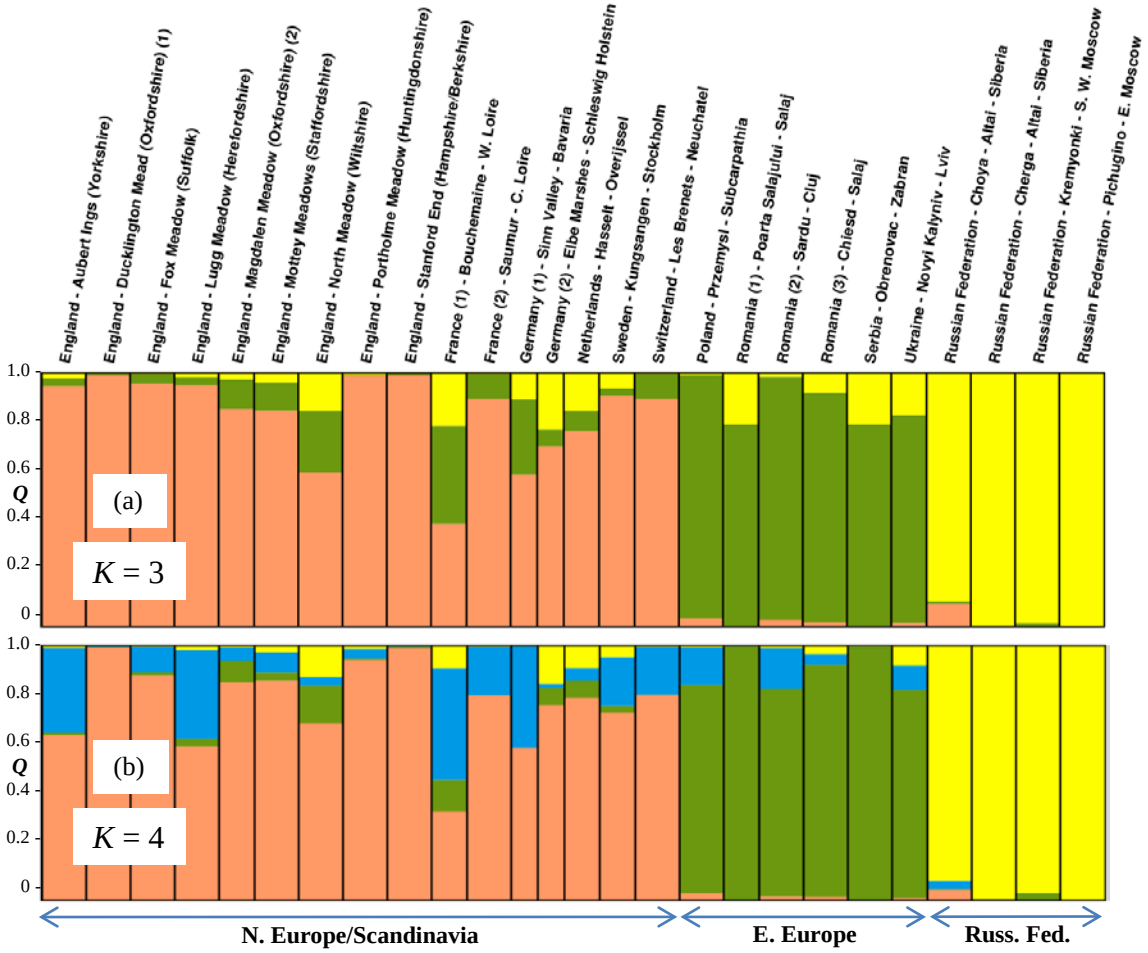

Figure 5.7 - Graphical representation of population structure for 826 genotypes recovered from 19 loci representing 115 individuals contained within 26 *F. meleagris* populations, estimated by posterior probability of membership for (a) K = 3 and (b) K = 4 clusters. Analysis parameters: *no admixture* model, *independent allele frequencies*, population geographic location provided. Each population is represented by a column stacked with a maximum of K coloured segments, each proportionally indicating the population's estimated assigned allele membership coefficients (Q) of K clusters. Vertical lines separate different populations, each of which is labelled above the figure with the sampled population country of origin and location. Major geographic associations are labelled below the figure. The bar plots shown at (a) and (b) are each based on the assigned allele membership coefficient means (AAMC) of fifteen STRUCTURE runs, which produced strongly matching population membership coefficients, having pairwise similarity (h') = 87%.

5.5 Discussion

Work to elucidate genetic structure within extant wild populations of *F. meleagris* across Eurasia has revealed a distinct geographic divide between populations at the extremes of the plant's natural range in N. Europe/Scandinavia, E. Europe and Russian Federation (Figures 5.3, 5.7). However, the results need to be interpreted with the following in mind: anomalies and errors in results can emanate from several factors, the

most significant effects coming from (1) sampling error due to inadvertently selecting unrepresentative genic specimens from the population for analysis; (2) errors in genotyping, from incorrect allele scoring during measurement of electropherogram images; (3) small sample sizes, leading to biased statistical estimates through lack of heterogeneity in alleles.

In the pair-wise correlation of population genetic variation (distance) and geographic distance (Figures 5.4, 5.5), there is substantial evidence of the effect of isolation by distance (Wright, 1943). In addition, the intra-population structures for the $K = 3$ and $K = 4$ clusters (Figure 5.7(a) and (b)) appear to show evidence of either a geographic east to west limited genetic interchange (gene flow) or the possible effects of incomplete lineage sorting of alleles. However, given the current geographic distribution of these populations (Figure 5.1, Table 5.2), it appears unlikely that there would be natural gene flow between the three clusters or between the populations within them. Thus genetic drift coupled with incomplete lineage sorting is the most likely explanation for the population structure shown within the PCoA and STRUCTURE plots. The significant among-population heterogeneity in common allele frequencies (mean of 13.8 alleles (7 - 33) (Table 5.4) for all loci, might also suggest that genetic drift has played an important part in influencing genetic variation between these populations. This explanation almost certainly also best explains the pattern shown within the $K = 4$ clusters bar plot (Figure 5.7(b)) with the additional division of AAMC population substructure recovered within STRUCTURE. It is also possible, however, that since the late 16th century human mediated transport of *F. meleagris* bulbs could have enabled gene flow among populations and across clusters (see Chapter 3). It is the assigned membership characteristics indicated among the N. Europe/Scandinavia populations and to a lesser extent the E. Europe populations (Figure 5.7(a)) that lends credence to this possibility.

With all populations in HWE, none with evidence of LD between loci, and none (putatively) displaying evidence of bottlenecks (Table 5.4), it can be assumed that these are ancient or at least very well established populations. The apparent geographic west to east cline of increasing numbers of private alleles (Table 5.5) and increase in genetic distance correlated with geographic distance (Mantel test) (Figure 5.4), both indicate isolation of populations. Slatkin (1985) argues that when gene flow between populations

is restricted to zero or close to zero through isolation, frequency of private (rare) alleles will increase in the different populations to moderate levels through genetic drift. Conversely, Slatkin (1985) suggests that when gene flow between populations is extensive, there is a much higher probability that a migrant will possess a rare allele, thereby effectively reducing the frequencies of private alleles in the emigrant populations. Wright (1943), on the other hand, also argued that random genetic drift acting within localised isolated populations can act to create increasing genetic differentiation with increasing distance. In consequence, from the results obtained (Table 5.5) there is inference that Russian Federation populations are more isolated from each other than are populations within E. Europe, which in turn are generally more isolated from each other than are populations within N. Europe/Scandinavia. The conservation implication of these distinct populations are discussed in section 5.5.3.

It is possible that leading up to the Last Glacial Maximum at about 18,000 BP (Prentice and Jolly, 2000), a Eurasian *F. meleagris* population was driven into refugia and fragmented into the Iberian, Italian and Balkan peninsulas, Greece, the Caucasus/Caspian (Hewitt, 1999), and possibly the oceanic-continental land-mass proposed by Stewart *et al.* (2010) (Figure 5.8). From here these populations may have diverged from bottlenecked populations by genetic drift, to later expand and possibly fragment again with climate warming into Western and Northern Europe from Iberia and Italy, into Eastern and Northern Europe from the Balkans and into Central Asia (modern day Russian Federation) from the Caucasus/Caspian and oceanic-continental regions. Such northerly migrations of both plants and animals have been similarly inferred previously, for example in trees, mammals, insects and fish (Hewitt, 1999; Fuentes-Utrilla *et al.*, 2014; Gouskov and Vorburger, 2016).

5.5.1 Movement of bulbs

The implications of this study might be that *F. meleagris* is viewed as a species composed of a number of large ancient populations of different historical lineages, with abundant genetic diversity. The loss of ecological connectivity between populations, which restricts or even precludes migration, can result in loss of genetic connectivity and diversity reduction and increased differentiation through genetic drift, with a concomitant susceptibility to inbreeding and therefore inbreeding depression (Lacy, 1987; Lancaster *et al.*, 2011). However, the results indicate that this might not actually

be the case, as the STRUCTURE plots for *K* = 3 and K = 4 (Figure 5.7(a) and (b)) infers a number of inter-population genetic relationships, indicating there has been gene flow between them. From the first records of *F. meleagris* in Orléans in 1567, history has shown (Chapter 3) that the plant was known to have been coveted as a very attractive addition to the gardens of the land-owning classes across Europe, and therefore new populations might well have originated from the flow of plant bulbs between countries to satisfy demand.

No verifiable records have been found of the source locations from which such early anthropogenically mediated bulb migration might have occurred, but various early records describe the presence of *F. meleagris* in gardens from the early to mid 17th century, three examples of which are of interest:

In Chapter 3, the relationship between Thomas Penny and Conrad Gessner describes the journey Penny undertook to Zurich in 1568 to stay with Gessner. The close genetic affinity of the France (2) population near Orléans and that in Switzerland inferred by the STRUCTURE plot (Figure 5.7(a)), perhaps indicates a tangible affinity between the two populations.

In England, one of the early gardens to be catalogued for plants and published as a list was that of Richard Shanne (1561-1627), who lived in Woodrowe near Methley in Yorkshire, which lists *F. meleagris* as cultivated in his garden in 1615 (Gunther, 1922). As he was known to be an eminent physician and horticulturalist, with knowledge that 'there was not allmost anie herbe growinge but he did knowe the severall names therof, and the nature and operation of the same' (Gunther, 1922), it is unlikely he would have cultivated a wild plant in his garden. It might therefore be assumed that this plant must have arrived as an 'exotic' bulb from somewhere, and from someone he trusted. This may well have been John Goodyer (1592-1664), a considerable and extremely well known scholar and botanist/horticulturalist for his time (Potter, 2006), well travelled in England and Europe and well known to Shanne (Gunther, 1922). Specimens of *F. meleagris* established in gardens such as Shanne's may well have been exchanged with other estate gardens or escaped from them, creating the now wild populations across lowland England. It might be no coincidence that many of the known ancient *F. meleagris* populations have arisen near sites of the great estates of the time (Grigson, 1955; Harvey, 1996).

In Sweden, Sernander and Sandberg (1948) records *F. meleagris* (Swedish 'Kungsängslilja') established as an ornamental in the garden of the natural scientist Olof Rudbeck in Uppsala and catalogued in his collection of cultivated plants (Rudbeck, 1658). The collection was established in 1657, four years after Rudbeck had travelled to Holland, where he collected plants and quite possibly bulbs of *F. meleagris*, as by now it was a commercially well established plant (Sernander and Sandberg, 1948). The affinity between the Swedish and Netherlands STRUCTURE plots however appears possibly to support such a proposition, and the proximity of the two populations in the PCoA plot (Figure 5.3) lends credence to the idea.

Further analysis of the STRUCTURE plot (Figure 5.7(a)) also suggests that English populations may have originated from multiple origins. The affinities between Germany (1), France (1) and North Meadow (Figure 5.7(a)) might be indicative of such multiple routes of introduction.

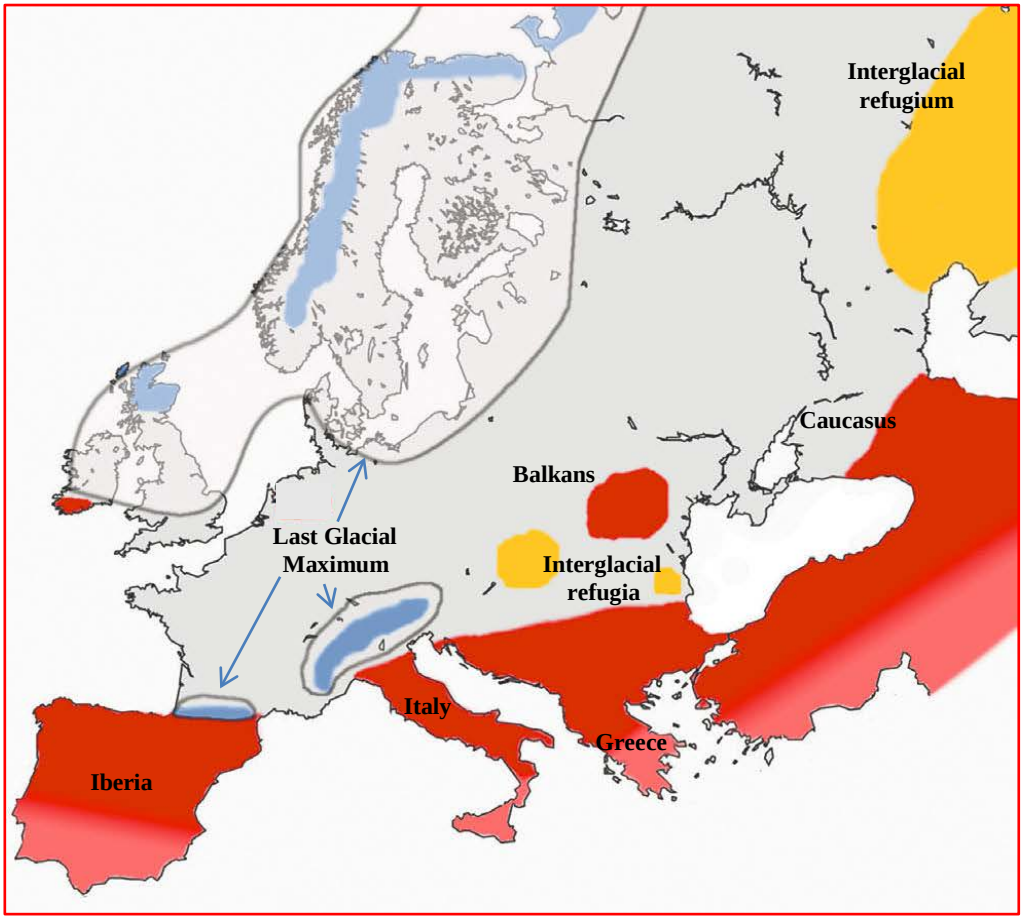

Figure 5.8 – Schematic map showing the two main categories of estimated refugia for Eurasia during the Last Glacial Maximum (LGM) (18,000 BCE). Refugia for northern temperate species are in red; interglacial refugia are in yellow; grey lines delimit ice sheets during the LGM. Adapted from Ehlers and Gibbard (2004).

5.5.2 History of *F. meleagris* in the British Isles.

The data arising from the population genetics study clearly reveal that the British representatives of *F. meleagris* form part of the broader Northern Europe population and are not distinct from them (Chapter 5). However, it has long been debated as to whether *F. meleagris* is (i) a native species of the British flora, having crossed the North-Sea via the land bridge from Northern Europe before it was closed by rising sea levels with post-Ice Age melt, (ii) an archaeophyte (naturalised in the British Isles prior to 1500, perhaps in association with agriculture), or (iii) a neophyte (introduced after 1500).

The typical habitat of *F. meleagris* of unimproved hay meadows forming part of a low-lying circum-neutral alluvial flood-plain community is defined under the National Vegetation Classification (NVC) as type MG4 (Mesotrophic grassland) (Rodwell, 1992). Robinson (2011) suggests that this habitat type is likely to have arisen in Britain with the farming practices brought by the Romans (43-410 CE), because the relevant hydrological and soil conditions for MG4, winter flooding and covering of fine alluvial sediments above the river terrace gravel, did not become widespread on the floodplain of the Uppper Thames Valley until that period. Robinson (2011) further states that it is possible there was insufficient depth of alluvial soil over much of the floodplain to support MG4 before the end of the Roman period. However, it appears that once grassland on the floodplain was managed for hay production in this early period, a distinctive hay-meadow flora seems to have developed rapidly, which potentially is when *F. meleagris* could have first been found in Britain. However, as shown in Chapter 3, no records of *F. meleagris* in published herbals or floras prior to 1574 (Dodoens (1574) could be discovered, despite exhaustive searches of the available literature. Given the charismatic nature of the plants, it is most unlikely the plant could have been overlooked by botanists or horticulturalists if they were present, and suggests that the species actually arrived in Britain for commercial or ornamental purposes through European plant collectors or sellers after *c.*1567 or 1568, when the species had been first described (Chapter 3). The plant was subsequently made popular by those managing or owning the gardens of stately homes and other large houses with land in the Thames valley.

