Phylogenetic Analysis of 47 Chloroplast Genomes Clarifies the Contribution of Wild Species to the Domesticated Apple Maternal Line

Svetlana V. Nikiforova,[1] Duccio Cavalieri,[1] Riccardo Velasco,[1] and Vadim Goremykin*[,1]
[1]Research and Innovation Centre, Fondazione E. Mach, San Michele all'Adige (TN), Italy
***Corresponding author:** E-mail: Vadim.Goremykin@fmach.it.
Associate editor: Shizhong Xu

Abstract

Both the origin of domesticated apple and the overall phylogeny of the genus *Malus* are still not completely resolved. Having this as a target, we built a 134,553-position-long alignment including two previously published chloroplast DNAs (cpDNAs) and 45 de novo sequenced, fully colinear chloroplast genomes from cultivated apple varieties and wild apple species. The data produced are free from compositional heterogeneity and from substitutional saturation, which can adversely affect phylogeny reconstruction. Phylogenetic analyses based on this alignment recovered a branch, having the maximum bootstrap support, subtending a large group of the cultivated apple sorts together with all analyzed European wild apple (*Malus sylvestris*) accessions. One apple cultivar was embedded in a monophylum comprising wild *M. sieversii* accessions and other Asian apple species. The data demonstrate that *M. sylvestris* has contributed chloroplast genome to a substantial fraction of domesticated apple varieties, supporting the conclusion that different wild species should have contributed the organelle and nuclear genomes to the domesticated apple.

***Key words: Malus domestica*, chloroplast genome phylogeny, base compositional heterogeneity, hybridization.**

Introduction

The domesticated apple, *Malus domestica* (MD) Borkh., is one of the most important temperate fruit crops. The origin of the crop from wild progenitors is, for several reasons, relevant both to the breeders and to taxonomists. Yet, fruit tree domestication is still poorly understood process (Miller and Gross 2011), which differs from domestication of the selfing perennial crops in many aspects. Long juvenile development and self-incompatibility of fruit trees lead to highly diverse offspring, rendering breeding long, expensive, and laborious. Weak domestication syndrome (Pickersgill 2007; Velasco et al. 2010) and limited domestication bottleneck (Miller and Gross 2011; Cornille et al. 2012) are likely consequences of a limited number of tree generations, which underwent artificial selection.

Invention of grafting practice, undoubtedly, revolutionized tree breeding in general, and apple breeding, in particular, as it allowed maintaining desirable lines indefinitely by vegetative propagation. When and where the grafting was invented is not known; however, evidence of widespread cultivation of apples in Europe can be traced back to antiquity, when grafting has become already well-established practice (Zohary and Hopf 1994). Before that time, the only apple "sorts" men could grow were variable "crab" apples resulting from uncontrolled open pollination (Zohary and Hopf 1994). It has even been suggested that spreading of the grafting technique, termed "instant domestication" (Zohary and Spiegelroy 1975), and not diffusion of cultivars, has led to apple domestication (Hokanson et al. 2001; Robinson et al. 2001). Planting apple trees from forest to gardens using root suckers, which was a widespread practice in central Asia (Ponomarenko 1983), might have contributed to the "instant domestication" scenario too. In this scenario, apple individuals to be propagated were chosen from local forests (Hokanson et al. 2001; Robinson et al. 2001). This eventual scenario, if true, could be evidenced, for example, by appearance of multiple unrelated branches subtending cultivars in a phylogenetic tree (Robinson et al. 2001).

At such circumstances, parental contribution in the origin of apple might be quite diverse. Nuclear DNA, which is inherited biparentally, and chloroplast DNA, which in Rosaceae, is inherited by maternal line (Hu et al. 2008), represent independent sources of evidence of evolution in the case of outcrossing, closely related *Malus* species, which can give fertile progeny (Harris and Ingram 1991). Here, eventual incongruence in nuclear and chloroplast DNA-based tree topologies would indicate strong influence of cross-pollination. In contrast, clonal propagation would lead to congruency of the nuclear and chloroplast DNA-based trees. Obviously, such comparison can yield meaningful results only if the trees themselves are stable and allow unequivocal explanation.

Attaining high resolution of the phylogenetic relationships within the genus *Malus* is, however, still problematic (e.g., Luby 2003; Li et al. 2012). For example, high dependence of the outcome of the simple sequence repeat (SSR) analyses within the genus *Malus* on the marker set can be seen in Zhang et al. (2012): Profound changes in the phylogenetic tree topology were observed when two sets of the SSR markers were applied to the same *Malus* accessions.

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1... on 19 April 2021
Article
Fast Track

© The Author 2013. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

In general, improvement of the phylogenetic methodology so far offered partial improvements in tree resolution within the genus. Nuclear ribosomal ITS, used in phylogenetic reconstruction of the Rosaceae (Campbell et al. 1995, 1997; Oh and Potter 2003; Lo et al. 2007), for the genus *Malus* yielded trees with many unresolved branches and low bootstrap support (Feng et al. 2007; Li et al. 2012). Attempts to combine internal transcribed spacer (ITS) data with the chloroplast *mat*K gene sequences (Robinson et al. 2001; Harris et al. 2002; Juniper 2007) did not improve the situation, as *mat*K gene contains only 16 informative characters across the genus *Malus*. The chloroplast *atp*B-*rbc*L spacer, also widely used as a phylogenetic marker in Rosaceae (Wissemann and Ritz 2005; Campbell et al. 2007; Lo and Donoghue 2012), contains only five polymorphic sites in *Malus* (Savolainen et al. 1995).

