Genetic clues to the origin of the apple

Stephen A. Harris, Julian P. Robinson and Barrie E. Juniper

Molecular genetic markers complement archaeological, breeding and geographical investigations of the origins, history and domestication of plants. With increasing access to wild apples from Central Asia, along with the use of molecular genetic markers capable of distinguishing between species, and explicit methods of phylogenetic reconstruction, it is now possible to test hypotheses about the origin of the domesticated apple. Analyses of nuclear rDNA and chloroplast DNA (cpDNA) sequences indicate that the domesticated apple is most closely related to series Malus species. Moreover, the occurrence of a shared 18-bp duplication in the cpDNAs of wild and cultivated apple supports the close relationship between them. Hypotheses about the hybridization and the origin of the domesticated apple cannot be rejected completely until more variable, phylogenetically informative markers are found.

The domesticated apple is one of the most important fruit crops of the colder and temperate parts of the world [1]. Vavilov suggested that the wild apple of Turkestan (Kazakhstan, Kyrgyzstan, Uzbekistan, Turkmenistan and Tajikistan) and its close relatives were the progenitors of the domesticated apple [2]; the whole process of wild apple domestication being traced to Almaty (Kazakhstan). Vavilov reasoned that, because the wild apple bears similar fruits to the domesticated apple, it must have been the progenitor. Recent fieldwork in the region appears to confirm the similarity between wild and cultivated apples [2,3]. Furthermore, Janick et al. suggested that ‘this area (Central Asia) is the area of greatest diversity and the centre of origin’ of the domesticated apple [4]. This is linked to Vavilov’s ‘oversimplified’ idea that the centre of diversity is the place of origin [5,6]. The Central Asian wild apple (Box 1) is a diverse species with a wide range of forms, colours and flavours [7], whilst its allelozyme diversity is significantly greater than that found in four widely distributed North American wild apples [8,9].

The Central Asian wild apple is closely related to a group of apples that have different-sized fruits. One of these, Malus baccata (Siberian Crab), has small, red fruits that hang in clusters and are bird-dispersed (Fig. 1). Malus baccata might have had a wider distribution than that of the present day, and we think that populations became ‘trapped’ as the Tien Shan began to rise out of the Tethys Ocean. Over seven million years, perhaps up to ten million years, mammals, such as bears, acted as distribution vehicles by selecting the largest and juiciest fruits; a small, bird-distributed, cherry-like delicacy giving way to a large mammal-distributed form. Small apples have even been observed to pass intact through a bear’s jaw and guts; it should be noted, however, that apple seeds retained in the apple core do not germinate.

By the time humans began to occupy the area –5000–8000 years ago, the early evolution of the apple was almost complete, and its migration, ably assisted by the now domesticated horse [10], was under way. Over several more thousands of years, from within this migrating flow came the many thousands of apple cultivars now known, as a result of both unconscious and conscious selection [11]. Based on combined archaeological and molecular data, it seems likely that, in the late Neolithic or early Bronze Age, travellers on the great trade routes that ran from central China to the Danube (Fig. 2), carried the seed of the Central

Box 1. Taxonomic hierarchy and apple classification

Species are arranged according to a taxonomic hierarchy, in which they are placed into larger groups or divided into smaller groups to account for observed variation. Each of these groups is given a particular name. Thus, the family Rosaceae includes the genus Malus (apples) with its ~55 species, which are grouped into infrageneric groups (section, series), and each species can be divided into intraspecific groups (cultivar). Malus spp. are arranged according to Phipps et al.’s classification [a].

The genus Malus comprises some 55 species [a], although between eight and 79 species have been recognized. Part of the problem of Malus taxonomy is the intimate association that humans have with apples, where the distinction between wild and cultivated species could become blurred and hence the recognition of distinct categories difficult. Furthermore, the scientific names that have been applied to the domesticated apple are legion, and include Malus pumila Miller, M. communis Desf., M. sylvestris (L.) Miller and M. domestica Borkh. Recently, it has been argued that M. pumila is the correct name for the domesticated apple and that this also includes the presumed wild relative, M. sieversii (Ledeb.) M. Roem [b]. We use M. domestica to refer to ‘domesticated apple’ and M. sieversii to refer to the Central Asian ‘wild apple’. However, the term ‘wild apple’ has been applied to all Malus spp. other than the domesticated apple. For example, in Europe, the common name ‘wild apple’ refers to M. sylvestris Miller, whilst in North America the term has been applied to species including M. angustifolia (Ait.) Michx. and M. coronaria (L.) Miller. The situation is further complicated by the application of the name ‘wild apple’ to hybrids between domesticated apple and other members of the genus Malus.