Within Britain, the first presence of *F. meleagris* in cultivation is putatively 1597 (Griffiths, 2015) (depicted on the frontispiece of John Gerarde's (1545-1612) *Herball* (Gerarde, 1597) (Figure 5.9), but personally discovered as 1596, as recorded by Gerarde in his list of 1039 plants in his Holborn garden (Gerarde,1596). John Tradescant the Elder (1570-1638) is known to have purchased '40 Frittelarias' from Cornellis Helin in Haarlem in 1611, which he sent to London to be added to the garden of Lord Salisbury at Hatfield (Potter, 2006). The Yorkshire garden of Richard Shanne of 1615 (Gunther, 1922) is the next earliest personally discovered recorded presence of *F. meleagris* in cultivation.

With this evidence, it would seem that this species is, in all likelihood, a neophyte that has been introduced from Europe through multiple routes.

Figure 5.9 - Hand-coloured frontispiece from John Gerarde's 'Herball' (Gerarde, 1597) showing for the first time *Fritillaria meleagris* ('Fritillary' - highlighted) in cultivation (ref. Gerarde (1596)). The image purportedly depicts Gerarde with Queen Elizabeth I in Lord Burghley's garden at Theobolds Palace at Cheshunt in Hertfordshire, where Gerarde was superintendent from 1577-1598. William Shakespeare is thought to be holding the 'Fritillary' (Griffiths, 2015). NWGS (2017).

5.5.3 Conservation implication

The results from the population genetics study of *F. meleagris* (Chapter 5) have implications for the conservation of this enigmatic species, and provide a strong indication that the distinct genetic variation discovered in the three separate population clusters should be preserved. From a practical point of view this would mean that movement of bulbs between populations within individual population clusters would have a positive effect on genetic diversity, as the introduction of migrants would mitigate genetic drift. However, movement of bulbs between populations within different clusters would have the effect of diluting extant diversity between the clusters. If these three separate population clusters are indeed the product of post-glaciation migration from refugia (Chapter 5), the distinct genetic variation evident between them might infer that we are witnessing incipient speciation. More specifically, *F. meleagris* is a member of the Red Data List in Britain and classed as **A2c**[4] Vulnerable according to non-IUCN data (see footnote) (Cheffings *et al*., 2005; RBGKew, 2017a). According to IUCN Red List criteria (IUCN, 2017), the plant is also nationally assessed as **Vulnerable** in Croatia, Germany, Netherlands, Poland, Russian Federation and Ukraine; **Endangered** in Hungary and Switzerland; **Critically Endangered** in Austria, Romania, Slovakia and Slovenia and **Extinct** in Belgium and Czech Republic (Velchev, 1984: Schnittler and Günther, 1999; Nikolić and Topić, 2005; Bardunov and Novikov, 2008; Stpiczyńska *et al*., 2012; RBGKew, 2017b).

The results suggest that the extensive horticultural trade in *F. meleagris* within Europe should be restricted to trading material of the appropriate *Fritillaria* population. In brief, that means the Eastern European trade is best restricted to trading populations that are genetically similar to genotypes found in wild populations from Eastern Europe, any Russian Federation trade should be conducted with plants containing genotypes similar to Russian Federation populations, and the Northern European and Scandinavian trade to genotypes from their associated wild populations.

[4] Note: **A2c**. **(A)** Reduction in population size based on **(2)** an observed, estimated, inferred or suspected population size reduction of ≥30% over the last 10 years or three generations, whichever is the longer, where the reduction or its causes may not have ceased OR may not be understood OR may not be reversible, based on **(c)** a decline in area of occupancy, extent of occurrence and/or quality of habitat.

5.6 Data archiving

All forward and reverse primer pair sequences for 19 *F. meleagris* loci (see Table 4.1) were submitted to ProbeDB (NCBI) for archiving (https://www.ncbi.nlm.nih.gov) on 30 September 2017 (acc. nos: Pr032826426 - Pr032826444).

Chapter 6

Conclusions and future research

6.1 Thesis aims

Within this thesis the genus *Fritillaria* has been studied for its importance to genome size studies and traditional Chinese medicinal properties. In addition, the characterisation of genetic diversity of *F. meleagris* across its distribution in Eurasia has been studied because of the species conservation importance, and the findings have been compared against the earliest records of the occurrence of *Fritillaria* in Europe and the Middle East.

In Chapter 2 'Evolutionary relationships in the medicinally important genus *Fritillaria* L.', the most extensive study to date of the evolutionary relationships was undertaken using a combined plastid dataset from a broad range of species that include most (66%) of the currently recognised species diversity in the genus. The analysis also included nine other genera of the Liliaceae family as outgroups to determine their phylogenetic relationships to *Fritillaria*. The analysis is significant because of its importance for interpreting how the large genome sizes of *Fritillaria* have evolved, given that genome sizes in the genus range from 29-98 Gb/1C (Kelly *et al*., 2015) and J. Pellicer, pers.comm.). The phylogenetic tree also includes species which are commercially valuable in the Traditional Chinese Medicine (TCM) market, and which are currently under serious threat of over-harvesting. The recovered relationships infer that the most important species used in TCM are closely related to species in widespread cultivation as ornamentals. Future research using bioprospecting methods informed by the new robust phylogenetic tree presented here, are likely to contribute to helping alleviate pressure on the TCM species that are most vulnerable.

Chapter 3, entitled, 'The introduction of *Fritillaria* to Europe', investigates the history of how and when *Fritillaria* species first came to the public's attention in Europe. The opportunities for undertaking serious research afforded by the multitude of facilities available online enabled creation of a comprehensive account to be constructed. Mining of historic texts for information relating to *Fritillaria*, and more recent texts on the habitat in which *F. meleagris* is found, has provided new insights into the likely movements of *Fritillaria* in Europe by humans since the time of the Romans.

The invention in *c.*1439 of the mechanical printing press by Johannes Gutenberg provided the broader medical and culinary professions and wider public the means through which to be enlightened about not only the medicinal properties of plants, but also the joy of their visual appeal rather than just the practical.

The early to mid-16th century witnessed European expeditions to the Levant for purposes of diplomatic venture as well as commerce. The returning travellers brought with them new and exotic varieties of flora which they had discovered growing horticulturally, and which they recognised had intrinsic value commercially and as objects of desire. By this route the first species to be discovered and brought to Europe were the *Crown Imperial* (*F. imperialis*) and *Lilium persicum* (*F. persica*) in 1553 and 1576 respectively.

Internet search facilities made available for important herbaria and collections of botanical images, enabled research to be conducted in a specifically targeted and detailed fashion. This raised personal questions to the archivists responsible, as much about what material was currently available online as to what else might not be publicly enabled for either personal or online access. The images of *Hiacinto de homero* (*F. imperialis*) and *Lilium susianum* (*F. persica*) from the Biblioteca Nazionale Marciana in Venice, and both the Conrad Gessner image and Thomas Penny description of '*Narcissus variegatus*' (*F. meleagris*) from University Library Erlangen-Nürnberg, were discovered through this process. Results which would not have been possible by any other means.

Chapters 4 and 5, focus on *F. meleagris***,** with the aim of resolving the population dynamics of the species to help inform debate on its conservation status in nature. The aim of Chapter 4 **'**Microsatellite development and analysis for *F. meleagris***'** was to develop markers which could be used in such an analysis. Marker development proved

successful, and were applied in Chapter 5 ‘Genetic diversity in Eurasian populations of *F. meleagris*’ to determine the extent of genetic diversity within populations of *F. meleagris.*

The use of microsatellite loci to inform genetic diversity within populations of animals and plants has been well established by other workers. The employment of a high throughput sequencing (HTS) method to produce a library of sequences from a vouchered DNA sample, was combined with the collection of *F. meleagris* leaf material sampled from individuals within the Eurasian populations to produce the source material with which to work. Using a panel of individuals from the different populations, customised screening and genotyping methods were developed to overcome PCR limitations imposed by the large genome size of *F. meleagris* during development of microsatellite loci and recovery of genotypes. Successful identification of potentially variable loci enabled a panel of individuals to be used to find microsatellites most suitable for the detection of genetic variation between individuals.

To determine the genetic diversity of all the sampled populations of *F. meleagris* across Eurasia, a population genetics study was conducted on the remaining population individuals. This revealed a distinct population structure, divided between populations across the extremes of the species natural range, with consequences for the conservation status of this iconic and much revered species. The positive results emanating from this study show that working with plants of extreme genome size should not be perceived as a barrier to obtaining successful outcomes in population genetics analysis.

6.2 Future work

Future work to build on the findings here could include:

(1) Improved taxon sampling, including multiple individuals of each *Fritillaria* species, to generate a more comprehensive phylogenetic tree of species relationships. Phylogenetic trees that most closely resemble species trees are most valuable for resolving directions and patterns of evolutionary change of any plant trait, and of phenotypes at all levels, including in morphology, physiology, genetics and genomics. Improved phylogenies will also be important for understanding patterns of divergence across all species of *Fritillaria*, while the addition of data that will enable the dates of

nodes to be estimated will be important for reconstructing, for example, the post-glacial range changes in *Fritillaria* across its entire range.

(2) Improved marker development in population genetic analysis. During the course of the PhD studies (6 years), marker technologies have become more powerful and costs have fallen dramatically. However, to address the large genome size of *F. meleagris*, some technologies, like RAD-seq remain too expensive for population genetics studies. Nevertheless, a study starting today could well exploit DNA baiting approaches, potentially using the *c*. 350 gene baits developed in RBG, Kew's Plant and Fungal Tree of Life Project (PAFTOL; https://www.kew.org/science/who-we-are-and-what-we-do/strategic-outputs-2020/plant-and-fungal-trees-life), and then search for single nucleotide polymorphisms (SNPs) amongst the data. One additional advantage of using genes is that functional significance can be explored from alleles that associate with morphological, genetic, physiological or ecological traits. Indeed, such is the potential power of this approach, and because of the clear population structure in *F. meleagris* presented here, *F. meleagris* would be an ideal candidate for use in a pilot study seeking to test the applicability of the PAFTOL baits to population genetic studies.

(3) To extend the markers developed into other *Fritillaria* species to ascertain their population substructures. Furthermore, deeper sampling of individuals within populations of *F. meleagris*, and sampling of further natural populations of *F. meleagris* across Eurasia could provide finer resolution of population substructure within the three main clusters of populations of *F. meleagris* identified in this work. In particular, sampling more populations from Eastern Europe and the Russian Federation may prove to be informative. Use of these markers to plants in the horticultural trade could also be used to determine their genetic origin and inform conservation programmes.

Bibliography

Abdelkrim J, Robertson BC, Stanton J-AL, Gemmell NJ (2009). Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. *BioTechniques*, **46**: 185-192.

Abdulina SA (1999). *Checklist of the vascular plants of Kazakhstan*. Ministry of Science, Academy of Sciences of the Republic of Kazakhstan, Institute of Botany and Phytointroduction: Alma Ata, Kazakhstan, Russian Federation. p 115.

Almeida PRM, López-Roberts MC, Vigna BBZ, Souza AP, Góes-Neto A, van den Berg C (2013). Microsatellite markers for the endangered orchids *Cattleya labiata* Lindl. and *C. warneri* T. Moore (Orchidaceae). *Conservation Genetic Resources*, **10**: 1007/s12686-013-9909-5.

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F (2004). Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics*, **20**: 407-415.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**: 3389-3402.

Altshuler ML (2013). *PCR troubleshooting: the essential guide*. Caister Academic Press: Poole, Dorset, UK. p 53.

Álvarez I, Wendel JF (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**: 417-434.

Ambreen H, Kumar S, Variath MT, Joshi G, Bali S, Agarwal M, Kumar A, Jagannath A, Goel S (2015). Development of genomic microsatellite markers in *Carthamus tinctorius* L. (Safflower) using Next Generation Sequencing and assessment of their cross-species transferability and utility for diversity analysis. *Plos One*, **10**: e0135443.

Ambrožová K, Mandáková T, Bureš P, Neumann P, Leitch IJ, Koblížková A, Macas J, Lysak MA (2011). Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria* lilies. *Annals of Botany*, **107**: 255-268.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Resources*, **7**: 10-14.

Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993). Optimising parental selection for genetic linkage maps. *Genome*, **36**: 181-186.

Andreson R, Möls T, Remm M (2008). Predicting failure rate of PCR in large Genomes. *Nucleic Acids Research*, **36**: e66.

Andrews S (2012). FastQC. A quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (last accessed June 2015).

Anguillara ML (1561). *Semplici dell'eccellente*. Venice: Italy. p 153.

Applied Biosystems (2004). *Evaluating genetic analysis systems: microsatellite analysis.* Applied Biosystems: Foster City, California, USA. pp 1-4.

Arber A (1986). Herbals: their origin and evolution. Cambridge University Press: Cambridge, UK.

Atta-ur-Rahman AF, Choudhary MI, Gilani AH, Shaheen F, Ali RA, Noor-e-ain F, Sener B (1994). A new anticholinergic steroidal alkaloid from *Fritillaria imperialis* of Turkish origin. *Planta Medica*, **60**: 377-379.

Azmat MA, Khan IA, Cheema HMN, Rajwana IA, Khan AS, Khan AA (2012). Extraction of DNA suitable for PCR applications from mature leaves of *Mangifera indica* L. *Journal of Zhejiang University-Science B (Biomedicine & Biotechnology)*, **13**: 239-243.

Balao F, Tannhäuser M, Lorenzo MT, Hedrén M, Paun O (2016). Genetic differentiation and admixture between sibling allopolyploids in the *Dactylorhiza majalis* complex. *Heredity*, **116**: 351-361.

Balloux F, Lugon-Moulin N (2002). The estimation of population differentiation with microsatellite markers. *Molecular Ecol*ogy, **11**: 155-165.

Bardunov LV, Novikov VS (eds) (2008). *Red Book of Russian Federation (plants and fungi)*. KMK Scientific Press Ltd: Moscow, Russian Federation.

Bartolucci F, Caparelli KF, Peruzzi L (2009). A biometric study of *Fritillaria montana* Hoppe ex W.D.J. Koch s.l. (Liliaceae) shows a single polymorphic species, with no infraspecific taxa. *Plant Biosystems*, **143**: 516-527.

Bastias A, Correa F, Rojas P, Almada R, Muñoz C, Sagredo B (2016). Identification and Characterization of Microsatellite loci in Maqui (*Aristotelia chilensis* [Molina] Stunz) using Next-Generation Sequencing (NGS). *PloS One*, **11**: e0159825.

Baumgarten JCG (1816). E*numeratio Stirpium Magno Transsilvaniae Principatui*. vol. 1. Libraria Camesinae: Vienna, Austria. p 302.

Bebee TJC, Rowe G (2008). *An introduction to molecular ecology*. Oxford University Press: Oxford, UK.

Beckmann J (1823). A concise history of ancient institutions, inventions and discoveries in science and mechanic art. G & W. B. Whittaker: London. p 297.

Bell G, Lechowicz MJ, Waterway MJ (2000). Environmental heterogeneity and species diversity of forest sedges. *Journal of Ecology*, **88**: 67-87.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**: 289-300.

Bennett MD, Leitch IJ (2011). Nuclear DNA amounts in angiosperms - targets, trends and tomorrow. *Annals of Botany*, **107**: 467-590.

Bennett MD, Leitch IJ (2012). Plant DNA C-values Database (Release 6.0, Dec. 2012). Available at: http://data.kew.org/cvalues (last accessed July 2016).

Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003). Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in

Arabidopsis to be ~157 Mb and thus ~25 % larger than the *Arabidopsis* Genome Initiative estimate of ~125 Mb. *Annnals of Botany*, **91**: 547-557.

Bensky D, Clavey S, Stöger E, Gamble A (eds) (2004). *Chinese Herbal Medicine: Materia Medica*, 3rd ed. Eastland Press: Vista, California, USA.

Björklund, M (2005). A method for adjusting allele frequencies in the case of microsatellite allele drop-out. *Molecular Ecology Notes*, **5**: 676-679.

Blamey M, Fitter R, Fitter A (2003). *Wild flowers of Britain & Ireland*. A & C Black: London, UK. p 308.

Blunt W, Raphel S (1979). *The illustrated herbal*. Frances Lincoln Publishers Ltd.: London, UK.

Bonin AE, Bellemain PB, Eidesen F, Pompanon F, Brochmann C, Taberlet P (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, **13**: 3261-3273.

Boreau A (1857). *Flore du Centre de la France*. Librairie Encyclopédic de Roret: Paris.

Borges A, Rosa MS, Recchia GH, Queiroz-Silva JR, Bressan E de A, Veasey EA (2009). CTAB methods for DNA extraction of sweetpotato for microsatellite analysis. *Scientia Agricola (Piracicaba, Brazil)*, **66**: 529-534.

Botstein D, White RL, Skolnick M, Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32**: 314-331.

Bouvier L (1878). *Flore des Alpes de la Suisse et La Savoie*. Alphonse Picard: Paris, France. p 656.

BRC (2015). Biological Records Centre, Centre for Ecology and Hydrology. Wallingford, Oxfordshire, UK. Available at: http://www.brc.ac.uk/plantatlas/index.php?q=node/1321 (last accessed April 2015).

Brian A, Thomson P (2002). *The history and natural history of Lugg Meadow*. Logaston Press: Herefordshire, UK. p 31.

Britannica (2017). Printing press. Available at: https://www.britannica.com/technology/printing-press (last accessed September 2017).

Brookfield JFY (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**: 453-455.

Brownie J, Shawcross S, Theaker J, Whitcombe D, Ferrie R, Newton C, Little S (1997). The elimination of primer-dimer accumulation in PCR. *Nucleic Acids Research*, **25**: 3235-3241.