Recently, Velasco et al. (2010) and Micheletti et al. (2011) sequenced and analyzed the largest data set utilized so far to elucidate intergeneric phylogeny of *Malus*. The authors maintained that cumulative evidence (Neighbor-net from *p* distances plus maximum likelihood [ML] analyses of a subset of species) from these analyses points to the common ancestry of MD and *M. sieversii*. However, a certain degree of affinity between MD and *M. sylvestris* was also suggested based on single markers utilized by Velasco et al. (Harrison N and Harrison RJ 2011) and distribution of cpDNA polymorphisms (Coart et al. 2006). In the latter study, assignment of polymorphic polymerase chain reaction (PCR) products amplified from total DNA to chloroplast "haplotypes" was done without considering the influence (Arthofer et al. 2010) of sequences of the chloroplast origin residing in nuclear and mitochondrial genomes, which might have caused not genome-specific amplification.

Introgression of *M. sylvestris* DNA into the nuclear genome of MD was cited (Harrison N and Harrison RJ 2011) as a factor obscuring the results of phylogenetic inference based on the concatenation of sequences amplified from various nuclear genome regions (Velasco et al. 2010). Recent SSR analysis (Cornille et al. 2012) indicated that contribution of European crab apple, *M. sylvestris* into MD gene pool was at about 61%. This was explained by introgression of genetic material from *M. sylvestris* into the nuclear genome of domesticated apple, originated from *M. sieversii*.

However, evidence of wild species introgression is of complex interpretation, considering that, although nuclear genome is inherited biparentally, chloroplast and mitochondrial genomes are maternally transmitted. Given the situation, phylogenetic relationships among closely related plant species, particularly of those of economic interest that underwent multiple cycles of conventional breeding, should be investigated independently for the different cell genomes. The target of this article was to investigate the phylogenetic relationships of wild and domesticated apples based on chloroplast genome data.

Monoparental mode of cpDNA transmission (Hu et al. 2008) offers advantages for phylogeny reconstruction: When cpDNA is inherited uniparentally, exchanges between the genomes of different individuals are rare, and chloroplast fusion is even more rare (Gillham et al. 1991; Kuroiwa 1991). A nearly perfect cpDNA colinearity among even unrelated angiosperm species (e.g., Goremykin et al. 2003) also speaks in favor of rarity of recombination in cpDNA. Thus, introgression of sequence material from different species into the chloroplast genome molecule cannot obscure the inference of chloroplast genome-based phylogeny.

Complete chloroplast genomes have already been successively used in systematic studies at the shallow taxonomic level in seed plants (Bortiri et al. 2008; Parks et al. 2009). The level of statistical support for the branches observed was very high (Bortiri et al. 2008; Parks et al. 2009). We analyze in this article a data set including 46 completely or nearly completely sequenced chloroplast genomes sampled across the genus *Malus*, with emphasis on the sampling within the *domestica-sylvestris-siversii* lineage. Phylogenetic analyses of the chloroplast genome data have resulted in a tree topology characterized by a resolution previously unattained within the genus *Malus*.

Results

Overall Data Properties

The dot plot of the evolutionary versus observed distances among the OTUs based on the 134,553 position long alignment of 47 chloroplast genomes (fig. 1) showed a nearly perfect linear distribution. The mean ML distance among all the operation taxonomic units (OTUs), estimated using the settings of the best-fitting TVM + I + G model in PAUP*, was 0.00134, which is only marginally different from the corresponding uncorrected *p* distance (0.00128). Thus, superimposed substitutions, causing on deeper taxonomic levels model-misspecification and related tree-building artifacts in phylogenetic analyses based on cpDNA data (Zhong et al. 2011; Goremykin et al. 2013), should not pose a problem in the current analysis. The 5% χ^2 test, implemented in Tree-Puzzle program, was adopted to determine whether the base composition of sequences in the alignment was uniform. All accessions, except *M. mandjurica*, passed the test. Overall results of Bowker's test of matched pairs symmetry, as implemented in SeqVis program, indicated that, out of 1,081 pairwise comparisons, only 50 (~4.6%) showed significant compositional heterogeneity (P value < 0.05). Thus, the null hypothesis of evolution under stationary, reversible, and homogeneous conditions could not be rejected for the majority of the sequences under analysis.

The 134,553-position-long alignment of cpDNA sequences from *Pyrus* and 46 *Malus* species and cultivars contains 773 informative positions (in the sense of Maximum Parsimony). Of all informative positions, only three had three character states, the rest contained two character states. The data structure indicates no erosion of the historical signal in the cpDNA sequences under analysis. Good resolution of the overall tree topology (fig. 2) can thus be attributed to the fact that phylogenetic signal is well preserved in the data and is not distorted by multiple substitutions and strong compositional bias.