The taxonomic hierarchy for the domesticated apple ‘Bramley’s Seedling’ would be:

- **Family:** Rosaceae
- **Subfamily:** Maloideae
- **Genus:** Malus
- **Section:** Malus
- **Series:** Malus
- **Species:** domestica
- **Variety/cultivar:** ‘Bramley’s seedling’

References

- b Maberley, D.J. et al. (2001) The name of the apple. Telopea 9, 421–430

Fig. 1. Range of variation in apple fruit types: (a) Malus baccata, a bird-dispersed species; and (b) domesticated apple, a mammal-dispersed species.

---

Asian wild apple west, either in saddle bags or horses’ guts. We now know that the important technique of grafting (described as ‘instant domestication’ [12]), where the variety can be preserved forever, was probably discovered in Mesopotamia at Mari, as early as 3800 years ago (S. Dalley, pers. commun.). The description, on cuneiform tablets, concerns vines (Vitis spp.), but the techniques are easily transferable. From there, the fruits and the necessary technology passed through the Persians and Greeks to the Romans, who perfected orchard economies. The Romans brought the whole package to Western Europe and, for the last 2000 years, the domesticated apple has diversified and flourished worldwide. Archaeological evidence for the collection of apples from the wild in Europe can be found in Neolithic (11 200 years ago) and Bronze Age (c. 4500 years ago) sites throughout Europe [13, 14], and for apple cultivation as early as 1000 BC in Israel [1].

In terms of apple domestication, series Malus is the most important group of species. However, the nomenclature of the species in series Malus is complex. With few discrete characters to differentiate species, the difficulty of species delimitation has hampered investigations into the origins of apples [15]; morphological characters used to delimit species in series Malus are continuous and overlapping. Such problems perhaps contributed to Zohary and Hopf’s statement that ‘it is therefore futile to try to delimit the area of initial domestication on the basis of the evidence available from the living plant’ [1]. However, the molecular systematics revolution potentially provides an excellent source of potential characters for analyses of these type.

The origin of the domesticated apple and relations with other Malus spp.

Combinations of nuclear (nDNA) and cytoplasmically (chloroplast (cpDNA) and mitochondrion (mtDNA)) inherited molecular genetic markers provide important clues about the origins of domesticated plants, for example, Citrus spp. [16], beans [17] and potatoes [18]. In Malus, the chloroplast genome provides data about evolutionary relationships of the maternal line, whilst biparentally inherited nDNA provides independent data, which, combined with cpDNA, could enable the origin of the domesticated apple to be determined. Hybridization and introgression are expected to have been important factors in the development of domesticated apple cultivars. Therefore, for investigations that aim to determine the earliest origins of the domesticated apple, it is necessary to use cultivars that are as ancient as possible.

matK, a cpDNA-encoded region ~1800-bp long, of which 1341 bp was sequenced, showed only 16 phylogenetically informative characters across the genus Malus, and thus poor resolution in the phylogenetic tree [19] (Fig. 3). However, two duplications were found 39 bp from the 3’ end of the matK coding region. Duplication I is an imperfect 8-bp duplication that differs by a T residue, whereas duplication II is a perfect 18-bp duplication. Duplications I and II were always present together, suggesting that each has arisen only once during the evolution of Malus spp. (Fig. 3). Duplication II was polymorphic in both the Central Asian wild apple and the domesticated apple, suggesting that the Central Asian wild apple could be the major maternal contributor, as proposed by Watkins et al. [20].