Buntaine MT, Mullen RB, Lassoie JP (2007). Human use and conservation planning in Alpine areas of northwest Yunnan, China. *Environmental Development and Sustainability*, **9**: 305-324.

Buza L, Young A, Thrall P (2000). Genetic erosion, inbreeding and reduced fitness in fragmented populations of the endangered tetraploid pea *Swainsona recta*. *Biological Conservation*, **93**: 177-186.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC, Sutherland GR (1993). Incidence and origin of 'null' alleles in the $(AC)_n$ microsatellite markers. *American Journal of Human Genetics*, **52**: 922-927.

Capperon N (1571). *Letter of 12th December 1571 from Clusius in Malines (Belgium) to Nöel Capperon in Örleans (France)*. Ctlg no. VUL 101/Capperon, N_001. Clusius Correspondence database. Available at: http://media.leidenuniv.nl/legacy/clusius_correspondence_database.pdf (last accessed September 2016).

Chakraborty R, De Andrade MN, Daiger SP, Budowle B (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics*, **56**: 45-57.

Chakraborty R, Jin L (1992). Heterozygote deficiency, population substructure and their implications in DNA fingerprinting. *Human Genetics*, **88**: 267-272.

Chapuis MP, Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**: 621-631.

Cheffings CM, Farrell L (eds) Dines TD, Jones TD, Leach RA, McKean DR, Pearman DA, Preston CD, Rumsey FJ, Taylor I (2005). *Species status 7: The vascular plant red data list for Great Britain*. Joint Nature Conservation Committee: Peterborough, UK. p 53.

Chen D-X, Li YD, Luo TS, Lin MX, Sun XY (2004). Study on biomass and net primary productivity of *Podocarpus imbricatus* plantation in Jianfengling, Hainan Island. *Forest Research*, **17**: p 598.

Chen S (2005). Sustaining herbal supplies: China. In: *Sharing Innovative Experience*, vol. 10, *Examples of the Development of Pharmaceutical Products from Medicinal Plants*. United Nations Office for South–South Cooperation: New York, USA. pp 157-163.

Chen S-C, Hsia K-C (1977). A taxonomic study of the Chinese drug bei-mu (*Fritillaria*). *Acta Phytotaxonomica Sinica*, **15**: 31-46.

Cheng WC, Fu LK (1978). *Flora Republicae Popularis Sinicae*. Science Press: Beijing, China. p 398.

Chiucchi JE, Gibbs HL (2010). Similarity of contemporary and historical gene flow among highly fragmented populations of an endangered rattlesnake. *Molecular Ecology*, **19**: 5345-5358.

Christenhusz MJM, Govaerts R, David JC, Hall T, Borland K, Roberts PS, Tuomisto A, Buerki S, Chase MW, Fay MF (2013). Tiptoe through the tulips – cultural history, molecular phylogenetics and classification of *Tulipa* (Liliaceae). *Botanical Journal of the Linnean Society*, **172**: 280-328.

Cipriani G, Testolin R (2004). Isolation and Characterization of microsatellite loci in *Fragaria*. *Molecular Ecology Notes*, **4**: 366-368.

Clapham AR, Tutin TG, Moore DM (1989). *Flora of the British Isles*. Cambridge University Press: Cambridge, UK. p 536.

Clusius C (1601). *Rariorum plantarum historia*. Plantin: Antwerp, Belgium. pp 130, 183.

Comber HF (1949). A new classification of the genus *Lilium. Lily Yearbook; Royal Horticultural Society*, **13**: 86-105.

Cornuet J-M, Luikart G (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**: 2001-2014.

Corporaal A, Horsthuis MAP, Schaminee JHJ (1993). Oecologie, verspreiding en planten sociologische postie van de kievitsbloem (*Fritillaria meleagris L.*) in Nederland en Noorwest-Europa. *Stratiotes*, **6:** 14-39.

Cumo C (ed.) (2013). Encyclopedia of plants: from Acacia to Zinnia, vol. 1. ABC-CLIO: California, USA.

Dakin EE, Avise JC (2004). Microsatellite null alleles in parentage analysis. *Heredity*, **93**: 504-509.

Day PD, Berger M, Hill L, Fay MF, Leitch AR, Leitch IJ, Kelly LJ (2014). Evolutionary relationships in the medicinally important genus *Fritillaria* L. (Liliaceae). *Molecular Phylogenetics and Evolution*, **80**: 11-19.

De Barba M, Miquel C, Lobreaux S, Quenette PY, Swenson JE, Taberlet P (2017). High-throughput microsatellite genotyping in ecology: improved accuracy, efficiency, standardization and success with low-quantity and degraded DNA. *Molecular Ecology Resources*, **17**: 492-507.

Dempster AP, Laird NM, Rubin DB (1977). Maximum Likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society B*, **39**: 1-38.

De Toni E (1910). Luigi Anguillara e Pietro Antonio Michiel. In: Pirotta R (ed.) *Annali di Botanica*. Tipografia Enrico Voghera: Rome, Italy. p 662.

Deutsch P (1988). *Ghostscript*. Artifex Software: Novato, California, USA.

De Woody J, Hipkins V (2012). *Genetic differentiation among species and populations of Fritillaria in northern California: identification of a possible hybrid swarm.* National Forest Genetics Laboratory Report for Project Number 251, Forest Service: United States Department of Agriculture: Placerville, California, USA.

De Woody J, Nason JD, Hipkins VD (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes*, **6**: 951-957.

Diniz-Filho JAF, Soares TN, Lima JS, Dobrovolski R, Landeiro, VL, Telles MP de C, Rangel TF, Bini LM (2013). Mantel test in population genetics. *Genetics and Molecular Biology*, **36**: 475-485.

Di Rienzo A, Peterson AC, Garza JC, Valdez AM, Slatkin M, Freimer NB (1994). Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA*, **91**: 3166-3170.

Dodoens R (1574). *Purgantium aliarumque eo facientium, tum et radicum convolvulorum ac deleteriarum herbarum historiae*. Plantin: Antwerp, Belgium. pp 395, 397-398.

Dodoens R (1583). *Stirpium historiae pemptades sex*. Plantin: Antwerp, Belgium. p 202.

Doležel J, Bartoš J, Vogelmayr H, Greilhuber J (2003). Nuclear DNA content and genome size of trout and human. *Cytometry Part A*, **51A**: 127-128.

Doody JP (2008). *Portholme Meadow. A celebration of Huntingdonshire's grassland.* In: Arnold HR, Dickerson BP, Drew KL, Walker PEG (eds) The Huntingdonshire Fauna and Flora Society, 60th Anniversary Report: Huntingdon, UK. p 13.

Douglas NA, Wall WA, Xiang Q-Y, Hoffmann WA, Wentworth TR, Gray JB, Hohmann MG (2011). Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses. *Molecular Ecology*, **20**: 2901-2915.

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19:** 11-15.

Drechsler A, Geller D, Freund, K, Schmeller DS, Künzel S, Rupp O, Loyau A, Denoël M, Valbuena-Ureña E, Steinfarz S (2013). What remains from a 454 run: estimation of success rates of microsatellite loci development in selected newt species (*Calotriton asper*, *Lissotriton helveticus*, and *Truturus cristatus*) and comparison with Illumina-based approaches. *Ecology and Evolution*, **3**: 3947-3957.

Druce GC (1886). *The flora of Oxfordshire*. Parker and Co.: Oxford, UK. p 305.

Druce GC (1897). *The flora of Berkshire*. Clarendon Press: Oxford, UK. pp 493-494.

Earl DA, VonHoldt BM (2012). Structure Harvester: a website and program for visualising STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**: 359-361.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics*, **49**: 746-756.

Egmond F (2008). Apothecaries as experts and brokers in the sixteenth-century network of the naturalist Carolus Clusius. In: Dupré S, Kusukawa S (eds) *History of Universities* vol. XXIII/2. Oxford University Press: Oxford, UK. p 78.

Egmond F (2010). *The world of Carolus Clusius: natural history in the making, 1550-1610*. Pickering & Chatto: London, UK. p 118.

Egmond F (2013). Observing nature. The correspondence network of Carolus Clusius (1526-1609). In: van Miert D (ed.) *Communicating observations in early modern letters (1500-1675): epistolography and epistemology in the age of the scientific revolution*. Warburg Institute School of Advanced Study, University of London: UK. pp 43-72.

Egmond F (2017). *Eye for detail*: images of plants and animals in art and science, 1500 - 1630. Reaktion Books: London, UK. pp 90, 220.

Ehlers J, Gibbard P (2004). *Quaternary glaciations – extent and chronology*. Elsevier: London, UK.

Engels WR (2009). Exact tests for Hardy-Weinberg proportions. *Genetics*, **183**: 1431-1441.

Estoup A, Cornuet J-M (1999). Microsatellite evolution: inferences from population data. In: Goldstein DB, Schlötterer C (eds) *Microsatellites: evolution and applications*. Oxford University Press: Oxford, UK. pp 49-64.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**: 2611-2620.

Ewing B, Green P (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Resources*, **8**: 186-194.

Ewing B, Hillier L, Wendl MC, Green P (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Resources*, **8**: 175-185.

Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**: 564-567.

Falush D, Stephens M, Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**: 1567-1587.

Farrelly V, Rainey FA, Stackebrandt E (1995). Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Applied and Environmental Microbiology*, **61**: 2798-2801.

Fay MF, Chase MW (2000). Modern concepts of Liliaceae, with a focus on the relationships of *Fritillaria*. *Curtis's Botanical Magazine*, **17**: 146-149.

Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **63**: 783-791.

Fernandez-Silva I, Toonen RJ (2013). Optimizing selection of microsatellite loci from 454 pyrosequencing via post-sequencing bioinformatic analyses. *Methods in Molecular and Cellular Biology*, **1006**: 101-120.

Floron (2016). Verspreidingsatlas vaatplanten. Available at: https://www.verspreidingsatlas.nl/0532# (last accessed June 2016).

FMP (2016). *Snakeshead fritillary* (*Fritillaria meleagris*). Available at: https://www.floodplainmeadows.org.uk (last accessed October 2016).

Fontúrbel FE, Murúa MM, Vega-Retter C (2016). Development of ten microsatellite markers from the keystone mistletoe *Tristerix corymbosus* (Loranthaceae) using 454 next generation sequencing and their applicability to population genetic structure studies. *Molecular Biology Reports*, **43**: 339-343.

Forster CT, Daniell FHB (1884). *The life and letters of Ogier Ghiselin de Busbecq. Seigneur of Bousbecqe, Knight, Imperial Ambassador*. C. Kegan Paul & Co: London. pp 62, 107-108.

Forsyth AA (1954). Liliaceae. In: *British poisonous plants*, Bulletin no.161. HMSO: London. p 86.

Fowler J, Cohen L, Jarvis P (1998). *Practical statistics for field biology*. John Wiley & Sons: Chichester, West Sussex, UK. pp 211-220.

Frankham R, Ballou JD, Briscoe DA (2010). *Introduction to conservation genetics*. Cambridge University Press: Cambridge, UK. p 83.

Fratti C, Segarizzi A (eds) (1909). *Catologo dei Codici Marciani Italiani a cura della direzzione della R. Bibliotecha Nazionale di S. Marco in Venezia*. vol. 1. G.Ferraguti et C.: Modena, Italy. pp 212-214.

Fu LG, Jin JM (1992). *Red List of endangered plants in China*. Science Press: Beijing, China.

Fuentes-Utrilla P, Venturas M, Hollingsworth PM, Squirrell J, Collada C, Stone GN, Gil L (2014). Extending glacial refugia for a European tree: genetic markers show that Iberian populations of white elm are native relicts and not introductions. *Heredity*, **112**: 105-113.

Funk DJ, Omland KE (2003). Species-level paraphyly and polyphyly: frequency, causes and consequences with insights from animal mitochondrial DNA. *Annual Review of Ecology Evolution and Systematics*, **34**: 397-423.

Gao Y-D, Harris AJ, Zhou S-D, He X-J (2013). Evolutionary events in *Lilium* (including *Nomocharis*, Liliaceae) are temporally correlated with orogenies of the Q–T Plateau and the Hengduan Mountains. *Molecular Phylogenetics and Evolution*, **68**: 443-460.

Gao Y-D, Hohenegger M, Harris AJ, Zhou S-D, He X-J, Wan J (2012a). A new species in the genus *Nomocharis Franchet* (Liliaceae): evidence that brings the genus *Nomocharis* into Lilium. *Plant Systematics and Evolution*, **298**: 69-85.

Gao Y-D, Zhou S-D, Wan J (2012b). Chromosome diversity and evolution in tribe Lilieae (Liliaceae) with emphasis on Chinese species. *Journal of Plant Research,* **125:** 55-69.

Gardner MG, Fitch AJ, Bertozzi T, Lowe AJ (2011). Rise of the machines - recommendations for ecologists when using next generation sequencing for microsatellite development. *Molecular Ecology Resources*, **11**: 1093-1101.

Garner TWJ (2002). Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome*, **45**: 212-215.

Gerarde J (1596). *Catalogus arborum, fruticum, ac plantarum tam indigenarum, quam exoticarum, in horto Johannis Gerardi civis et chirurgi Londinensis nascentium.* Johannis Norton: London. p 8.

Gerarde J (1597). *The herball or general historie of plantes*. John Norton: London, UK. Frontispiece.

Gessner C (1561). De hortis Germaniae liber recens. In: *Valerius C. Annotationes in Pedacii Dioscoridis Anazarbel de medica materia libros V.* Josias Rihel: Strasbourg, France. p 212.

Gessner C (1751). *Opera botanica* v. 1. Casimir Christoph Schmeidel: Nuremberg, Germany.

Gessner C (1771). *Opera botanica* v. 2. Casimir Christoph Schmeidel: Nuremberg, Germany.

Gibbons B (1990). North Meadow, Cricklade, Wiltshire. *British Wildlife*, **1**: 216-218.

Gilbert KJ (2016). Identifying the number of population clusters with STRUCTURE: problems and solutions. *Molecular Ecology Resources*, **16**: 601-603.

Gillam B (ed.) (1993). Liliaceae. In: *The Wiltshire flora*, Pisces Publications: Newbury, Berkshire, UK. pp 345-346.

Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F (2011). Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*, **12**: 245-255.

Ginot F, Bordelais I, Nguyen S, Gyapay G (1996). Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research*, **24**: 540-541.

Goldstein DB, Schlötterer C (1999). *Microsatellites: evolution and applications*. Oxford University Press: Oxford, UK. p 278.

Goudet J (2002). Fstat, a program to estimate and test gene diversities and fixation indices (Version 2.9.3.2). Available at: http://www2.unil.ch/popgen/softwares/fstat.htm (last accessed February 2017).

Gouskov A, Vorburger C (2016). Postglacial recolonisations, watershed crossings and human translocations shape the distribution of chub lineages around the Swiss Alps. *BMC Evolutionary Biology*, **16**: 185-197.

Gowing DJG, Tallowin JRB, Dise NB, Goodyear J, Dodd ME, Lodge RJ (2002). A review of the ecology, hydrology and nutrient dynamics of floodplain meadows in England. *English Nature Research Reports No. 446*. English Nature: Peterborough, UK.

Greene EL (1983). *Landmarks of botanical history* v. 2. Stanford University Press: Stanford, USA. p 851.

Greilhuber J, Doležel J, Lysak MA, Bennett MD (2005). The origin, evolution and proposed stabilization of the terms ‘Genome size’ and ‘C-value’ to describe nuclear DNA contents. *Annals of Botany*, **95**: 255-260.

Greilhuber J, Leitch IJ (2013). Genome size and the phenotype. In: Leitch IJ, Greilhuber J, Doležel J, Wendel JF (eds) *Plant genome diversity,* vol. 2, *Physical structure, behaviour and evolution of plant genomes.* Springer-Verlag: Vienna, Austria. pp 323-344.

Grieve M (1984). *A modern herbal*. Savvas Publishing: Adelaide, South Australia. p 328.

Griffiths M (2015). The true face of Shakespeare: Dioscorides and the fourth man. *Country Life*, 20 May: 103-138.

Grigson G (1955). Snake's Head, Fritillary. In: *The Englishman's flora*. Phoenix House: London. pp 402-404.

Groom Q, Walker K, McIntosh J (2011). Annex 1: Guidance on sampling approaches. In: *BSBI recording the British and Irish flora 2010-2020.* Botanical Society of the British Isles: Bristol, UK. pp 1-9.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevan C, Malausa T, Revardel E, Salin F, Petit RJ (2011). Current trends in microsatellite genotyping. *Molecular Ecology Res*ources, **11**: 591-611.

Guillot G and Rousset F (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4**: 336-344.

Gunther RT (1922). *Early British botanists*. Oxford University Press: Oxford, England, UK. pp 264, 311.

Guo SW, Thompson EA (1992). Performing the Exact Test of Hardy-Weinberg Proportion for multiple alleles. *Biometrics*, **48**: 361-372.

Gusmão L, Sánchez-Diz P, Calafell F, Martin P, Alonso CA, Álvarez-Fernández F, Alves C, Borjas-Fajado L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Lonardo AM, Espinheira R, Fagundes de Carvalho E, Farfán MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza Góes AC, Santapa OA, Sumita DR, Vallejo G, Vidal Rioja L, Vide MC, Vieira da Silva CI, Whittle MR, Zabala W, Zarrabeitia MT, Alonso A, Carracedo A, and Amorim A (2005). Mutation rates at Y chromosome specific microsatellites. *Human Mutation*, **26**: 520-528.