At the same time, unresolved clusters with zero or nearly zero branch lengths at the crown part of the tree (fig. 2) point

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

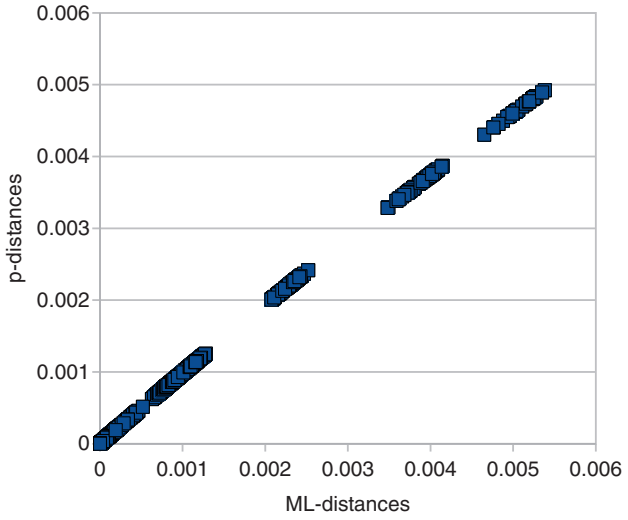

FIG. 1. Plot showing the distribution of the uncorrected *p* distances versus ML distances estimated using the settings of the best-fitting TVM + I + G model. The distances were calculated based on the 134,553-position-long alignment of chloroplast genomes, including *Pyrus*.

at the resolution limit that chloroplast genome sequences have at the shallowest taxonomic range. For example, our data provided no resolution for B2 branch, containing related sorts with known maternal pedigree (*domestica* cv. Gala and MD cv. Florina diverged from the common ancestor of the maternal line, Red delicious, which originated 143 years ago). Based on the observation that chloroplast genomes contain no informative characters to distinguish pedigree of apple cultivars in the six-species monophyletic cluster, including cultivars Gala and Florina (fig. 2), one can conclude that cpDNA data might be of limited use for intraspecific, population-based studies of plant biodiversity.

Tree Structure

Apple species *M. angustifolia* and *M. ioensis* of the Malus section Chloromeles, as defined in the Germplasm Resources Information Network, which we use as taxonomic reference, form the most basal clade on the *Malus* subtree (Branch E on fig. 2). This placement receives maximum bootstrap proportion (BP) support. Thus, among the species tested, *M. angustifolia* and *M. ioensis* can be considered the ancestral lineage of *Malus*.

Next representatives of the section Sorbomalus (*M. kansuensis*, *M. honanensis*, *M. prattii*, and *M. yunnanensis*)