The nuclear ribosomal internal transcribed spacer (ITS) showed 89 phylogenetically informative characters from the 617 bp sequenced. A strongly supported group comprised the Central Asian wild apple and the domesticated apple, as well as M. asiatica, M. orientalis, M. niedzwetzkyana and M. prunifolia (all from section Malus spp.), although there was little resolution within this group (Fig. 4). Together, data from the chloroplast and nuclear genomes support the view that the domesticated apple is most closely related to the series Malus spp. These data also indicate that the Central Asian wild apple could be most closely related to the domesticated apple. Savolainen et al. were unable to determine the origin of the domesticated apple because of a lack of variation in the cpDNA-encoded atpB-rbcL spacer and the absence of the Central Asian wild apple in their sample [21]. However, additional phylogenetically useful DNA sequence information should be sought, so that this hypothesis can be confirmed; additional sampling of Central Asian wild apples and domesticated apples is needed to ensure that rare hybridization events with other Malus spp. that might have
reconstruction of phylogeny because of RAPD data are not appropriate for the [23]. However, it has been argued that accessions from 23 species analysed determining the relationships among the [et al.]

Fig. 3. Phylogeny of the chloroplast DNA-encoded gene matK in the genus Malus (a). A cartoon of matK is shown in (b) and the positions of the two duplications are indicated. Wild apple refers to the Central Asian wild apple, M. sieversii. Circles indicate the level of statistical support for particular groups. The distribution of the two matK duplications is shown on the tree.

Two contrasting types of molecular genetic marker can be used to characterize apples. The first type is those that identify few, high information content loci, such as microsatellites. The second are those that identify many, low information content loci, such as RAPD. Studies of RAPD variation in Malus spp. have focused on the diversity present in domesticated apples and germplasm collections. Dunemann et al. investigated RAPD variation at 52 putative loci in a collection of 27 domesticated apple cultivars and were able to differentiate between cultivars [22]. Furthermore, RAPD markers have also been used to analyse the maternal and paternal contributions to pedigrees [22,28,29]. In a large study by Oraguzie et al., 43 RAPD markers were used to assess the variation in a worldwide collection of 50 old domesticated apple varieties, current breeding lines and a collection of 105 other apples, including wild species [30].

Variation within the domesticated apple

Many thousands of named domestic apple varieties (both dessert and cider) have been selected for hundreds of years in Europe, Asia and North America, and more recently in the Southern Hemisphere. Together with wild species, these are maintained in national collections as genetic resources for breeding, particularly as sources of resistance to apple scab [Venturia inaequalis (Cooke) Winter], powdery mildew [Podosphaera leucotricha (Ellis and Everh) Salmon] and fireblight [Erwinia amylovora (Burill) Winslow et al.] [26]. As with wild species, the taxonomy of apple varieties is fraught with problems, including environmental and developmental variation within varieties, and limited numbers of morphological and chemical characters to distinguish varieties [11]. In addition the triad of decreasing budgets, but increasing collection size and running cost, traps a collection manager whose goal is to have a well-characterized but easily utilized collection, and forces choices to be made to try and minimize genotypic redundancy. Thus, molecular genetic markers provide a valuable aid to the identification of duplicates (possible cultivar synonyms), correctly naming mis-identified plants and enabling managers to make effective decisions about the reduction of collection size without reducing genetic variation (i.e. creation of core collections) [27].

The need for additional DNA markers has led to analyses of Malus phylogeny using randomly amplified polymorphic DNA (RAPD) [22] and nuclear microsatellites [single sequence repeats (SSR)] [23]. Dunemann et al. ’s results indicated that M. pumila and M. sylvestris (section Malus) were involved in the origin of cultivated apples [23]. However, they suggested that M. sylvestris, M. floribunda or M. dasypyla could be the female parent of M. domestica. By contrast, Hokanson et al. ’s SSR studies were not useful in determining the relationships among the 142 accessions from 23 species analysed [23]. However, it has been argued that RAPD data are not appropriate for the reconstruction of phylogeny because of reproducibility, primer structure, dominance, product competition, homology, allelic variation, genome sampling and non-independence of loci, whilst similar problems have been highlighted for the use of SSRs in phylogenetic analyses [23–25].

http://tig.trends.com
Eight apple SSR loci (seven dinucleotide repeat loci and one trinucleotide repeat locus) were developed by Szewc-McFadden and colleagues [33,34] and used by Hokanson [33] to identify the genotypes of a core collection of 66 domesticated apple varieties. All but seven pairs of accessions (five of these were either mutations or their progenitors) of the 2145 possible pair-wise varietal combinations could be distinguished. The probability of any two single locus genotypes matching by chance in this study ranged from 0.027 to 0.970, although this dropped to $0.156 \times 10^{-8}$ for multilocus comparisons. The high discrimination power shows the value of SSR markers for varietal genotyping, whilst their co-dominance and high repeatability makes them more reliable markers than RAPDs. SSR loci also appear to be very useful anchor markers in apple genome-mapping projects. The occurrence of low frequency alleles potentially makes SSRs valuable markers for pedigree analysis, although large numbers of co-dominant markers are likely to be necessary if complex crossing patterns are to be disentangled.