Haeupler H, Montag A, Wöldecke K (1976). Verschollene und gefährdete Gefässpflanzen in Niedersachsen. In: *30 jahre Naturschutz und Landschaftspflege in Niedersachsen.* Niedersächsischer Minister für Ernährung, Landwirtschaft und Forsten: Hannover, Federal Republic of Germany. pp 1-24.

Hale ML, Burg TM, Steeves TE (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. *PloS ONE*, **7:** e45170.

Hao D-D, Gu X-J, Xiao P-G, Peng Y (2013). Phytochemical and biological research of *Fritillaria* medicinal resources. *Chinese Journal of Natural Medicines*, **11**: 330-344.

Harmon LJ, Glor RE (2010). Poor statistical performance of the Mantel test in phylogenetic comparative analyses. *Evolution*, **64**: 2173-2178.

Hartl DL, Clark AG (2007). *Principles of population genetics*. Sinauer Associates Inc.: Sunderland, Massachusetts, USA. p 68.

Harvey JH (1996). Fritillary and Martagon - wild or garden? *Garden History*, **24**: 30-38.

Hastings WK (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**: 97-109.

Hayek A (1933). *Prodromus Florae Peninsulae Balcanicae*. Verlag Des Repertoriums: Dahlem, Germany. p 63.

Hedrick PW (2005). Genetics of populations. Jones and Bartlett Publishers: Sudbury, Massachusetts, USA. pp 504-505.

Henrey B (1975). *British botanical and horticultural literature before 1800*. Oxford University Press: Oxford, UK.

Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**: 87-112.

Hill CR, Butler JM, Vallone PM (2009). A 26plex autosomal STR assay to aid human identity testing. *Journal of Forensic Sciences*, **54**: 1008-1015.

Hill L (2011). A taxonomic history of Japanese endemic *Fritillaria* (Liliaceae). *Kew Bulletin*, **66**: 227-240.

Hill L (2013). *Fritillaria*: A list of published names. version 5.7. Available at: http://www.fritillariaicones.com/info/names/frit.names.pdf (last accessed November 2013).

Hoffman JI, Amos W (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**: 599-612.

Horton PJ, Jefferson EG (2006). The Wiltshire Fritillary meadows: a case study in habitat degradation. *British Wildlife*, **17**: 176-184.

Huang K, Ritland K, Dunn DW, Qi X, Guo S, Li B (2016). Estimating relatedness in the presence of null alleles. *Genetics*, **202**: 247-260.

Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**: 754-755.

Hultén E, Fries M (1985). *Atlas of Northern European vascular plants north of the Tropic of Cancer*. Koeltz Scientific Books: Königstein, Federal Republic of Germany. p 70.

Hutchinson DW, Templeton AR (1999). Correlation of pairwise genetic and geographic distances measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution*, **53**: 1898-1914.

İkinci N (2011). Molecular phylogeny and divergence times estimates of *Lilium* section Liriotypus (Liliaceae) based on plastid and nuclear ribosomal ITS sequence data. *Turkish Journal of Botany*, **35**: 319-330.

IUCN (2017). *The IUCN Red List of Threatened Species.* Version 2016-3. Available at: http://www.iucnredlist.org/static/categories_criteria_3_1 (last accessed April 2017).

Jaffe´ R, Castilla A, Pope N, Imperatriz-Fonseca VL, Metzger JP, Arias MC, Jha S (2016). Landscape genetics of a tropical rescue pollinator. *Conservation Genetics*, **17**: 267-278.

Jakobsson M, Rosenberg NA (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**: 1801-1806.

Jasra A, Holmes CC, Stephens DA (2005). Markov Chain Monte Carlo methods and the label switching problem in Bayesian mixture modelling. *Statistical Science*, **20**: 50-67.

Jefferson RG, Pinches CE (2011). The conservation of floodplain meadows in Great Britain: an overview. *Fritillary*, **5**: 11-24.

JNCC (2017). *National Vegetation Classification*. Available at: http://jncc.defra.gov.uk/page-4259 (last accessed April 2017).

Jones AG, Stockwell CA, Walker D, Avise JC (1998). The molecular basis of a microsatellite null allele from the White Sands Pupfish. *Journal of Heredity*, **89**: 339-342.

Jones N, Ougham H, Thomas H, Pašakinskiené I (2009). Markers and mapping revisited: finding your gene. *New Phytologist*, **183**: 935-966.

Kalinowski ST (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity*, **106**: 625-632.

Kalinowski ST, Taper ML (2006). Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics*, **7**: 991-995.

Keller K, Chintauan-Marquier IC, Veltsos P, Nichols RA (2006). Ribosomal DNA in the Grasshopper *Podisma pedestris*: escape from concerted evolution. *Genetics*, **174**: 863-874.

Kelly LJ, Leitch AR, Clarkson JJ, Knapp S, Chase MW (2013). Reconstructing the complex evolutionary origin of wild allopolyploid tobaccos (*Nicotiana* section Suaveolentes). *Evolution*, **67**: 80-94.

Kelly LJ, Leitch AR, Fay MF, Renny-Byfield S, Pellicer J, Macas J, Leitch IJ (2012). Why size really matters when sequencing plant genomes. *Plant Ecology and Diversity*, **5**: 415-425.

Kelly LJ, Leitch IJ (2011). Exploring giant plant genomes with next-generation sequencing technology. *Chromosome Research*, **19**: 939-953.

Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novák P, Neumann P, Lysak MA, Day PD, Berger M, Fay MF, Nichols RA, Leitch IJ (2015). Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. *New Phytologist*, **208**: 596-607.

Kent DH (2001). The history of *Fritillaria meleagris* in Britain to 1900. *The London Naturalist*, **80**: 29-41.

Kimura M, Crow JF (1964). The number of alleles that can be maintained in a finite population. *Genetics,* **49:** 725-738.

Kimura M, Weiss GH (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, **49**: 561-576.

King N, Wells DA (1993). *Fritillaria meleagris* L. Fritillary. In: Gillam B (ed.) *The Wiltshire Flora*. Pisces Publications: Newbury, UK. pp 95-97.

Knight CA, Ackerley DD (2002). Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecology Letters*, **5**: 66-76.

Knight CA, Molinari NA, Petrov DA (2005). The large genome hypothesis: evolution, ecology and phenotype. *Annals of Botany*, **95**: 177-190.

Konchar K, Li X-L, Yang Y-P, Emshwiller E (2011). Phytochemical variation in *Fritillaria cirrhosa* D. Don (Chuan Bei Mu) in relation to plant reproductive stage and timing of harvest. *Economic Botany*, **65**: 283-294.

Kwok S, Kellog DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990). Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus 1 model studies. *Nucleic Acids Research*, **18**: 999-1005.

Lacy RC (1987). Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection and population subdivision. *Conservation Biology*, **1**: 143-158.

Lancaster M, Taylor AC, Cooper SJB, Carthew SM (2011). Limited ecological connectivity of an arborial marsupial across a forest plantation landscape despite apparent resiliance to fragmentation. *Molecular Ecology*, **20**: 2258-2271.

Lang R (2009). *A Ghostscript graphical interface*. Ghostgum Software Pty Ltd: Melbourne, Australia.

Langella O (2012). Populations 1.2.32. Available at: https://www.bioinformatics.org/~tryphon/populations (last accessed May 2016).

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**: 2947-2948.

Lavor P, van den Berg C, Versieux LM (2013). Transferability of 10 nuclear microsatellite primers to *Vriesea minarum* (Bromeliaceae), a narrowly endemic and threatened species from Brazil. *Brazilian Journal of Botany*, **36**: 165-168.

Legendre P, Fortin MJ (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**: 831-844

Lehmann T, Hawley WA, Collins FH (1996). An evaluation of evolutionary constraints on microsatellite loci using null alleles. *Genetics*, **144**: 1155-1163.

Leitch IJ, Beaulieu JM, Cheung K, Hanson L, Lysak MA, Fay MF (2007). Punctuated genome size evolution in Liliaceae. *Journal of Evolutionary Biology*, **20**: 2296-2308.

Leon C, Lin Y-L (2017). *Chinese medicinal plants, herbal drugs and substitutes*. Kew publishing: Royal Botanic Gardens, Kew, UK. pp 78, 332.

Lewontin RC, Kojima K (1960). The evolutionary dynamics of complex polymorphisms. *Evolution*, **14**: 450-472.

Li H-J, Jiang Y, Li P, Ye WC (2006). Puqienine F, a novel veratramine alkaloid from the bulbs of *Fritillaria puqiensis*. *Chemical and Pharmaceutical Bulletin*, **54**: 722-724.

Li N, Deng Q, Huang L, Su Y, Wang T (2014). Isolation and characterization of ten polymorphic microsatellite loci for a vulnerable species *Dacrycarpus imbricatus* (Podocarpaceae) in China. *Biochemical Systematics and Ecology*, **54**: 83-87.

Li S, Chou HH (2004). Lucy2: an interactive DNA sequence quality trimming and vector removal tool. *Bioinformatics*, **20**: 2865-2866.

Lin G, Li P, Li S-L, Chan S-W (2001). Chromatographic analysis of *Fritillaria* isosteroidal alkaloids, the active ingredients of beimu, the antitussive traditional Chinese medicinal herb. *Journal of Chromatography* A, **935**: 321-338.

Linnaeus C (1753). *Fritillaria*. In: *Species Plantarum* vol. 2. L. Salvius: Stockholm, Sweden. pp 303-304.

Lloyd BD (2003). The demographic history of the New Zealand short-tailed bat *Mystacina tuberculata* inferred from modified control region sequences. *Molecular Ecology*, **12**: 1895-1911.

Löytynoja A, Goldman N (2010). WebPrank: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*, **11**: 579-585.

Lozina-Lozinskaya AS (1935). Genus 271. *Fritillaria* L. In: Komarov VL (ed.) *Flora of the USSR*, vol 4. Leningrad: Izd-vo Akademii Nauk SSSR. p 236.

Luikart G, Cornuet JM (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology*, **12**: 228-237.

Mabey R (1996). *Flora Britannica*. Chatto & Windus: London, UK. pp 403-409.

MacDougall-Shackleton EA, MacDougall-Shackleton SA (2001). Cultural and genetic evolution in mountain white-crowned sparrows: song dialects are associated with population structure. *Evolution*, **55**: 2568-2575.

Macfarlane RM (1978). On the taxonomic status of *Fritillaria phaeanthera* Eastw. (Liliaceae). *Madroño*, **25**: 93-100.

Maddison DR, Maddison WP (2005). *MacClade 4: Analysis of Phylogeny and Character Evolution*, Version 4.08 OS X. Sinauer Associates: Sunderland, Massachusetts, USA.

Malyschev LI, Peschkova GA (eds) (2001). *Flora of Siberia* vol. 4. Science Publishers, Inc.: Enfield, New Hampshire, USA. pp 95, 195.

Mantel N (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**: 209-220.

Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, Kelly JK (2009). Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity*, **103**: 310-317.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**: 639-655.

Maruyama T, Fuerst PA (1985). Population bottlenecks and non-equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics*, **111**: 675-689.

Mason P (2007). Americana in the Exoticorum libri decem of Charles de L' Écluse. In: Egmond F, Hoftijzer P, Visser R (eds) *Carolus Clusius: Towards a cultural history of a Renaissance naturalist*. Royal Netherlands Academy of Arts and Sciences: Amsterdam, Netherlands. p 199.

Matschiner M, Salzburger W (2009). TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics*, **25**: 1982-1983.

McPherson M, Møller S (2006). *PCR*. Taylor & Francis Group: Abingdon, Oxfordshire, UK. pp 67-68.

Meglécz E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, Martin JF (2010). QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*, **26**: 403-404.

Meglécz E, Martin J-F (2011). *QDD2 for Windows and Linux version 2 (28 June 2011).* Aix-Marseille Université: Marseille, France. pp 1-19.

Metzker ML (2010). Sequencing technologies - the next generation. *Nature Reviews: Genetics*, **11**: 31-46.

Meyers SC, Liston A, Meinke RJ (2006). *The hybrid origin of Fritillaria gentneri* Gilkey (Liliaceae). Botany 2006 Abstracts. California State University, Chico, California, USA. Available at: http://2006.botanyconference.org/engine/search/index.php?func=detail&aid=629 (last accessed March 2016).

Morey M, Fernández-Marmiesse A, Castiñeiras D, Fraga JM, Couce ML, Cocho JA (2013). A glimpse into past, present, and future DNA sequencing. *Molecular Genetics and Metabolism*, **110**: 3-24.

Muratović E, Hidalgo O, Garnatje T, Siljak-Yakovlev S (2010a). Molecular phylogeny and genome size in European Lilies (Genus *Lilium*, Liliaceae). *Advanced Science Letters*, **3**: 180-189.

Muratović E, Odil R, Bogunić F, Šoljan D, Siljak-Yakovlev S (2010b). Karyotype evolution and speciation of European lilies from *Lilium* sect. Liriotypus. *Taxon*, **59**: 165-175.

Nei M (1972). Genetic distance between populations. *American Naturalist*, **106**: 283-392.

Nei M, Maruyama T, Chakraborty R (1975). The bottleneck effect and genetic variability in populations. *Evolution*, **29**: 1-10.

NHGRI (2017). National Human Genome Research Institute. Available at: www.genome.gov/10001691/nhgri-genome-sequencing-program-gsp (last accessed April 2015.

Nikolić T, Topić J (eds) (2005). *Red Book of vascular flora of Croatia*. Ministry of Culture: Zagreb, Croatia. p 440.

Nissen C (1958). *Herbals of five centuries*. Art Ancien: Zurich, Switzerland.

Nixon KC (1999). The parsimony ratchet, a new method for rapid parsimony analysis. *Cladistics*, **15**: 407-414.

NWGS (2017). *Coloured frontispiece of John Gerarde's herbal of 1597*. Available at: http://nantwichwalledgarden.org.uk/wp-content/uploads/2015/11/Theobalds-image.jpg (last accessed October 2017).

Nylander JAA (2004). *MrModeltest* version 2. Program Distributed by the Author. Evolutionary Biology Centre: Uppsala University, Sweden.

Nylander JAA, Wilgenbusch JC, Warren DL, Swofford DL (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, **24**: 581-583.

Ohta T, Kimura M (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetics Research*, **22**: 201-204.

Orlóci L (1978). *Multivariate analysis in vegetation research*. Junk B. V.: The Hague, Netherlands.

Oswald P (1992). The Fritillary in Britain - a historical perspective. *British Wildlife*, **3**: 200-210.

Paetku D and Strobeck C (1995). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology*, **4**: 519-520.

Palmer MW (2008). Ordination methods – an overview. Available at: http://www.okstate.edu/artsci/botany/ordinate/overview (last accessed March 2015).

Palm-Silva C, Lexer C, Paggi GM, Barbara T, Bered F, Bodanese-Zanettini MH (2009). Range-wide patterns of nuclear and chloroplast DNA diversity in *Vriesea gigantea* (Bromeliaceae), a neotropical forest species. *Heredity*, **103**: 503-512.

Pante E, Abdelkrim J, Viricel A, Gey D, France SC, Boisselier MC, Samadi S (2015). Use of RAD sequencing for delimiting species. *Heredity*, **114**: 450-459.

Park K, Daston L (eds) (2003). *The Cambridge History of Science* vol. 3. *Early modern science*. Cambridge University Press: Cambridge, UK. p 256.

Parkin EJ, Butlin RK (2004). Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Molecular Biology and Evolution*, **21**: 1595-1601.

Parkinson J (1629). *Paradisi in sole paradisus terrestris*. London. pp 28, 30, 45, 420.

Patwardhan A, Ray S, Roy A (2014). Molecular markers in phylogenetic studies – a review. *Phylogenetics & Evolutionary Biology*, **2**: 1-9.

Pavord A (1999). *The tulip*. Bloomsbury Publishing: London, UK. pp 58-59.

Pavord A (2005). *The naming of names: the search for order in the world of plants.* Bloomsbury Publishing: London, UK. pp 17, 212.

Peakall R, Smouse PE (2012). GenAlex 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics*, **28**: 2537-2539.

Pearman DA (2007). 'Far from any house - assessing the status of doubtfully native species in the flora of the British Isles. *Watsonia*, **26**: 271-290.

Pellicer J, Kelly LJ, Leitch IJ, Zomlefer WB, Fay MF (2014). A universe of dwarfs and giants: genome size and chromosome evolution in the monocot family Melanthiaceae. *New Phytologist*, **201**: 1484-1497.

Pharmacopoeia of the People's Republic of China (2010). vol. 1: Chinese Pharmacopoeia Commission. China Science Medical Press: Beijing, China.

Piry S, Luikart G, Cornuet JM, (1999). Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity*, **90**: 502-503.

Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo A, Lareu MV (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*, **4**: 1-13.

Potter J (2006). *Strange Blooms: the curious lives and adventures of the John Tradescants.* Atlantic Books: London, UK. p 134.

Potts WTW, Fear L (2000). Thomas Penny the first English Entomologist In: *Contrebis* vol XXV (2000). Lancaster Archaeological and Historical Society: Lancaster, UK. pp 21-30.