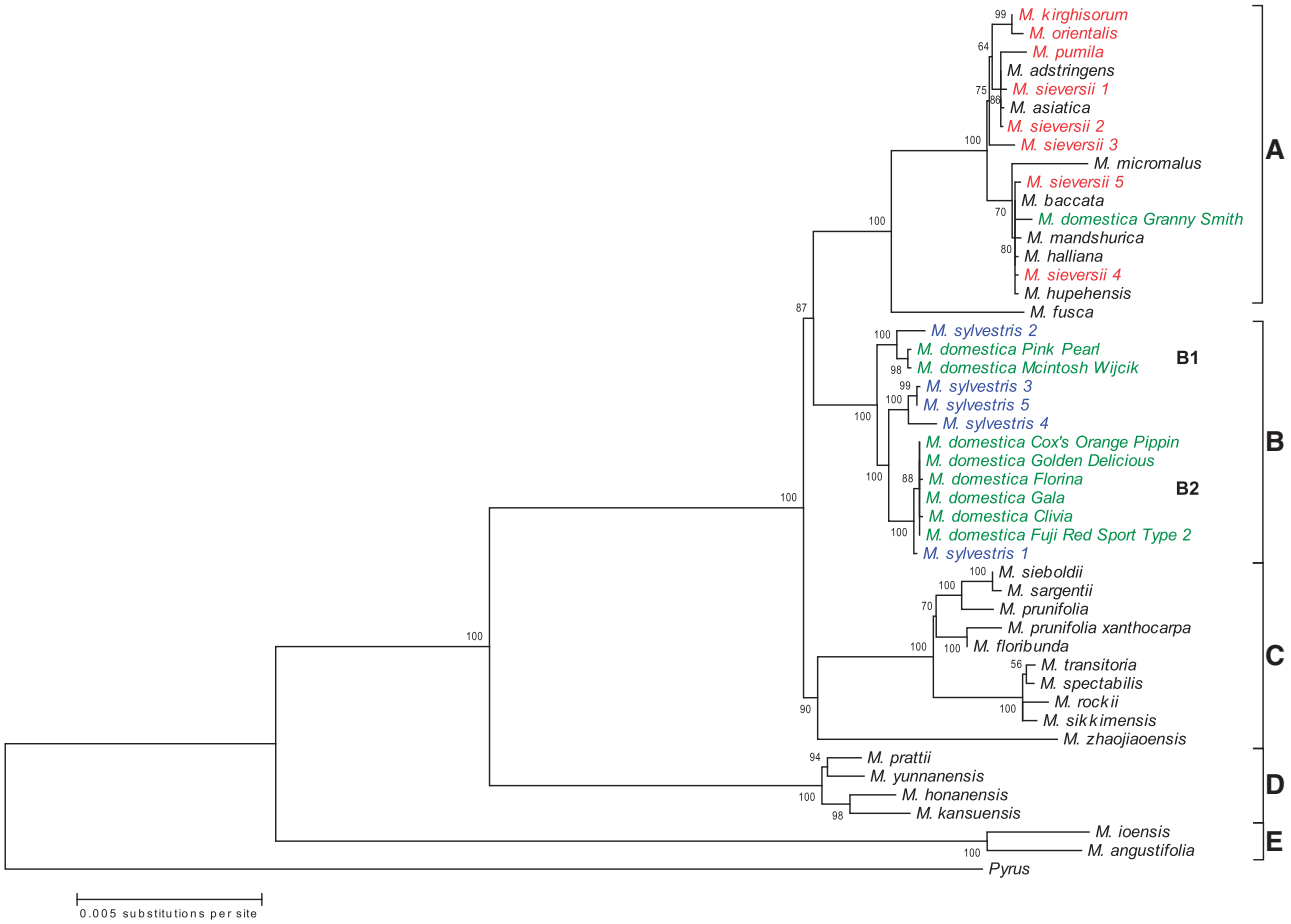

FIG. 2. Tree reconstructed from ML analysis using the settings of the optimal substitution model (TVM + I + G model) found by double-fitting procedure (Goremykin et al. 2010) for the 134,553-position-long alignment of chloroplast genomes. The numbers next to the tree branches represent bootstrap support values.

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

are branching off (Branch D on fig. 2, 100% BP support). Affinity of *M. fusca* to *M. kansuensis* (which are sometimes recognized within series Kansuensis [Robinson et al. 2001]) was not confirmed in our analysis: *M. fusca* cpDNA line has a sister group relationship in all bootstrap replicas made, to the broad assemblage, uniting *M. sieversii* and related species. Further up in the tree, resemblance between tree topology and taxonomy of the genus was lost. The next strongly supported (90% BP) branch C unites species attributed to: section Malus (*M. floribunda*, *M. prunifolia*, *M. spectabilis*, *M. xanthocarpa*, and *M. zhaojiaoensis*); section Sorbomalus (*M. sargentii*, *M. sieboldii*, and *M. transitoria*); and section Gymnomeles (*M. rockii* and *M. sikkimensis*). A further, strongly supported (100% BP) large branch uniting a number of wild species (Branch A), contains representatives of section Malus (*M. asiatica*, *M. sieversii*, *M. kirghisorum*, *M. orientalis*, and *M. pumila*), section Sorbomalus (*M. mandshurica*) and section Gymnomeles (*M. baccata*, *M. halliana*, *M. hupehensis*, and two hybrids with the chloroplasts deriving from different accessions of *M. baccata*–*M. adstringens* and *M. micromalus*). A conclusion based on these results is that the overall taxonomic subdivision of the genus *Malus* does not correspond to the phylogeny of the maternal line of the species analyses.

Accessions of *M. sieversii*, a central Asian species, whose nuclear genome was suggested to be the ancestor of domesticated apple, were scattered across branches containing other wild species. The clade supported by 80% BP, subtending *M. sieversii*, 4 and 5, included also *M. baccata*, *M. mandshurica*, *M. halliana*, *M. hupihensis*, and MD cv. Granny Smith, is clearly separated from a second well-supported (86% BP) branch subtending, among other OTUs, *M. sieversii* 1 and 2. These data indicate that genetic diversity of chloroplast genomes within *M. sieversii* exceeds that between other species and might justify its splitting onto at least two species.

Eight of nine *Malus x domestica* cultivars analyzed formed a branch with accessions of European crab apple, *M. sylvestris*, which was recovered in all 100 bootstrap replicas made (Branch B on fig. 2). Within this large branch, *M. x domestica* chloroplasts have polyphyletic origin, evidenced by two strongly supported monophyla, comprising *M. x domestica* accessions only, each sharing a strongly supported sister group relationship with different accessions of *M. sylvestris*. Polyphyly of *M. x domestica* maternal line is further evidenced by MD cv. Granny Smith embedded within a strongly supported (80% BP) branch with five Asian species including *M. sieversii*.

Dating Results

To estimate when separation of three cpDNA lines of MD occurred, we conducted two experiments. In the first, the age of the diversification of *Malus* from *Pyrus* was assumed to be about 45 My; in the second, the age of *Malus* was constrained with the earliest possible date based on molecular dating experiments (about 20 My, Lo and Donoghue 2012). The results of our dating experiments are presented in figure 3. The separation of the cpDNA line shared by Asian species and *Malus x domestica* cv. Granny Smith (fig. 2, Branch A) from the *M. x domestica*/*M. sylvestris* lineage (Branch B) occurred somewhere between 17.54 and 10.76 Ma. Within the lineage including *M. x domestica* and *M. sylvestris*, the separation between the cpDNA line shared by Pink Pearl and McIntosh Wijcik from the cpDNA line of other apple cultivars occurred within the 8.11–5.46 Ma range. Separation between the wild *M. sieversii* specimen and apple cultivars forming branches B1 and B2 occurred, correspondingly, 3.45–2.36 and 1.69–1.19 Ma.

Discussion

The main conclusion of this article is that the chloroplast genome of *Malus x domestica* derives from at least two wild species, with *M. sylvestris* being the main contributor. The common origin of cpDNA of *M. sylvestris* and the majority of *M. x domestica* cultivars analyzed was supported by 100% BP. The evidence provided opens a major question: Apparently, the nuclear and chloroplast genomes of a large part of apple cultivated varieties (fig. 3) have different phylogenies.