Limited surveys of apple varieties have revealed two cpDNA mutations that together provide evidence for multiple origins of the maternal component of the domesticated apple, because apple cultivars are propagated vegetatively. In a survey of 40 cultivars (36 named and four unnamed), Savolainen et al. showed that cultivars could be divided into two cpDNA haplotype groups depending on a C→T transition at position 17 of the atpB-rbcL spacer [21]. Similarly, Robinson et al., in a survey of nine cultivars, showed that these could be divided into two groups based on the occurrence of an 18-bp duplication in the matk-3′trnK spacer region [19]. The duplication was found only in the domesticated apple and one accession of the Central Asian wild apple following a broad survey of the genus, highlighting the close relationship between the Central Asian wild and domesticated apple. Unfortunately, because the objectives of these two studies were different, complementary cultivars were not included. However, combining these two markers into the same study might allow major lineages in cultivated and wild apples to be identified.

Investigations of restriction fragment length polymorphism provide additional evidence for uncharacterized variation in both the chloroplast and mitochondrial genomes of domesticated apples [34,35].

**Conclusions**

Two stages appear to have been important in the domestication of apples across the range of the genus; the initial introduction of apples into Western Europe and later hybridizations between cultivars and between cultivars and wild species. Based on molecular data, it was members of section M. alus that were important, particularly the wild apple, in the initial introduction. The morphological, biochemical and molecular variation within wild apple indicates that the earliest selections of domesticated apples could have come directly from the wild apple, without the involvement of other species. However, later hybridizations could have been important in the creation of new cultivars that carry economically important characteristics. The detailed analysis of these stages requires more sampling of the variation within Central Asia. Given the extensive range of the wild apple, detailed analysis of known old apple cultivars using highly variable molecular genetic markers such as SSRs, identification of DNA sequences that allow resolution of the section M. alus spp. a comprehensive taxonomic account of the genus M. alus and additional archaeological evidence of human use of apples in Central Asia and the Near East will be important. Clearly, there is still much to be discovered about the origin of the domesticated apple, the processes that led to its domestication and the origins of both dessert and cider apples.

**Acknowledgements**

We thank The Leverhulme Foundation for funding the work described in this article; the Merlin Trust for travel grants; the staff at the USDA-ARS Plant Genetic Resources Unit, Cornell University, in particular P. Forsline, H. S. Aldwinkle, A. Szewc-McFadden and W. Lamboy, for practical help and support; K. Tobutt and the staff at Horticultural Research International, East Malling, for providing wild species material from their collections; I. Belolipov and L. Samieva of...
the University of Tashkent for assistance in the field collecting in Uzbekistan and Kirghistan; the staff of LES-I-C, the Kirghiz Swiss Forestry Support Programme and the Forestry Institute of the Republic of Kirghistan, J alal-Abad, Kirghistan; N. Ajtkhozina and N. Ajtkhozina of the Ministry of Sciences and the University of Almaty for assistance in field collecting in Kazakhstan. In particular, we thank R. Wise of the Dept of Plant Sciences, Oxford, UK for all the artwork. B.E.J. is currently a Leverhulme Emeritus Research Fellow in the Dept of Plant Sciences, Oxford.

References
15 Rohrer, J.R., et al. (1994) Floral morphology of Malus baccata (Rosaceae) and its systematic relevance. J. Bot. 81, 574–581
22 Dunemann, F., et al. (1994) Genetic relationships in Malus evaluated by RAPD fingerprinting of cultivars and wild species. Plant Breed. 113, 150–159

Stephen A. Harris
Barrie E. Juniper*
Dept of Plant Sciences, University of Oxford, South Parks Road, Oxford, UK OX1 3RB
*e-mail: barrie.juniper@plants.ox.ac.uk

J ulian R Robinson
Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford, UK OX1 3RE.

TRENDS in Genetics Vol.18 No.8 August 2002