Prentice C, Jolly D (2000). Mid-Holocene and glacial-maximum vegetation geography of the northern continents and Africa. *Journal of Biogeography*, **27**: 507-519.

Price JM (1983). *Fritillaria meleagris* – aspects of conservation. *Proceedings of the Birmingham Natural History Society*, **25:** 36-40.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics*, **155**: 945-959.

Pritchard JK, Wen X, Falush D (2010). *Documentation for STRUCTURE software: Version 2.3.* Available at: http://pritch.bsd.uchicago.edu/structure (last accessed April 2016).

Puskás LG, Bottka S (1995). Reduction of mispriming in amplification reactions with restricted PCR. *Genome Research*, **5**: 309-311.

Puttick MN, Clark J, Donoghue PCJ (2015). Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proceedings of the Royal Society of London B: Biological Sciences*, **282**: 20152289.

Rambaut A, Drummond AJ (2009). Tracer: version 1.5. Available at: http://tree.bio.ed.ac.uk/software/tracer (last accessed March 2013).

Raven CE (1947). *English Naturalists from Neckam to Ray*. Cambridge University Press: Cambridge, UK. pp 162-163.

RBGKew (2017a). *Fritillaria meleagris* (*Snake's-head fritillary*). Available at: http://www.kew.org (last accessed November 2015).

RBGKew (2017b). State of the world's plants. Available at: https://www.kew.org/science/data-and-resources/plants (last accessed June 2017).

Reusch TBH, Stam TW, Olsen JL (2000). A microsatellite-based estimation of clonal diversity and population subdivision in *Zostera marina*, a marine flowering plant. *Molecular Ecology*, **9**: 127-140.

Reuter JA, Spacek D, Snyder MP (2015). High-throughput sequencing technologies. *Molecular Cell*, **58**: 586-597.

Richards AJ (1986). *Plant breeding systems*. George Allen & Unwin: London, UK. p 456.

Rico C, Normandeau E, Dion-Côté A-M, Rico MI, Côté G, Bernatchez L (2013). Combining next-generation sequencing and online databases for microsatellite development in non-model organisms. *Scientific Reports*, **3:** 3376.

Rix EM (1968). *Fritillaria* L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM (eds) *Flora Europaea* vol 5. Cambridge Univ Press: Cambridge, UK. pp 31-34.

Rix E.M (2000). Plate 396. *Fritillaria crassifolia* subsp. *kurdica*. *Curtis's Botanical Magazine*, **17**: 173-175.

Rix EM (2001). *Fritillaria*: *A revised classification: together with an updated list of species*. *Fritillaria* Group of the Alpine Garden Society: Worcestershire, UK. pp 1-14.

Robinson M (2011). *The paleoecology of alluvial hay meadows in the Upper Thames valley*. *Fritillary*, **5**: 47-57.

Rodwell JS (ed.) (1992). *British Plant Communities*. vol. 3, *Grasslands and Montane Communities.* Cambridge University Press: Cambridge, UK.

Rohde ES (1974). *The old English herbals*. Minerva Press Limited: London, UK.

Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**: 1572-1574.

Rønsted N, Law S, Thornton H, Fay MF, Chase MW (2005). Molecular phylogenetic evidence for the monophyly of *Fritillaria* and *Lilium* (Liliaceae; Liliales) and the infrageneric classification of *Fritillaria*. *Molecular Phylogenetics and Evolution*, **35**: 509-527.

Rosenberg NA (2004). Distruct: a program for the graphical display of population struture. *Molecular Ecology Notes*, **4**: 137-138.

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW (2005). Clines, clusters, and the effect of study design on the influence of human population structure. *PloS Genetics* **1**: e70.

Rothero E, Jefferson R, Gowing D (2011). Floodplain Meadows in Great Britain. In: *In Practice, December*. CIEEM: London, UK. pp 4-7.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-Statistics under isolation by distance. *Genetics*, **145**: 1219-1228.

Rousset F (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, **8**: 103-106.

Rousset F, Raymond M (1997). Statistical analyses of population genetic data: new tools, old concepts. *Trends in Ecology and Evolution*, **12**: 313-317.

Rozen S, Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular and Cellular Biology*, **132**: 365-386.

Ruano G, Fenton W, Kidd KK (1989). Biphasic amplification of very dilute DNA samples via a 'booster' PCR. *Nucleic Acids Research*, **17**: 5407.

Rudbeck O (1658). *Catalogus plantarum tam exoticarum quam indigenarum quibus Hortum Academicum Ubsaliensem primum instruxit anno MDCLVII*. Johannes Pauli: Uppsala, Sweden. p 17.

Rychlik W (1993). Selection of primers for polymerase chain reaction. In: White BA (ed.) *Methods in molecular biology, v 15: PCR protocols: current methods and applications*. Humana Press: Totowa, New Jersey, USA. pp 31-40.

Rychlik W (1995). Priming efficiency in PCR. *BioTechniques*, **18**: 84-90.

Saccardo PA (1909). *Cronologia della flora italiana*. Tipografia del Seminario: Padua, Italy. pp 45-46, 377-378.

Sanford M, Fisk R (2010). *A flora of Suffolk*. Henry Ling: Dorchester, UK. p 436.

Saslis-Lagoudakis CH, Savolainen V, Williamson EM, Forest F, Wagstaff SJ, Baral SR, Watson MF, Pendry C, Hawkins JA (2012). Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proceedings of the National Academy of Sciences USA*, **109**: 15835-15840.

Schlumpberger BO, Renner SS (2012). Molecular phylogenetics of *Echinopsis* (Cactaceae): polyphyly at all levels and convergent evolution of pollination modes and growth forms. *American Journal of Botany*, **99**: 1335-1349.

Schmalhausen IF (1897). *Flora of Central and Southern Russia, Crimea and Northern Kazakhstan.* I. N. Kushnerev and Co.: Moscow, USSR. p 508.

Schnittler M, Günther K-F (1999). Central European vascular plants requiring priority conservation measures - an analysis from national Red lists and distribution maps. *Biodiversity and Conservation*, **8**: 891-925.

Schoebel CN, Brodbeck S, Buehler D, Cornejo C, Gajurel J, Hartikainen H, Keller D, Leys M, Říčanová Š, Segelbacher G, Werth S, Csencsics D (2013). Lessons learned from microsatellite development for nonmodel organisms using 454 pyrosequencing. *Journal of Evolutionary Biology*, **26**: 600-611.

Schwartz MK, McKelvey KS (2009). Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**: 441-452.

Selkoe KA, Toonen RJ (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**: 615-629.

Sell P, Murrell G (1996). *Flora of Great Britain and Ireland* vol. 5: *Butomaceae – Orchidaceae*. Cambridge University Press: Cambridge, UK. pp 258-259.

Sernander JR, Sandberg G (1948). *Uppsala Kungsäng*. Almqvist and Wiksell, Uppsala, Sweden.

Shang ZJ, Liu XL (1994). The history of medicinal use of beimu and investigation of its original plants. *Chinese Journal of Medicinal History*, **25**: 38–42.

Shea KL, Furnier GR (2002). Genetic variation and population structure in central and isolated populations of balsam fir, *Abies balsamea* (Pinaceae). *American Journal of Botany*, **89**: 783-791.

Sikes DS, Lewis PO (2001). *PAUPRat: PAUP* Implementation of the Parsimony Ratchet (Beta Software, version 1).* Distributed by the Authors. Department of Ecology and Evolutionary Biology, University of Connecticut: Storrs, USA.

Simpson FW (1982). *Simpson's flora of Suffolk*. Suffolk Naturalists' Society: Ipswich, Suffolk, UK. pp 358-359.

Singer C (1927). The herbal in antiquity and its transmission to later ages. *Journal of Hellenic Studies*, **47**: 1-52.

Singer C (1928). *From magic to science: essays on the scientific twilight*. Boni and Liveright: New York, USA.

Sinngrund (2017). Naturpark Spessart. Available at: https://www.naturpark-spessart.de (last accessed April 2017).

Slatkin M (1985). Rare alleles as indicators of gene flow. *Evolution*, **39**: 53-65.

Slatkin M (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**: 264-279.

Slatkin M, Excoffier L (1996). Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity*, **76**: 377-383.

Sokal RR, Rohlf FJ (1995). *Biometry*. W. H. Freeman and Company: New York, USA. pp 813-819.

Stace CA (2010). Liliaceae. In: *New Flora of the British Isles*. Cambridge University Press: Cambridge, UK. p 857.

Stephens M (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B*, **62**: 795-809.

Stewart JR, Lister AM, Barnes I, Dalén L (2010). Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B*, **277**: 661-671.

Stone C (2015). Flowers from the East. *AramcoWorld*, **65**: 38-46.

Stpiczyńska M, Nepi M, Zych M (2012). Secretion and composition of nectar and the structure of perigonal nectaries in *Fritillaria meleagris* L. (Liliaceae). *Plant Systematics and Evolution*, **298**: 997-1013.

Sun Y, Wang A, Wan D, Wang Q, Liu J (2010). Rapid radiation of *Rheum* (Polygonaceae) and parallel evolution of morphological traits. *Molecular Phylogenetics and Evolution*, **63**: 150-158.

Suzuki JI, Herben T, Maki M (2004). An under-appreciated difficulty: sampling of plant populations for analysis using molecular markers. *Evolutionary Ecology*, **18**: 625-646.

Swofford DL (2003). *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Version 4. Sinauer Associates: Sunderland, Massachusetts, USA.

Szczecińska M, Kwaśniewski M, Chwiałkowska K, Sawicki J (2013). Isolation and characterization of microsatellite loci in *Pulsatilla patens* (L.) Mill. (Ranunculaceae) a rare and endangered plant species in Europe. *Conservation Genetics Resources*, **5**: 421-423.

Szpiech ZA, Rosenberg NA (2011). On the size distribution of private microsatellite alleles. *Theoretical Population Biology*, **80**: 100-113.

Temsch EM, Temsch W, Ehrendorfer-Schratt L, Greilhuber J (2010). Heavy metal pollution, selection and genome size: The species of the žerjav study revisited with flow cytometry. *Journal of Botany*, **2010**: 1-11.

Thermo Fisher Scientific (2009). *NanoDrop 2000/2000c Spectrophotometer version 1.0 User Manual*. Thermo Fisher Scientific: Wilmington, Delaware, USA. pp 3-4.

Thermo Fisher Scientific (2014). *DNA fragment analysis by capillary electrophoresis*. Wilmington, Delaware, USA. pp 165-168, 175.

Thermo Fisher Scientific (2015). *Genome copy number*. Available at: https://www.thermofisher.com/us/en/home/references/protocols/nucleic-acid-amplification-and-expression-profiling/pcr-protocol/pcr-and-rt-pcr.html (last accessed April 2015).

Thick M (1990). Garden seeds in England before the late eighteenth century. I. Seed growing. *Agricultural History Review*, **38**: 58-71.

Thuillet A-C, Bru D, David J, Roumet P, Santoni S, Surdille P, Bataillon T (2002). Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp *durum* desf. *Molecular Biology and Evolution*, **19**: 122-125.

Tigner A (2008). The Flowers of paradise: botanical trade in Sixteenth and seventeenth-century England, In: Sebeck B, Deng S (eds) *Global traffic: discourses and practices of*

trade in English literature and culture from 1550 to 1700. Palgrave: Basingstoke, UK. pp 137-156.

Tomović G, Vukojičić S, Niketić M, Zlatković B, Stevanović V (2007). *Fritillaria* (Liliaceae) in Serbia: distribution, habitats, and some taxonomic notes. *Phytologica Balcanica*, **13**: 359-370.

Tourlet E-H (1908). Catalogue raisonée des plantes vasculaires du Département D'Indre-et-Loire. Paul Klincksieck: Paris, France. p 465.

Trist PJO (1981). *Fritillaria meleagris* L.: its survival and habitats in Suffolk, England. *Biological Conservation*, **20**: 5-14.

Türktas M, Aslay M, Kaya E, Ertugğrul F (2012). Molecular characterization of phylogenetic relationships in *Fritillaria* species inferred from chloroplast trnL-trnF sequences. *Turkish Journal of Botany*, **36**: 552-560.

Turrill WB (1951). The Snake's Head *(Fritillaria meleagris).* In: Synge PM (ed.) *The Lily Year Book 1951-2 (number fifteen).* Royal Horticultural Society: London, UK. pp 108-116.

Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA (1980). *Flora Europaea*. *Alismataceae to Orchidaceae*, vol. 5. Cambridge University Press: Cambridge, UK.

van Asch B, Pinheiro R, (2010). A framework for the development of STR genotyping in domestic animal species: characterisation and population study of 12 canine X-chromosome loci. *Electrophoresis*, **31**: 303-308.

van Oosterhout C, Weetman D, Hutchinson WF (2006). Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes*, **6**: 255-256.

van Paridon BJ, Goater CP, Gilleard JS, Criscione CD (2016). Characterization of nine microsatellite loci for *Dicrocoelium dendriticum*, an emerging liver fluke of ungulates in North America, and their use to detect clonemates and random mating. *Molecular & Biochemical Parasitology*, **207**: 19-22.

Velchev V (ed.) **(**1984**).** *Red data book of Bulgaria. Extinct, threatened, and rare plants and animals. Plants.* vol. 1. Publishing House of the Bulgarian Academy of Sciences: Sofia, Bulgaria.

Vigouroux Y, Jaqueth JS, Yoshihiro M, Smith OS, Beavis WD, Smith JSC, Doebley J (2002). Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, **19**: 1251-1260.

Vinogradov AE. (2003). Selfish DNA is maladaptive: evidence from the plant Red List. *Trends in Genetics*, **19:** 609-614.

von Humboldt FWH, Bonpland AJA, Kunth CS (1815). *Nova genera et species plantarum quas in peregrinatione ad plagam aequinoctialem orbis novi collegerunt Bonpland et Humboldt*. vol. 1. Paris, France. p 288.

Wagner A, Blackstone N, Cartwright P, Dick M, Misof B, Snow P**,** Wagner GP, Bartels J, Murtha M, Pendelton J (1994). Surveys of gene families using polymerase chain-reaction - PCR selection and PCR drift. *Systematic Biology*, **43**: 250-261.

Waldbieser GC, Bilodeau AL, Nonneman DJ (2003). Complete sequence and characterization of the channel catfish mitochondrial genome. *DNA Sequence*, **14**: 265-277.

Walsh PS, Fildes NJ, Reynolds R (1996). Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Research*, **24**: 2807-2812.

Wang C, Schroeder KB, Rosenberg NA (2012). A maximum-likelihood method to correct for allelic dropout in microsatellite data with no replicate genotypes. *Genetics*, **192**: 651-669.

Ward SS, Jasieniuk M (2009). Sampling weedy and invasive plant populations for genetic diversity analysis. *Weed Science*, **57**: 593-602.

Warren J, Mackenzie S (2001). Why are all colour combinations not equally represented as flower-colour polymorphisms. *New Phytologist*, **151**: 237-241.

Watterson GA (1984). Allele frequencies after a bottleneck. *Theoretical Population Biology*, **26**: 387-407.

WCSP (2014). *World Checklist of Selected Plant Families*. Facilitated by the Royal Botanic Gardens, Kew. Available at: http://apps.kew.org/wcsp (last accessed March 2013).

WCSP (2017). *World Checklist of Selected Plant Families*. Facilitated by the Royal Botanic Gardens, Kew. Available at: http://wcsp.science.kew.org (last accessed September 2017).

Wells DA (1994). *Fritillaria meleagris* L. In: Stewart A, Pearman DA, Preston CD (eds) *Scarce plants in Britain*. Joint Nature Conservation Committee: Peterborough, UK. pp 173-175.

Wells DA, King N (1975). *The flora of neutral grasslands in Wiltshire*. Nature Conservancy Council: Huntingdon, UK.

Wenz HM, Robertson JM, Menchen S, Oaks F, Demorest DM, Scheibler D, Rosenblum BB, Wike C, Gilbert DA, Efcavitch JW (1998). High-precision genotyping by denaturing capillary electrophoresis. *Genome Research*, **8**: 69-80.

Wietsma W, van den Berg RG (2012). Cultivating *Fritillaria eduardii*. *The Plantsman*, **11**: 222-225.

Wigginton JE, Cutler DJ, Abecasis GR (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, **76**: 887-893.

Wilke ABB, Vidal PO, Suesdek L, Marrelli MT (2014). Population genetics of neotropical *Culex quinquefasciatus* (Diptera: Culicidae). *Parasites and Vectors*, **7**: 468-476.

Willes M (2011). *The making of the English Gardener: plants, books and inspiration, 1560-1660*. Yale University Press: New Haven, Connecticut, USA. p 109.

Wogan GOU, Kapelke J, Feldheim KA, Papenfuss TJ, Bowie RCK (2015). Isolation and characterization of nine tetranucleotide microsatellite loci for the secretive limbless lizards of the genus *Anniella* (Anguidae). *Biochemical Systematics and Ecology*, **62**: 155-58.

Wotton E, Gessner C, Penny T (1634). *Insectorum sive minimorum animalium theatrum*. Thomas Cotes for Benjamin Allen: London, UK. p 269.

Wright S (1943). Isolation by distance. *Genetics*, **28**: 114-138.

Xiao P-G, Jiang Y, Li P, Luo Y-B, Liu Y (2007). The botanical origin and pharmacophylogenetic treatment of Chinese materia medica beimu. *Acta Phytotaxonomica Sinica*, **45**: 473-487.