The nuclear genome donor seems to be *M. sieversii*, as supported by the data of Velasco et al. (2010) and Micheletti et al. (2011) which 1) compared 74 accessions, including 12 *M. x domestica*, 10 *M. sieversii*, and 21 *M. sylvestris*, based on resequencing of 23 gene amplicons for a total length of 11,300 bp. The data were analyzed by a split-tree planar graph (Velasco et al. 2010) and by a maximum-likelihood method under general time reversible (GTR) model (Micheletti et al. 2011), as suggested by Harrison N and Harrison RJ (2011); 2) by comparing the same accessions (excluding putative *M. x domestica*/*M. sylvestris* hybrids) and using 27 SSR markers (Micheletti et al. 2011) unrelated to the aforementioned 23 amplicons; the phylogenetic tree was computed based on the "shared allele" distance index and the neighbour-joining (NJ) clustering algorithm. All three phylogenies were based on nuclear genes; the separation of *M. sylvestris* from *M. sieversii* was clear and highly supported by bootstrapping. *Malus x domestica* varieties clustered together with *M. sieversii*.

It is true, however, that also *M. sylvestris* has been recurrently indicated as a possible contributor to the nuclear genome of *M. x domestica* (summarized in Juniper and Mabberley [2006] and in Harrison N and Harrison RJ [2011] and Micheletti et al. [2011]), but this was, almost always, discussed considering the possibility that introgression resulted in nuclear genes private to *M. sylvestris* and not to *M. sieversii* (Micheletti et al. 2011; Harrison N and Harrison RJ 2011). Apple has been introduced to Europe by Romans and Greeks, and then from Europe it spread all over the world (Juniper and Mabberley 2006). It was proposed to have originated either in Europe, from *M. sylvestris*, a European crab apple bearing small astringent and acidulate fruits (Zohary and Hopf 1994; Coart et al. 2006; Harrison N and Harrison RJ 2011) or in Asia, from *M. sieversii* (Velasco et al. 2010; Micheletti et al. 2011; Cornille et al. 2012), a diverse central Asian species, characterized by a wide range of forms, colors, and flavors (Way et al. 1990). Abundant reports of hybridization among domesticated apple, *M. sieversii* and *M. sylvestris*,