Xinqi C, Mordak HV (2000). *Fritillaria.* In: Wu ZY, Raven PH (eds) *Flora of China*, vol. 24 (Flagellariaceae through Marantaceae). Science Press: Beijing, China and Missouri Botanical Garden Press: St. Louis, Missouri, USA. pp 127-133.

Yan ZH, Yan L, Ding WG (1999). Studies on the chemical constituents of *Fritillaria Cirrhosa* D. Don. *Acta Universitatis Medicinalis Secondae Shanghai*, **19**: 487-489.

Zaharof E (1987). *Fritillaria rhodocanakis* subsp. *argolica* (Liliaceae), a new subspecies from Peloponnese, Greece. *Willdenowia*, **16**: 343-348.

Zhang D-Q, Gao L-M, Yang, Y-P (2010). Genetic diversity and structure of a traditional Chinese medicinal plant species, *Fritillaria cirrhosa* (Liliaceae), in southwest China and implications for its conservation. *Biochemical Systematics and Ecology*, **38**: 236-242.

Zhang L (1983). *Vegetation ecology and population biology of Fritillaria meleagris L. at the Kungsängen Nature Reserve, Eastern Sweden*. In: Acta Phytogeographica Suecica 73. Almqvist and Wiksell International: Stockholm, Sweden. pp 53, 55, 78-79.

Zonneveld BJM (2010). New record holders for maximum genome size in eudicots and monocots. *Journal of Botany*: Article ID 527357.

Zych M, Stpiczyńska M (2012). Neither protogynous nor obligatory out-crossed: pollination biology and breeding system of the European Red List *Fritillaria meleagris*. *Plant Biology*, **14**: 285-294.

Appendix A

Supplementary Tables

Table S1 - Plant material sampled.

Species Name[a] (sample number)[b]	Living Collection Accession[c]	Voucher Details[d]	DNA Bank Number(s)[e]
***Fritillaria* L.**			
subgenus *Davidii* Rix			
Fritillaria davidii Franch.[f]	LH 739	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones739.pdf)	33589
subgenus *Fritillaria*			
Fritillaria acmopetala subsp. *acmopetala* Boiss.	Kew 1959-59401	Chase 2565 K	2565
Fritillaria alburyana Rix	Kew 1994-3139	Chase 3470 K	3470
Fritillaria alfredae Post subsp. *glaucoviridis* (Turrill) Rix	LH 744	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones744.pdf)	37858
Fritillaria amana (Rix) R.Wallis & R.B.Wallis (1)	Kew 1974-2043	Chase 2563 K	2563
Fritillaria amana (Rix) R.Wallis & R.B.Wallis (2)	LH 430	Photo (www.fritillariaicones.com/icones/ic400/Fritillaria_Icones430.pdf)	31519
Fritillaria armena Boiss[f]	LH 726	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones726.pdf)	40598
Fritillaria assyriaca subsp. *assyriaca* Baker (1)	LH 609	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones609.pdf)	37767
Fritillaria assyriaca subsp. *assyriaca* Baker (2)	LH 668	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones668.pdf)	40661
Fritillaria aurea Schott	Kew 1973-21448	Chase 3487 K	3487
Fritillaria bithynica Baker	Kew 1989-3037	Photo (Fig. S1)	24332
Fritillaria carica Rix	Kew 1999-2082	Chase 24330 K	24330
Fritillaria caucasica Adam	Kew 1989-1112	Chase 3488 K	3488
Fritillaria cirrhosa D. Don	n/a	TCMK 727 K	25486
Fritillaria conica Boiss.	LH 490	Photo (www.fritillariaicones.com/icones/ic400/Fritillaria_Icones490.pdf)	31540
Fritillaria crassifolia Boiss. & Huet subsp. *kurdica* (Boiss. & Noë) Rix (1)	Kew 1985-926	Chase 2559 K	2559
Fritillaria crassifolia Boiss. & Huet subsp. *kurdica* (Boiss. & Noë) Rix (2)	LH 610	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones610.pdf)	31556
Fritillaria crassifolia subsp. *crassifolia* Boiss. & Huet	LH 065	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones065.pdf)	37862
Fritillaria davisii Turrill	Kew 1979-885	s.n. 1985 K	31558
Fritillaria drenovskii Degen & Stoj.	LH 727	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones727.pdf)	40600
Fritillaria ehrhartii Boiss. & Orph.	LH 105	Photo (www.fritillariaicones.com/icones/ic100/Fritillaria_Icones105.pdf)	31521
Fritillaria elwesii Boiss.	LH 605	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones605.pdf)	31541
Fritillaria fleischeriana Steud. & Hochst. ex Schult.f.	LH 804	Photo (www.fritillariaicones.com/icones/ic800/Fritillaria_Icones804.pdf)	40663
Fritillaria forbesii Baker	LH 714	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones714.pdf)	40601
Fritillaria frankiorum R.Wallis & R.B.Wallis	LH 091	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones091.pdf)	31523/37769
Fritillaria graeca Boissier & Spruner	Kew 1959-59103	s.n. 1975 K	31560
Fritillaria gussichiae (Degen & Dorfler) Rix	Kew 1999-2085	Chase 21896 K	24336
Fritillaria involucrata Allioni (1)	Kew 2000-3114	Photo (Fig. S1)	24337
Fritillaria involucrata Allioni (2)	LH 076	Photo (Fig. S1)	31542
Fritillaria kotschyana Herbert	LH 500	Photo (www.fritillariaicones.com/icones/ic500/Fritillaria_Icones500.pdf)	31326
Fritillaria latakiensis Rix	LH 090	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones090.pdf)	31544
Fritillaria latifolia Willd.	LH 888	Photo (www.fritillariaicones.com/icones/ic800/Fritillaria_Icones888.pdf)	40602
Fritillaria lusitanica Wikstr.	Kew 1980-3020	Chase 2603 K	2603
Fritillaria meleagris subsp. *meleagris* L.	Kew 1990-3088	Chase 2566 K	2566
Fritillaria meleagroides Patrin ex Schult.f.	Kew 2005-2045	Photo (Fig. S1)	24371
Fritillaria messanensis subsp. *messanensis* Rafin	Kew 1984-3711	Chase 21897 K	24341

Fritillaria michailovskyi Fomin	Kew 1981-3060	Chase 2583 K	2583
Fritillaria minuta Boiss. & Noe	Kew 1978-3492	Chase 2562 K	2562
Fritillaria montana Hoppe ex W. D. J. Koch (1)	Kew 2004-3474	Chase 24342 K	24342
Fritillaria montana Hoppe ex W. D. J. Koch (2)	LH 612	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones612.pdf)	40603
Fritillaria mutabilis Kamari	LH 491	Photo (www.fritillariaicones.com/icones/ic400/Fritillaria_Icones491.pdf)	31329
Fritillaria obliqua Ker Gawl. subsp. *tuntasia* (Heldr.ex Halácsy) Kamari (1)	Kew 1959-2703	s.n. 1995 K	31571
Fritillaria obliqua Ker Gawl. subsp. *tuntasia* (Heldr.ex Halácsy) Kamari (2)	LH 085	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones085.pdf)	37774
Fritillaria obliqua subsp. *obliqua* Ker Gawl.	LH 068	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones068.pdf)	40667
Fritillaria olivieri Baker	Kew 1963-51207	Chase 2569 K	2569
Fritillaria oranensis Pomel	n/a	Photo (Fig. S1)	35345
Fritillaria orientalis Adam (1)	Kew 2004-3475	Chase 24345 K	24345
Fritillaria orientalis Adam (2)	Kew 1994-1338	s.n. 1997 K	31501
Fritillaria pallidiflora Schrenk ex Fischer & C. A. Meyer	Kew 1959-1103	Chase 2567 K	2567
Fritillaria pinardii Boiss.	Kew 1989-3044	Chase 31564 K	31564
Fritillaria poluninii (Rix) G.Bakhshi Khaniki & K.Persson	LH 178	Photo (www.fritillariaicones.com/icones/ic100/Fritillaria_Icones178.pdf)	40599
Fritillaria pontica Wahlenb.	Kew 1975-2438	s.n. 1978 K	31565
Fritillaria pyrenaica subsp. *pyrenaica* L.	LH 948	Photo (www.fritillariaicones.com/icones/ic900/Fritillaria_Icones948.pdf)	38749
Fritillaria reuteri Boiss.	Kew 1969-6106	Chase 2568 K	2568
Fritillaria rixii Zaharof	Kew 2007-1836	Chase 40654 K	40654
Fritillaria ruthenica Wikstr.	Kew 2004-3479	Photo (Fig. S1)	24362
Fritillaria sibthorpiana Baker subsp. *enginiana* A.Byfield & Özhatay	LH 094	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones094.pdf)	40607
Fritillaria sibthorpiana subsp. s*ibthorpiana* Baker	LH 108	Photo (www.fritillariaicones.com/icones/ic100/Fritillaria_Icones108.pdf)	31530
Fritillaria sororum Jimmy Persson & K.Persson (1)	LH 814	Photo (www.fritillariaicones.com/icones/ic800/Fritillaria_Icones814.pdf)	40608
Fritillaria sororum Jimmy Persson & K.Persson (2)	Kew 2007-1838	Chase 40658 K	40658
Fritillaria stribrnyi Velen.	LH 398	Photo (www.fritillariaicones.com/icones/ic300/Fritillaria_Icones398.pdf)	31532
Fritillaria theophrasti Kamari & Phitos	LH 669	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones669.pdf)	40670
Fritillaria thessala (Boiss.) Kamari subsp. *reiseri* Kamari	LH 142	Photo (www.fritillariaicones.com/icones/ic100/Fritillaria_Icones142.pdf)	40671
Fritillaria thessala subsp. *thessala* (Boiss.) Kamari	Kew 1999-2084	Chase 24338 K	24338
Fritillaria thunbergii Miq.	LH 015	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones015.pdf)	37780
Fritillaria tortifolia X.Z.Duan & X.J.Zheng	n/a	Photo (Fig. S1)	31505
Fritillaria tubiformis subsp. *tubiformis* Gren. & Godr.	Kew 1966-109	Chase 2558 K	2558
Fritillaria usuriensis Maxim. (1)	Kew 2005-2048	Chase 31572 K	24369
Fritillaria usuriensis Maxim. (2)	LH 202	Photo (www.fritillariaicones.com/icones/ic200/Fritillaria_Icones202.pdf)	40660
Fritillaria uva-vulpis Rix	LH 821	Photo (www.fritillariaicones.com/icones/ic800/Fritillaria_Icones821.pdf)	24364/37781
Fritillaria verticillata Willd.	Kew 2005-2049	Photo (Fig. S1)	24363
Fritillaria walujewii Regel (1)	Kew 2005-2051	Chase 31573 K	24368
Fritillaria walujewii Regel (2)	LH 185	Photo (Fig. S1)	40609
Fritillaria whittallii Baker	n/a	Rix M. 17[9]	31331
Fritillaria zagrica Stapf	n/a	Zarrei, M. & Kamrani, A. 35188 TUH	23344
subgenus *Japonica* Rix			
Fritillaria amabilis Koidzumi (1)	LH 792	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones792.pdf)	35342/37766
Fritillaria amabilis Koidzumi (2)	LH 785	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones785.pdf)	40596
Fritillaria ayakoana I.Maruyama & N.Naruhashi	LH 712	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones712.pdf)	33590
Fritillaria japonica Miq.	LH 702	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones702.pdf)	33592
Fritillaria kaiensis N.Naruhashi	LH 847	Photo (Fig. S1)	37831

Fritillaria koidzumiana Ohwi[f]	LH 485	Photo (www.fritillariaicones.com/icones/ic400/Fritillaria_Icones485.pdf)	31496
Fritillaria muraiana Ohwi (1)	LH 699	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones699.pdf)	33594
Fritillaria muraiana Ohwi (2)	LH 945	Photo (www.fritillariaicones.com/icones/ic900/Fritillaria_Icones945.pdf)	40666
Fritillaria shikokiana N.Naruhashi (1)	LH 701	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones701.pdf)	33593
Fritillaria shikokiana N.Naruhashi (2)	LH 789	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones789.pdf)	37777
Fritillaria tokushimensis Akasawa, Katayama & T.Naito	n/a	Photo (Fig. S1)	37919
subgenus *Korolkovia* Rix			
Fritillaria sewerzowii Regel (1)	n/a	Chase 743 K	743
Fritillaria sewerzowii Regel (2)	Kew 1995-4397	Chase 37751 K	37751
Fritillaria sewerzowii Regel (3)	LH 029	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones029.pdf)	37863
subgenus *Liliorhiza* (Kellogg) Benth. & Hook. f.			
Fritillaria affinis (Schultes) Sealy (1)	Kew 1989-2020	Chase 446 K	446
Fritillaria affinis (Schultes) Sealy (2)	Kew 2010-905	Chase 31485 K	33601
Fritillaria agrestis Greene	LH 110	Photo (Fig. S1)	31534
Fritillaria atropurpurea Nutt.	Kew 1999-255	Chase 31488 K	31488
Fritillaria camschatcensis (L) Ker Gawl. (1)	n/a	Chase 3580 K	3580
Fritillaria camschatcensis (L) Ker Gawl. (2)	LH 019	Photo (Fig. S1)	31537
Fritillaria camschatcensis (L) Ker Gawl. (3)	LH 617	Photo (www.fritillariaicones.com/icones/ic600/Fritillaria_Icones617.pdf)	31539
Fritillaria camschatcensis (L) Ker Gawl. (4)	LH 738	Photo (www.fritillariaicones.com/icones/ic700/Fritillaria_Icones738.pdf)	33675
Fritillaria camschatcensis (L) Ker Gawl. (5)	LH 955	Photo (Fig. S1)	41093
Fritillaria dagana Turczaninow[f]	Kew 2005-2036	Photo (Fig. S1)	37856
Fritillaria eastwoodiae R.M.Macfarlane	Kew 1986-6357	Chase 8980 K	8993
Fritillaria falcata (Jeps.) D.F.Beetle	Kew 1989-3270	Chase 3478 K	3478
Fritillaria liliacea Lindl.	LH 305	Photo (Fig. S1)	37772
Fritillaria maximowiczii Freyn (1)[h]	Kew 1993-3390	Chase 8195 K	8195
Fritillaria maximowiczii Freyn (2)	Kew 2005-2043	Chase 31497 K	33600
Fritillaria micrantha A.Heller	LH 115	Photo (www.fritillariaicones.com/icones/ic100/Fritillaria_Icones115.pdf)	31526
Fritillaria pluriflora Torr. ex Benth.	LH 084	Photo (www.fritillariaicones.com/icones/ic001/Fritillaria_Icones084.pdf)	37775
Fritillaria pudica (Pursh) Sprengel	Kew 1986-6110	Chase 754 K	24359
Fritillaria recurva Benth.	Kew 1989-122	Chase 2560 K	2560
subgenus *Petilium* (L.) Endl.			
Fritillaria chitralensis (hort.) B. Mathew	Kew 1970-4019	Chase 3472 K	3472
Fritillaria eduardii A. Regel ex Regel	Kew 1992-27	Chase 31491 K	24350
Fritillaria imperialis L.	Kew 1970-3943	Chase 2557 K	2557
Fritillaria raddeana Regel	Kew 1973-54	Chase 745 K	745
subgenus *Rhinopetalum* Fisch.			
Fritillaria bucharica Regel	LH 488	Photo (www.fritillariaicones.com/icones/ic400/Fritillaria_Icones488.pdf)	37861
Fritillaria gibbosa Boiss.	Kew 2004-3469	Chase 31559 K	31559
Fritillaria karelinii (Fisch . ex D. Don) Baker	Kew 1994-218	Chase 3471 K	3471
Fritillaria stenanthera (Regel) Regel[f]	Kew 1990-304	Chase 33602 K	
subgenus *Theresia* Koch			
Fritillaria persica L.	Kew 1923-41201	Chase 3496 K	3496
Lilium L.			3697
Lilium davidii Duch. ex Elwes	Kew 1979-867	Chase 3697 K	3723
Lilium humboldtii subsp. *ocellatum* (Kellogg) Thorne	Kew 1994-561	Chase 3723 K	846

Lilium lophophorum (Bureau & Franch.) Franch.	Kew 1991-1997	Chase 846 K	33676
Lilium mackliniae Sealy[f]	LH L019	Photo (Fig. S1)	931
Lilium maritimum Kellogg	Kew 1977-1830	Chase 931 K	3698
Lilium martagon L.	Kew 1978-2452	Chase 3698 K	933
Lilium pensylvanicum Ker Gawl.	Kew 1987-2468	Chase 933 K	37918
Lilium pyrenaicum Gouan	Kew 1995-1667	Chase 8639 K	3696
Lilium regale E.H.Wilson	Kew 1979-879	Chase 3696 K	932
Lilium rubescens S.Watson	Kew 1989-197	Chase 932 K	3724
Lilium speciosum var. *speciosum* Thunb.	Kew 1993-332	Chase 3724 K	112
Lilium superbum L.	n/a	Chase 112 NCU	3688
Lilium washingtonianum subsp. *washingtonianum* Kellogg	Kew 1988-406	Chase 3688 K	
Outgroups			16314
Calochortus tolmiei Hook. & Arn.	Kew 1987-1469	Chase 16314 K	3689
Cardiocrinum giganteum var. *giganteum* (Wall.) Makino	Kew 1988-4907	Chase 3689 K	498
Clintonia borealis (Aiton) Raf.	Kew 1981-6330	Chase 498 K	35323
Gagea reticulata (Pall.) Schult. & Schult.f.	n/a	Zarrei, M. & Golzarian, K. 35258 K	934
Nomocharis pardanthina Franch.	Kew 1990-3521	Chase 934 K	14242
Notholirion macrophyllum (D.Don) Boiss.	Kew 1971-264	Chase 14242 K	448
Notholirion thomsonianum (Royle) Stapf	Kew 1970-4025	Chase 448 K	441
Scoliopus bigelovii Torr.	Kew 1976-1909	Chase 441 K	2782
Tricyrtis maculata (D.Don) J.F.Macbr.	Kew 1992-202	Chase 2782 K	23828
Tulipa heteropetala Ledeb.	Kew 1990-2587	Chase 23828 K	

[a]Names given for *Fritillaria* are those accepted according to the list compiled by Hill (2013; see: www.fritillariaicones.com/info/names/frit.names.pdf). Names for other genera are those accepted according to the World Checklist of Selected Plant Families (WCSP, 2014). World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://apps.kew.org/wcsp/). The following *Fritillaria* names are accepted according to Hill (2013) but treated as synonyms by the WCSP (accepted names according to the WCSP are given in parentheses): *Fritillaria armena* (*Fritillaria pinardii* subsp. *pinardii*), *Fritillaria chitralensis* (*Fritillaria imperialis* var. *chitralensis*), *Fritillaria crassifolia* subsp. *kurdica* (*Fritillaria kurdica*), *Fritillaria graeca* (*Fritillaria graeca* subsp. *graeca*), *Fritillaria meleagris* subsp. *meleagris* (*Fritillaria meleagris*), *Fritillaria poluninii* (*Fritillaria crassifolia* subsp. *poluninii*), *Fritillaria sororum* (*Fritillaria acmopetala* subsp. *acmopetala*), *Fritillaria thessala* subsp. *reiseri* (*Fritillaria graeca* subsp. *thessala*), *Fritillaria thessala* subsp. *thessala* (*Fritillaria graeca* subsp. *thessala*), *Fritillaria zagrica* (*Fritillaria pinardii* subsp. *pinardii*).