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

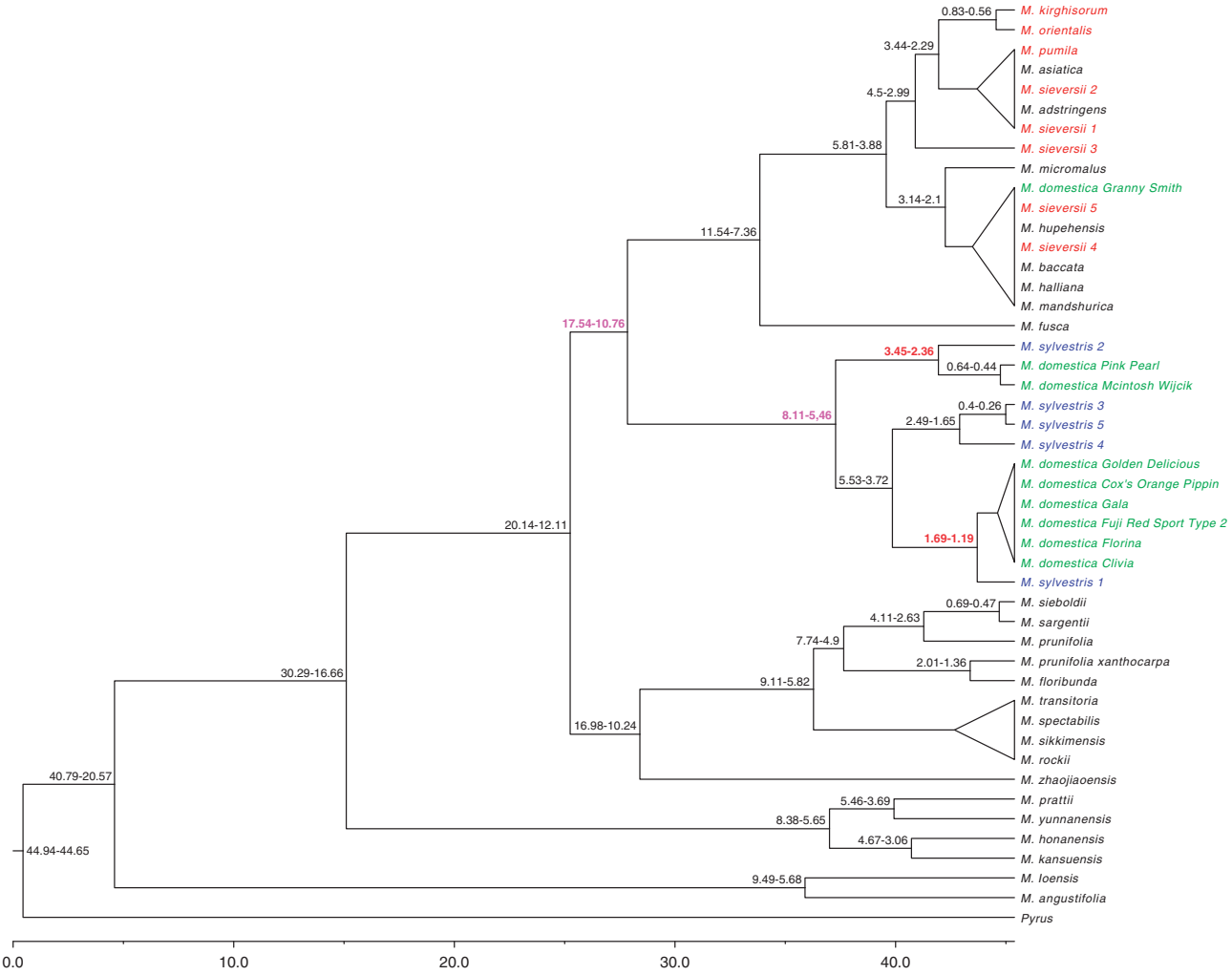

Fig. 3. Chronogram of *Malus* built employing Bayesian analysis as implemented in BEAST program from 134,553-position-long alignment of chloroplast genomes. Numbers on the left side of the tree nodes denote the age of the nodes in My. The numbers to the left of the dashes were obtained when constraining the root age to normal distribution with a mean of 45 and a standard deviation of 1. The numbers to the right of the dashes were obtained when, in addition, the age of *Malus* was constrained by a normal distribution with a mean of 20 and a standard deviation of 1. Dating for the clusters, which branching pattern could not be resolved in the ML analysis (fig. 2), was considered to be unreliable and is not shown here.

suggest polyphyletic origin of *M. x domestica* DNA loci. Cornille et al. (2012) found that 61% of the *M. x domestica* genome derives from *M. sylvestris*, which has been attributed to a recent massive introgression from the European wild apple. The introgression from *M. sylvestris* should be facilitated by self-incompatibility, long lifespan of the species, and cultural practices, including selection from open-pollinated seeds. On this subject, it must be considered that in interspecific Rosaceae hybrids, the chloroplast DNA is inherited from the maternal line (Hu et al. 2008). Thus, pollination of *M. x domestica* or *M. siversii* genotypes by *M. sylvestris* would not had led to the formation of the branch B (fig. 2), whereas the reciprocal cross remains a credible hypothesis. If the origin of the *M. x domestica* nuclear genome from *M. sieversii* is accepted, the apple varieties included in branch B would derive from hybridization events involving *M. sylvestris* as mother, followed by backcrossing with pollen from "sweet apple" genetic lines, under a strong selection to eliminate astringency components negative for fruit taste and to increase fruit size. Such a procedure was, for example, employed in the creation of scab-resistant apple cultivars, by incorporating the *Vf* gene from *M. floribunda* 821 into *M. x domestica* (Crosby et al. 1992). Studying the pedigrees of *M. x domestica* cultivars included in branch B (fig. 2) reveals that their maternal lines can be traced back to seven old "founders" (table 1). The oldest founder in branch B2, Ribston Pippin, derives from seeds brought from Rouen (Normandy) to England around 1700 (Cecil 1910). McIntosh, the oldest representative of the branch B1, was selected in Ontario, Canada, in 1792. Because, in apple, controlled breeding schemes were adopted only around 1800 (Sandlers 2010), intentional crossing and backcrossing to wild species preceding the origin of B1 and B2 branches are unlikely.

However, the same breeding outcome might have been facilitated by massive inclusion of local species into cultivation of *M. x domestica* (Hokanson et al. 2001; Robinson et al. 2001). In fact, planting apple trees from forest to gardens using root suckers was a widespread practice in central Asia

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

Table 1. Origins and Maternal Pedigrees of *M. x domestica* Cultivars Taken into Analysis.

Variety	Date of Origin
Clade B1	
Pink Pearl	
Surprise X	1944
McIntosh Wijick	
Discovered in Ontario, Canada	1796
Clade B2	
Florina	
PRI 612-1 × Jonathan	1977
Delicious × PRI 14-126	
Delicious originated in Iowa	1870
Fuji	
Ralls Janet × Red delicious	1939
Ralls Janet originated in Virginia	Late 1700s
Golden delicious	
Grimes Golden × Golden Reinette	1891
Grimes Golden was found in West Virginia	1804
Clivia	
Geheimrat Dr. Oldenburg × Cox Orane Pippin	1930
Minister von Hammerstein × Baumann's Reinette	1897
Landsberger Reinette X	1822
Cox Orange Pippin	
Ribston Pippin X	1825
Ribston Pippin originated from seeds brought from Rouen (France) in	1700
Gala	
Kidd's Orange Red × Golden delicious	1934
Delicious × Cox Orange Pippin	1924
Delicious originated in Iowa	1870
Clade A	
Granny Smith	
Eastwood, near Sydney, Australia	1868

(Ponomarenko 1983): The benefits of planting best apple individuals close to human dwellings were apparent enough that people from various places might have adopted this practice. Subsequent uncontrolled pollination among genetically heterogeneous apple cultivars, substantial proportion of which has had maternal *M. sylvestris* pedigree, would have produced results we obtained.

Consideration of the branching of the phylogenetic tree, in particularly, clear subdivision of clade B into subclades B1 and B2 (fig. 2), suggests that in *M. x domestica* the process of chloroplast genome substitution, which took place in historical time, before apple intentional breeding, occurred at least two times. Although the mechanisms responsible for the process of genome introgression are easily predicted, the forces that favored the result can only be speculated upon: Central roles may have played the selection for palatability and fruit size, unilateral compatibility in crosses, and even the fitness superiority of genotypes having cellular genomes deriving from different species. The testing of these hypotheses remains a subject of future studies. The finding that the *M. x domestica* variety Granny Smith has chloroplasts sharing a monophyletic origin with wild Asian accessions, *M. sieversii* included (Clade A, fig. 2), indicates that the process of chloroplast genome substitution in the *M. x domestica* did not affect all cultivated apple varieties.

In the study of Velasco et al. (2010), the distribution of synonymous substitution rates (K_s)—an indication of the relative age of gene duplication based on the number of synonymous substitutions in DNA coding sequences—peaked around 0.2 for recently duplicated genes, indicating that a (recent) genome-wide duplication (GWD) has shaped the genome of the domesticated apple. Dating of this GWD was based on the construction of penalized likelihood trees. Given a node of grape to rosids fixed at 115 Ma, the GWD has been dated to between 30 and 45 Ma (Fawcett et al. 2009; other references in Velasco et al. 2010). If similar rates of protein evolution are assumed for apple and poplar, the recent apple GWD may be as old as that of poplar, about 60 to 65 Ma (Tuskan et al. 2006). Because the genetic maps of *Malus* and *Pyrus* are colinear, this dating becomes the starting point for the radiation within the tribe Pyreae. At this time point, available molecular data indicate that the most probable ancestors of the event that generated the GWD were American Rosaceae species, corresponding to extant *Gillenia* and related taxa. In fact, the earliest fossils (48–50 Ma) of Pyreae genera are from North America (Wolf and Wehr 1988; Campbell et al. 2007).

Our dating results, based on chloroplast DNA (see Materials and Methods), are reported in figure 3. Under 45 Ma fossil-based constraint for the common origin of *Malus* and *Pyrus*, they indicate that radiation of extant *Malus* species might have already started 40 Ma. The basal-most branch in *Malus*, subtending *M. ioensis* and *M. angustifolia*, which natural habitats lie, respectively, in the central and eastern United States, indicates the Northern American origin of the genus. This is in good accordance with the fossil evidence (see Calibration for Estimating Divergence Times within *Malus*).

Under the same calibration, the origin of the Eurasian apple species, which progenitors, most likely, had spread to Asia via Bering Land Bridge, could be dated as approximately 30 Ma. Two basal branches of the Eurasian *Malus* subtree contain exclusively Eastern Asian, mostly Chinese species—30 Ma old branch D, subtending four extant Chinese species, survivors of the most ancient Eurasian line, and 20 Ma old branch C, bearing out a broad assemblage of species from China, Northern India, Bhutan, Japan, and Korea. *Malus zhaojiaoensis*, which ancient lineage diverged from the other apple species approximately 17 My, is the basal-most representative of the clade C.

A major subsequent diversification occurred approximately 6 Ma, likely in Central Asia, which is the center of origin of domesticated apple (Vavilov 1930): Between 25 and 47 different *Malus* species, including *M. x domestica*, are currently recognized there (Robinson et al. 2001), among which the Asiatic *M. x asiatica*, *M. baccata*, *M. micromalus*, *M. orientalis*, *M. halliana*, and *M. sieversii*. A Northern

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

American *M. fusca* was recovered as a sister group to this lineage (shown on fig. 2 as clade A) with the maximum bootstrap support. Interestingly, this species, native to the Pacific rim of North America, was considered (Routson et al. 2012) to be "the sole geographic, morphological (Van Eseltine 1933), chemical (Williams 1982), and genetic outlier among the North American taxa." Previously, amplified fragment length polymorphysm (AFLP) analyses (Qian et al. 2006) and nuclear ribosomal and chloroplast DNA phylogenetic analyses (Robinson et al. 2001) indicated its affinity to the species native to central Asia and China. *Malus fusca* was treated as belonging to the Asian section Kansuensis (Robinson et al. 2001) and was suggested to have relatively recently migrated to America across the Bering Strait (Williams 1982).

Around 17 Ma, the clade subtending European *M. sylvestris* and *M. x domestica* (clade B, fig. 2) separated from the lineage subtending the *M. fusca* plus the central Asian wild species, including *M. sieversii* (clade A). This subdivision corresponds to a major split among cpDNA lines of *M x. domestica*: the line (clade B) shared with the *M. sylvestris*, which later on, approximately 8 Ma, divided in the B1 and B2 haplotypes; and the other line, shared among the Asian wild *Malus* species, but also present in the gene pool of *M. x domestica* variety Granny Smith (Clade A).

Comparison of the topology of Branch B (fig. 2) with the geographic origin of the *M. sylvestris* reveals that the chloroplast genomes from the German *M. sylvestris* specimens (accessions 3, 4, and 5 in fig. 2) separated around 5 Ma from those present today in cultivated apple sorts. Moreover, cpDNAs of these accessions are not related to the chloroplast genomes of cultivated apple varieties, whereas southern European accessions are. Six *M. x domestica* cultivars share the chloroplast genome relationship with a *M. sylvestris* specimen collected on Monte Pollino, Calabria, Italy (*M. sylvestris* 1; fig. 2). Two other cultivars build a common branch with a *M. sylvestris* accession collected in Macedonia (*M. sylvestris* 2; fig. 1). With the limitations due to the number of accessions considered in this study, it suggests that the region where *M. sylvestris* introgressed *M. x domestica* was Southern Europe.