[b] For species with multiple samples, a sample number is given in parentheses after the species name.

[c] Accession numbers are for the Living Collection at the Royal Botanic Gardens, Kew (Kew) or Laurence Hill's collection at Petersham Lodge (LH; see www.fritillariaicones.com/projects/project.html for details).

[d] Herbarium abbreviations follow Index Herbariorum (Holmgren PK, Holmgren NH, Barnett LC (1990). *Index Herbariorum Part 1: The Herbaria of The World.* Regnum Vegetabile, Vol. 120. New York Botanical Garden: New York). Photographic vouchers are available via the URLs indicated, or are provided in Figure S1.

[e] Numbers correspond to accessions in the DNA Bank at the Royal Botanic Gardens Kew: http://data.kew.org/dnabank/homepage.html. Where two numbers are given, these refer to independent DNA extractions from the same living collection accession.

[f] Species included in screen of low-copy nuclear genes; for *F. koidzumiana* DNA Bank number 33599 was used instead of the one listed here (see http://data.kew.org/dnabank/homepage.html for further accession details).

[g] Personal herbarium of Martyn Rix. Contact details: Martyn Rix c/o The Herbarium, RBG Kew, Richmond, Surrey TW9 3AB, UK.

[h] Both accessions of *F. maximowiczii* derive from the same original collection and therefore represent a single population.

Table S2 - Primer sequences and amplification conditions for plastid regions.

Region	Primer Sequences (5′ - 3′)	References[b]	Amplification Conditions
matK	-19F - CGT TCT GAC CAT ATT GCA CTA TG 1565R - TCA CCA GGT CAT TGA CAC GAA 568F[a] - CCT TCA ATG CTG GAT TCA AGA	Molvray *et al*. (2000; -19F) Zarrei *et al*. (2009; 1565R) M. Zarrei personal communication (568F)	94°C 2.5 min; 30 cycles of 95°C 1 min, 50°C 1 min, 65°C 4 min; 65°C 5 min.
rbcL	1F - ATG TCA CCA CAA ACA GAA AC 1360R - CTT CAC AAG CAG CAG CTA GTT C	Fay *et al*. (1997; 1F) Reeves *et al*. (2001; 1360R)	94°C 3 min; 28 cycles of 94°C 1 min, 48°C 1 min, 72°C 1 min; 72°C 7 min.
rpl16	259F - TAA AGG GAT GTG GAT AAA TGG 1661R - CGT ACC CAT ATT TTT CCA CCA CGA C	This study[c] (259F) Jordan *et al*. (1996; 1661R)	94°C 2.5 min; 30 cycles of 95°C 1 min, 50°C 1 min, 65°C 4 min; 65°C 5 min.

[a] Internal primer; used for sequencing only.

[b] References:

Fay MF, Swensen SM, Chase MW (1997). Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin*, **52**: 111-120.

Jordan WC, Courtney MW, Neigel JE (1996). Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany*, **83**: 430-439.

Molvray MP, Kores PJ, Chase MW (2000). Polyphyly of mycoheterotrophic orchids and functional influences on floral and molecular characters. In: Wilson KL, Morrison D A (eds) *Monocots: Systematics and Evolution*. CSIRO, Melbourne, Australia, pp. 441-448.

Reeves G, Chase MW, Goldblatt P, Rudall P, Fay MF, Cox AV, Lejeune B, Souza-Chies T (2001). Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany*, **88**: 2074-2087.

Zarrei M, Wilkin P, Fay MF, Ingrouille MJ, Zarre S, Chase MW (2009). Molecular Systematics of *Gagea* and *Lloydia* (Liliaceae; Liliales): implications of analyses of nuclear ribosomal and plastid sequences for infrageneric classification. *Annals of Botany*, **104**: 125-142.

[c] The 259F primer was designed to avoid polyT and polyA regions downstream of the 158F primer previously used in *Fritillaria* (Rønsted *et al*., 2005; see main manuscript for full reerence).

Table S3 - Primer sequences and amplification conditions for genes included in the screen of low-copy nuclear genes.

Gene	Primer Sequences (5′ - 3′)	Reference[a]	Amplification Conditions
AKT1 (*AKT1*-like potassium channel gene)	AKT5f - AGA GAC TCT TGA TGC ACT TCC TAA A AKT10r - AAG AGA ACA ACA CAA CTT TCA TTC C	Douglas *et al*. (2011)	95°C 4 min; 35 cycles of 95°C 30 s, 58°C 30 s, 72°C 2.5 min; 72°C 4 min.
AP (*ALKALINE PHYTASE)*	AP8f - TCT CCT TGG GCT CTT TCT TG AP10r - GAA AAC CTC AAA TGG GCA GAG	Douglas *et al*. (2011)	95°C 4 min; 35 cycles of 95°C 30 s, 58°C 30 s, 72°C 2.5 min; 72°C 4 min.
CHS (*CHALONE SYNTHASE*)	CHS Lil 116F - ACT ACT ACT TCC GCA TCA CC CHSX 2RN - TTC AGT CAA GTG CAT GTA ACG	This study (116F) & Strand *et al*. (1997; CHSX 2RN)	95°C 2 min; 35 cycles of 95°C 1 min, 48°C 1.5 min, 72°C 2 min; 72°C 9 min.
G3PDH (*GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE*)	GPDX 7F - GAT AGA TTT GGA ATT GTT GAG G GPDX 9R - AAG CAA TTC CAG CCT TGG	Strand *et al*. (1997)	95°C 2 min; 35 cycles of 95°C 1 min, 44°C 1.5 min, 72°C 2 min; 72°C 9 min.
GS (*GLUTAMINE SYNTHETASE*, plastid expressed)	GSCP 687F - GAT GCT CAC TAC AAG GCT TG GSCP 994R - AAT GTG CTC TTT GTG GCG AAG	Emshwiller & Doyle (1999)	95°C 5 min; 35 cycles of 95°C 1 min, 48°C 1 min, 72°C 2 min; 72°C 7 min.
LFY/FLO (*LEAFY/FLORICAULA*)	LFY Lil 252F - TGT TGA TGG AAT GAT GGC GG LFY Lil 693R - CGT CAC GAT GAA CGG ATG C	This study	94°C 4 min; 35 cycles of 94°C 1 min, 50°C 1 min, 72°C 2 min; 72°C 10 min.
LFY/FLO (*LEAFY/FLORICAULA*)	LFY Lil 675F - GCA TCC GTT CAT CGT GAC G LFY Lil 992R - CAA GCT CCG ACG TTC TCC	This study	94°C 4 min; 35 cycles of 94°C 1 min, 50°C 1 min, 72°C 2 min; 72°C 10 min.
MS (MALATE SYNTHASE)	MS 400F - GGA AGA TGR TCA TCA AYG CNC TYA AYT C MS 943R - GTC TTN ACR TAG CTG AAD ATR TAR TCC C	Lewis & Doyle (2001)	94°C 4 min; 35 cycles of 94°C 1 min, 55°C 1 min, 72°C 2 min; 72°C 7 min.
PRK (*PHOSPHORIBULOKINASE*)	PRK 488F - AAY GAY TTT GAY CTY ATG TAT GAR CAR GT PRK 1167R - ATG GTY TGR AAN ARA CCN GTN CCR TTG TTG C	Lewis & Doyle (2002)	94°C 4 min; 35 cycles of 94°C 1 min, 55°C 1 min, 72°C 2 min; 72°C 7 min.
RPB2 (*RNA POLYMERASE II*)	RPB2-10F - CAR GAR GAT ATG CCA TGG AC RPB2 INT 23R - CCA CGC ATC TGA TAT CCA C	Denton *et al*. (1998; 10F) & Roncal *et al*. (2005; INT 23R)	94°C 4 min; 35 cycles of 94°C 1 min, 55°C 1 min, 72°C 2 min; 72°C 7 min.
XDH (*XANTHINE DEHYDROGENASE*)	X551F - GAA GAG CAG ATT GAA GAW WGC C XDH975R - TGC TCC WGC AGC CAT CCA GAG	Górniak *et al*. (2010; 551F) & Morton (2011; 975R)	94°C 2 min; 6 cycles of 94°C 45 s, 55°C-49°C 45 s (reduced by one degree per cycle), 72°C 1.5 min; 28 cycles of 94°C 45 s, 49°C 45 s, 72°C 1.5 min; 72°C 5 min.

[a] References:

Denton AL, McConaughy BL, Hall BD (1998). Usefulness of RNA Polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution*, **15**: 1082-1085.

Douglas NA, Wall WA, Xiang Q-Y, Hoffmann WA, Wentworth TR, Gray JB, Hohmann MG (2011). Recent vicariance and the origin of the rare, edaphically specialized Sandhills lily, *Lilium pyrophilum* (Liliaceae): evidence from phylogenetic and coalescent analyses. *Molecular Ecology*, **20**: 2901-2915.

Emshwiller E, Doyle JJ (1999). Chloroplast-expressed glutamine synthetase (ncpGS): potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). *Molecular Phylogenetics and Evolution*, **12**: 310-319.

Górniak M, Paun O, Chase MW (2010). Phylogenetic relationships within Orchidaceae based on a low-copy nuclear coding gene, *Xdh*: congruence with organellar and nuclear ribosomal DNA results. *Molecular Phylogenetics and Evolution*, **56**: 784-795.

Lewis CE, Doyle JJ (2001). Phylogenetic utility of the nuclear gene malate synthase in the palm family (Arecaceae). *Molecular Phylogenetics and Evolution*, **19**: 409-420.

Lewis CE, Doyle JJ (2002). Phylogenetics of tribe Areceae (Arecaceae) using low-copy nuclear genes. *Plant Systematics and Evolution*, **236**: 1-17.

Morton C (2011). Newly sequenced nuclear gene (*Xdh*) for inferring angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **98**: 63-89.

Roncal J, Francisco-Ortega J, Asmussen CB, Lewis CE, (2005). Molecular phylogenetics of tribe Geonomeae (Arecaceae) using nuclear DNA sequences of phosphoribulokinase and RNA polymerase II. *Systematic Botany*, **30**: 275-283.

Strand AE, Leebens-Mack J, Milligan BG (1997). Nuclear DNA based markers for plant evolutionary biology. *Molecular Ecology*, **6**: 113-118.

Table S4 - GenBank accession numbers.

Species Name (sample number)	*mat*K	*rbc*L	*rpl*16
Fritillaria acmopetala subsp. *acmopetala*	AY624426	LM992907	AY624372
Fritillaria affinis (1)	AY624427	LM992908	AY624373
Fritillaria affinis (2)	LM993042	LM992909	LM993151
Fritillaria agrestis	LM993043	LM992910	LM993152
Fritillaria alburyana	AY624429	LM992911	AY624375
Fritillaria alfredae subsp. *glaucoviridis*	LM993044	LM992912	LM993153
Fritillaria amabilis (1)	LM993045	LM992913	LM993154
Fritillaria amabilis (2)	LM993046	LM992914	LM993155
Fritillaria amana (1)	LM993047	LM992915	LM993156
Fritillaria amana (2)[a]	AY624440	LM992916	AY624387
Fritillaria armena	LM993048	LM992917	LM993157
Fritillaria assyriaca subsp. *assyriaca* (1)	LM993049	LM992918	LM993158
Fritillaria assyriaca subsp. *assyriaca* (2)	LM993050	LM992919	LM993159
Fritillaria atropurpurea	LM993051	LM992920	LM993160
Fritillaria aurea	AY624430	LM992921	AY624376
Fritillaria ayakoana	LM993052	LM992922	LM993161
Fritillaria bithynica	LM993053	LM992923	LM993162
Fritillaria bucharica	LM993054	LM992924	LM993163
Fritillaria camschatcensis (1)	AY624431	LM992925	AY624377
Fritillaria camschatcensis (2)	LM993055	LM992926	LM993164
Fritillaria camschatcensis (3)	LM993056	LM992927	LM993165
Fritillaria camschatcensis (4)	LM993057	LM992928	LM993166
Fritillaria camschatcensis (5)	LM993058	LM992929	LM993167
Fritillaria carica	LM993059	LM992930	LM993168
Fritillaria caucasica	AY624432	LM992931	AY624378
Fritillaria chitralensis	AY624433	LM992932	LM993169
Fritillaria cirrhosa	LM993060	LM992933	LM993170
Fritillaria conica	LM993061	LM992934	LM993171
Fritillaria crassifolia subsp. *crassifolia*	LM993062	LM992935	LM993172
Fritillaria crassifolia subsp. *kurdica* (1)	AY624434	LM992936	AY624380
Fritillaria crassifolia subsp. *kurdica* (2)	LM993063	LM992937	LM993173
Fritillaria dagana	LM993064	LM992938	LM993174
Fritillaria davidii	LM993065	LM992939	LM993175
Fritillaria davisii	LM993066	LM992940	LM993176
Fritillaria drenovskii	LM993067	LM992941	LM993177
Fritillaria eastwoodiae	LM993068	LM992942	AY624382
Fritillaria eduardii	LM993069	LM992943	LM993178
Fritillaria ehrhartii	LM993070	LM992944	LM993179
Fritillaria elwesii	LM993071	LM992945	LM993180
Fritillaria falcata	AY624436	LM992946	LM993181
Fritillaria fleischeriana	LM993072	LM992947	LM993182
Fritillaria forbesii	LM993073	LM992948	LM993183
Fritillaria frankiorum	LM993074	LM992949	LM993184
Fritillaria gibbosa	LM993075	LM992950	LM993185
Fritillaria graeca	LM993076	LM992951	LM993186
Fritillaria gussichiae	LM993077	LM992952	LM993187
Fritillaria imperialis	AY624441	LM992953	AY624388
Fritillaria involucrata (1)	LM993078	LM992954	LM993188
Fritillaria involucrata (2)	LM993079	LM992955	LM993189
Fritillaria japonica	LM993080	LM992956	LM993190
Fritillaria kaiensis	LM993081	LM992957	LM993191
Fritillaria karelinii	LM993082	n/a	AY624390
Fritillaria koidzumiana	LM993083	LM992958	LM993192
Fritillaria kotschyana	LM993084	LM992959	LM993193
Fritillaria latakiensis	LM993085	LM992960	LM993194
Fritillaria latifolia	LM993086	LM992961	LM993195

Fritillaria liliacea	LM993087	LM992962	LM993196
Fritillaria lusitanica	AY624443	LM992963	AY624391
Fritillaria maximowiczii (1)	AY624444	LM992964	AY624392
Fritillaria maximowiczii (2)	LM993088	LM992965	LM993197
Fritillaria meleagris subsp. *meleagris*	AY624445	LM992966	AY624393
Fritillaria meleagroides	LM993089	LM992967	LM993198
Fritillaria messanensis subsp. *messanensis*	LM993090	LM992968	LM993199
Fritillaria michailovskyi	AY624446	LM992969	LM993200
Fritillaria micrantha	LM993091	LM992970	LM993201
Fritillaria minuta	AY624448	LM992971	AY624396
Fritillaria montana (1)	LM993092	LM992972	LM993202
Fritillaria montana (2)	LM993093	LM992973	LM993203
Fritillaria muraiana (1)	LM993094	LM992974	LM993204
Fritillaria muraiana (2)	LM993095	LM992975	LM993205
Fritillaria mutabilis	LM993096	LM992976	LM993206
Fritillaria obliqua subsp. *obliqua*	LM993097	LM992977	LM993207
Fritillaria obliqua subsp. *tuntasia* (1)	LM993098	LM992978	LM993208
Fritillaria obliqua subsp. *tuntasia* (2)	LM993099	LM992979	LM993209
Fritillaria olivieri	AY624449	LM992980	LM993210
Fritillaria oranensis	LM993100	LM992981	LM993211
Fritillaria orientalis (1)	LM993101	LM992982	LM993212
Fritillaria orientalis (2)	LM993102	LM992983	LM993213
Fritillaria pallidiflora	AY624450	LM992984	AY624398
Fritillaria persica	AY624451	LM992985	AY624399
Fritillaria pinardii	LM993103	LM992986	LM993214
Fritillaria pluriflora	LM993104	LM992987	LM993215
Fritillaria poluninii	LM993105	LM992988	LM993216
Fritillaria pontica	LM993106	LM992989	LM993217
Fritillaria pudica	LM993107	LM992990	LM993218
Fritillaria pyrenaica subsp. *pyrenaica*	LM993108	LM992991	LM993219
Fritillaria raddeana	AY624454	LM992992	AY624402
Fritillaria recurva	AY624455	LM992993	AY624403
Fritillaria reuteri	AY624456	LM992994	LM993220
Fritillaria rixii	LM993109	LM992995	LM993221
Fritillaria ruthenica	LM993110	LM992996	LM993222
Fritillaria sewerzowii (1)	AY624457	LM992997	AY624405
Fritillaria sewerzowii (2)	LM993111	LM992998	LM993223
Fritillaria sewerzowii (3)	LM993112	LM992999	LM993224
Fritillaria shikokiana (1)	LM993113	LM993000	LM993225
Fritillaria shikokiana (2)	LM993114	LM993001	LM993226
Fritillaria sibthorpiana subsp. *enginiana*	LM993115	LM993002	LM993227
Fritillaria sibthorpiana subsp. *sibthorpiana*	LM993116	n/a	LM993228
Fritillaria sororum (1)	LM993117	LM993003	LM993229
Fritillaria sororum (2)	LM993118	LM993004	LM993230
Fritillaria stenanthera	LM993119	LM993005	LM993231
Fritillaria stribrnyi	LM993120	LM993006	LM993232
Fritillaria theophrasti	LM993121	LM993007	LM993233
Fritillaria thessala subsp. *reiseri*	LM993122	LM993008	LM993234
Fritillaria thessala subsp. *thessala*	LM993123	LM993009	LM993235
Fritillaria thunbergii	LM993124	LM993010	LM993236
Fritillaria tokushimensis	LM993125	LM993011	LM993237
Fritillaria tortifolia	LM993126	LM993012	LM993238
Fritillaria tubiformis subsp. *tubiformis*	LM993127	LM993013	LM993239
Fritillaria usuriensis (1)	LM993128	LM993014	LM993240
Fritillaria usuriensis (2)	LM993129	LM993015	LM993241
Fritillaria uva-vulpis	LM993130	LM993016	LM993242
Fritillaria verticillata	LM993131	LM993017	LM993243
Fritillaria walujewii (1)	LM993132	n/a	LM993244
Fritillaria walujewii (2)	LM993133	LM993018	LM993245

Fritillaria whittallii	LM993134	LM993019	LM993246
Fritillaria zagrica	LM993135	LM993020	LM993247
Lilium davidii	AY624460	LM993025	AY624410
Lilium humboldtii subsp. *ocellatum*[b]	AY624461	LM993026	AY624412
Lilium lophophorum	AY624462	LM993027	LM993252
Lilium mackliniae	LM993140	LM993028	LM993253
Lilium maritimum	AY624463	LM993029	AY624414
Lilium martagon	LM993141	LM993030	AY624415
Lilium pensylvanicum[c]	AY624466	LM993031	AY624419
Lilium pyrenaicum	LM993142	LM993032	LM993254
Lilium regale	AY624464	LM993033	AY624417
Lilium rubescens	LM993143	LM993034	AY624418
Lilium speciosum var. *speciosum*[d]	AY624467	LM993035	LM993255
Lilium superbum	AY624468	n/a	AY624421
Lilium washingtonianum subsp. *washingtonianum*	LM993144	LM993036	LM993256
Outgroups			
Calochortus tolmiei	LM993136	LM993021	LM993248
Cardiocrinum giganteum var. *giganteum*	LM993137	LM993022	LM993249
Clintonia borealis	LM993138	LM993023	LM993250
Gagea reticulata	LM993139	LM993024	LM993251
Nomocharis pardanthina	LM993145	LM993037	LM993257
Notholirion macrophyllum	LM993146	LM993038	LM993258
Notholirion thomsonianum	LM993147	LM993039	LM993259
Scoliopus bigelovii	LM993148	LM993040	LM993260
Tricyrtis maculata	LM993149	LM993041	LM993261
Tulipa heteropetala	LM993150	n/a	LM993262

[a] Sequences from this material (Kew DNA bank accession 2563) were previously published under the name *F. hermonis* subsp. *amana*, which is now considered to be a synonym and *F. amana* the accepted name.

[b] Sequences from this material (Kew DNA bank accession 3723) were previously published under the name *L. humboldtii* var. *bloomerianum*, which is now considered to be a synonym and *L. humboldtii* subsp. *ocellatum* the accepted name.

[c] Sequences from this material (Kew DNA bank accession 933) were previously published under the name *L. sachalinense*, which is now considered to be a synonym and *L. pensylvanicum* the accepted name.

[d] Sequences from this material (Kew DNA bank accession 3724) were previously published under the name *L. speciosum* var. *clivorum*, which is now considered to be a synonym and *L. speciosum* var.*speciosum* the accepted name.

Table S5 - Amplification results for screen of low-copy nuclear genes[a].

	Species						
Gene Region	*F. armena*	*F. dagana*	*F. davidii*	*F. eduardii*	*F. koidzumiana*	*F. stenanthera*	*Lilium mackliniae*
AKT1	X	X	X	X	X	X	X
AP	X	SB	SB	SB	SB	SB	SB
CHS	SM	X	SM	SM	SM	SM	SM
G3PDH	MB/SM	MB/SM	MB/SM	MB/SM	MB/SM	MB/SM	MB/SM
GS	SM	X	SM	MB	SM	MB	MB/SM
LFY/FLO (intron 1)	X	X	SM	SM	SM	SM	MB/SM
LFY/FLO (intron 2)	X	X	X	X	X	X	X
MS[b]	MB	X	MB	MB	MB	MB	X
PRK	SM	X	SM	SM	SM	MB	MB
RPB2	SM	X	MB	SM	SM	SM	MB
XDH	SM	X	SM	MB	SM	SM	SM

[a] Key to amplification results: X – no amplification; SB – single band; MB – multiple bands; SM – smear.

[b] For *MS*, samples that amplified generated two bands

Appendix B

Supplementary Figures


M
N
O
P
Q
R
S
T
U
V


W
Lilium mackliniae Sealy
Accession Number - LRH-L019
Photograph taken 14MAY2009
Bulbs from Jacques Armand Int. UK
Laurence Hill © 2009

Figure S1 - Photographic vouchers for plant material used in DNA sequencing (see Table S1). A – *Fritillaria agrestis* LH 110 (subgenus *Liliorhiza*; photo: Laurence Hill); B – *F. agrestis* LH 110 (subgenus *Liliorhiza*; photo: Laurence Hill); C – *F. bithynica* Kew 1989-3037 (subgenus *Fritillaria*; photo: Laura Kelly); D – *F. camschatcensis* LH 955 (subgenus *Liliorhiza*; photo: Laurence Hill); E – *F. camschatcensis* LH 019 (subgenus *Liliorhiza*; photo: Laurence Hill); F – *F. dagana* Kew 2005-2036 (subgenus *Liliorhiza*; photo: Laura Kelly and Madeleine Berger); G – *F. involucrata* Kew 2000-3114 (subgenus *Fritillaria*; photo: Laura Kelly); H – *F. involucrata* Kew 2000-3114 (subgenus *Fritillaria*; photo: Laura Kelly); I – *F. involucrata* LH 076 (subgenus *Fritillaria*; photo: Laurence Hill); J – *F. kaiensis* LH 847 (subgenus *Japonica*; photo: Laurence Hill); K – *F. kaiensis* LH 847 (subgenus *Japonica*; photo: Laurence Hill); L – *F. liliacea* LH 305 (subgenus *Liliorhiza*; photo: Susan Bond); M – *F. meleagroides* Kew 2005-2045 (subgenus *Fritillaria*; photo: Laura Kelly); N – *F. oranensis* RRW 8837 (subgenus *Fritillaria*; photo: Robert Wallis); O – *F. ruthenica* Kew 2004-3479 (subgenus *Fritillaria*; photo: Laura Kelly); P – *F. ruthenica* Kew 2004-3479 (subgenus *Fritillaria*; photo: Laura Kelly); Q – *F. tokushimensis* s.n. (subgenus *Japonica*; photo: Takato Natsui); R – *F. walujewii* LH 185 (subgenus *Fritillaria*; photo: Laurence Hill); S – *F. walujewii* LH 185 (subgenus *Fritillaria*; photo: Laurence Hill); T – *F. tortifolia* s.n. (subgenus *Fritillaria*; photo: Laura Kelly); U – *F. tortifolia* s.n. (subgenus *Fritillaria*; photo: Laura Kelly); V – *F. verticillata* 2005-2049 (subgenus *Fritillaria*; photo: Laura Kelly); W – *Lilium mackliniae* LH L019 (photo: Laurence Hill).


Rix 2001
F. subgenus Fritillaria
F. subgenus Japonica
F. subgenus Rhinopetalum
F. subgenus Petilium
F. subgenus Korolkovia
F. subgenus Theresia
F. subgenus Davidii
F. subgenus Liliorhiza
Lilium
Outgroup
A
B
F. amana (2)
F. latakiensis
F. alfredae subsp. glaucoviridis
F. elwesii
F. graeca
F. pinardii
F. pontica
F. sororum (1)
F. sororum (2)
F. crassifolia subsp. crassifolia
F. fleischeriana
F. latifolia
F. whittallii
F. crassifolia subsp. kurdica (1)
F. crassifolia subsp. kurdica (2)
F. michailovskyi
F. minuta
F. alburyana
F. armena
F. poluninii
F. aurea
F. caucasica
F. acmopetala subsp. acmopetala
F. amana (1)
F. assyriaca subsp. assyriaca (1)
F. frankiorum
F. obliqua subsp. obliqua
F. rixii
F. obliqua subsp. tuntasia (1)
F. obliqua subsp. tuntasia (2)
F. thessala subsp. thessala
F. davisii
F. mutabilis
F. conica
F. forbesii
F. sibthorpiana subsp. engiriana
F. sibthorpiana subsp. sibthorpiana
F. bithynica
F. stribrnyi
F. assyriaca subsp. assyriaca (2)
F. zagrica
F. oliveri
F. uva-vulpis
F. carica
F. drenovskii
F. ehrhartii
F. gussichiae
F. kotschyana
F. reuteri
F. theophrasti
F. involucrata (1)
F. tubiformis subsp. tubiformis
F. involucrata (2)
F. montana (1)
F. montana (2)
F. rutlhenica
F. messanensis subsp. messanensis
F. thessala subsp. reiseri
F. orientalis (1)
F. orientalis (2)
F. usuriensis (1)
F. usuriensis (2)
F. lusitanica
F. meleagris subsp. meleagris
F. meleagroides
F. oranensis
F. pyrenaica subsp. pyrenaica
F. tokushimensis
F. muraiana (1)
F. muraiana (2)
F. kaiensis
F. koidzumiana
F. shikokiana (1)
F. shikokiana (2)
F. amabilis (2)
F. ayakoana
F. amabilis (1)
F. japonica
F. bucharica
F. gibbosa
F. karelinii
F. stenanthera
F. imperialis
F. raddeana
F. chitralensis
F. eduardii
F. sewerzowii (1)
F. sewerzowii (2)
F. sewerzowii (3)
F. tortifolia
F. verticillata
F. pallidiflora
F. walujewii (1)
F. walujewii (2)
F. thunbergii
F. cirrhosa
F. persica
F. davidii
Lilium humboldtii subsp. ocellatum
Lilium martinum
Lilium rubescens
Lilium washingtonianum subsp. washingtonianum
Lilium superbum
Lilium macklniae
Nomocharis pardanthina
Lilium lophophorum
Lilium regale
Lilium speciosum var. speciosum
Lilium davidii
Lilium martagon
Lilium pensylvanicum
Lilium pyrenaicum
F. eastwoodiae
F. recurva
F. micrantha
F. affinis (1)
F. affinis (2)
F. atropurpurea
F. agrestis
F. liliacea
F. pluriflora
F. falcata
F. pudica
F. camschatcensis (1)
F. camschatcensis (2)
F. camschatcensis (3)
F. camschatcensis (4)
F. camschatcensis (5)
F. maximowiczii (1)
F. maximowiczii (2)
F. dagana
Cardiocrinum giganteum var. giganteum
Notholirion macrophyllum
Notholirion thomsonianum
Gagea reticulata
Tulipa heteropetala
Clintonia borealis
Caiochortus tolmiei
Scoliopus bigelovii
Tricyrtis maculata
Europe, Middle East, North Africa, & China
Japan
Central Asia, China & Middle East
North America
North America & N. E. Asia
N. E. Asia & China

Figure S2 - Strict consensus of 1986 most parsimonious trees (Length = 1313; CI = 0.717; RI = 0.868) from the parsimony ratchet analysis of the combined plastid dataset. Support values for nodes (bootstrap percentage (BP) of ≥70/posterior probabilities (PP) of ≥0.95) are shown above branches; a dash indicates a node with <70% BP/<0.95PP, or both. Filled coloured circles indicate different taxonomic groups (*Fritillaria* subgenera, *sensu* Rix, 2001, *Lilium* and outgroup; see top left-hand corner for key); bold uppercase letters (A and B) mark the two clades recovered for subgenus *Fritillaria*. Names of species used in traditional Chinese medicine are underlined. Symbols following taxon names indicate individuals from species resolved as non-monophyletic. Alternating black and white bars denote broad geographic distributions for species of *Fritillaria*, but note that the colours do not indicate groups with similar distributions.

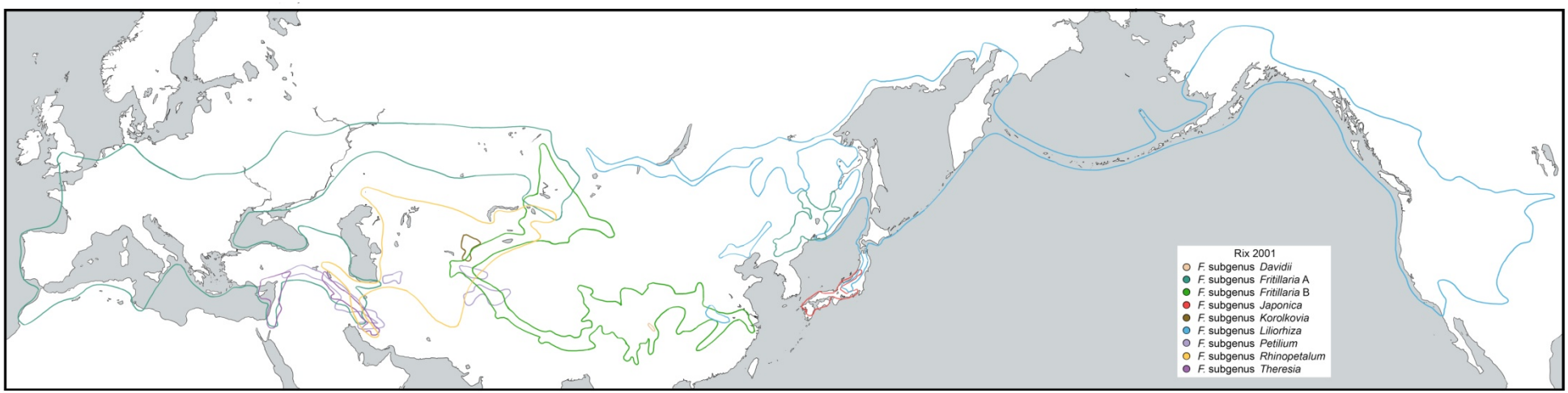

Figure S3 - The global distribution of *Fritillaria*. A different line colour is used for each of the subgenera (see bottom right for key); the two subgenus *Fritillaria* clades (A and B) are shown separately. Distributions are based on location data compiled from a comprehensive literature review (including journal articles, floras and monographs), from data extracted from specimens held in over 60 herbaria, from the field notes of LH and from personal communications to LH from numerous biologists and horticulturalists with knowledge of natural *Fritillaria* populations. Information from all species of *Fritillaria* was included for the purposes of drawing the distribution map, rather than from just those species sampled in the current study. For *F.* subgenus *Fritillaria*, species belonging to series 7-11 (*sensu* Rix, 2001) were included within the clade B distribution as all members of these series sampled to date (excepting *F. usuriensis*) have been shown to belong to *F.* subgenus *Fritillaria* clade B. Species from all other *Fritillaria* series, plus *F.usuriensis*, were included within the Clade A distribution.