We conclude that using *Malus* chloroplast genome data practically free from compositional heterogeneity and from substitutional saturation, we have been able to perform a reliable phylogeny reconstruction. Phylogenetic analyses based on this alignment demonstrate that *M. sylvestris* contributed cpDNA to a large fraction of the domesticated apple sorts, indicating that chloroplast and nuclear genomes of domesticated apple may have independent evolutionary histories.

Detailed comparative analysis of parental contribution requires a robust nuclear DNA-based tree. Analysis of the largest nuclear data set amassed so far (Velasco et al. 2010) supports separation of *M. sylvestris* from *M. x domestica*/*M. sieversii* complex yet yields an overall tree topology with a number of unresolved and weakly supported branches (e.g., Micheletti et al. 2011) and, thus, cannot be used for this purpose. The possibility of mosaic genome structure in domesticated apple (Cornille et al. 2012) suggests that 27 PCR amplificates from Velasco et al. (2010) may have come from the genome loci of different origin, contributed, for example, by *M. sylvestris*, *M. sieversii*, and *M. baccata*. Thus, one explanation for a weak resolution provided by Velasco et al. (2010) data is a questionable orthology of markers sampled.

Separation of conflicting signals in nuclear data can be achieved, for example, by identification of homologous bacterial artificial chromosome clones (which, in contrast to small PCR amplificates, will contain enough characters to resolve branches in single marker analysis) followed by phylogeny reconstruction based on each group of homologous clones. Trees congruent to the chloroplast tree presented here will represent the maternal line. The rest of the trees will represent paternal line or hybrid lines (if indicated by eventual tree incongruence). Comparison of these trees should help revealing complex evolutionary history of *M. x domestica* nuclear genome.

Materials and Methods

Sequencing

Fresh leaves of 45 wild and cultivated apple accessions, including 9 accessions of *Malus x domestica*, 5 accessions of *M. sieversii*, 5 accessions of *M. sylvestris*, and 26 samples of other species (see supplementary materials, Supplementary Material online, table 1) were gathered from the apple tree collection maintained at the Fondazione Edmund Mach. DNA was extracted using the DNeasy Plant Mini kit (Qiagen, The Netherlands) and subsequently quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Life Technology, USA). Shotgun genomic libraries were generated via fragmentation of 0.5 μg of genomic DNA as described in 454 Life Sciences (Branford, CT) protocol. Briefly, DNA was randomly sheared via nebulization, and Rapid Library adaptors were blunt ligated to fragment ends. The multiplex identifier adaptors were used to distinguish reads of different specimen. Libraries were quantified via quantitative PCR using Library quantification kit—Roche 454 titanium (KAPA Biosystems, Boston, MA).

Assembly of Chloroplast DNA from Single Reads

The chloroplast genome of MD, cultivar Golden Delicious was previously sequenced at FEM (Velasco et al. 2010). The reads from 454 sff files were mapped onto this genome sequence, wherein a copy of the inverted repeat region was deleted, by gsMapper (454 Life Sciences, Branford, CT). The selected reads were subjected to de novo assembly employing gsAssembler program from the same vendor. Assemblies were transferred into the Staden package (http://sourceforge.net/projects/staden/files/) and manually edited.

The high coverage of the cpDNA contigs obtained allowed to successfully assemble chloroplast genomes. As reported (Goremykin et al. 2012), the coverage values for nuclear, mitochondrial, and chloroplast genome assemblies built from the total MD DNA preparation are 15.4 X, 168 X, and 847X, respectively. Thus, the majority option for consensus sequence building used ensures correct representation of

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

the cpDNA sequence. Among the genomes assembled, 12 contained no gaps, for the others the mean number of gaps per sequence was 4.2, and the mean estimated gap length 237 bp.

Alignment and Phylogenetic Analyses

Assembled sequences were aligned manually with the help of Seaview alignment editor, because sequence similarity among the cpDNA sequences was no less than 99%. *Pyrus* cpDNA sequence was downloaded from the Genbank (accession no. NC_015996). The alignment of 47 OTUs—134,553 aligned positions in length, available from the Dryad database (datadryad.org, doi:10.5061/dryad.33817)—was subject to ML analysis employing the PAUP* program. The search for the best-fitting model was conducted with the help of the gamma_sorter.pl script (Goremykin et al. 2010). In the first stage, model parameters were fitted to the NJ tree and the best model was selected under Akaike information criterion (AIC); in the second stage, models were fitted to the ML tree built using parameters of the best model found at the first model-fitting stage, and the next best-fitting model was also selected employing AIC.

The ML tree (fig. 1) was computed in PAUP* using settings of the best-fitting TVM + I + G model and the Tree Bisection-Reconnection search option. Bootstrap support values for the tree branches were calculated using faster MPI version of Phyml 3.0 program, which was run with the specification of the 1) TVM + I + G model, 2) the BEST search option, and 3) the ML tree previously obtained employing PAUP*.

Matrices of *p* distances and of the ML-distances computed under specification of the optimal TVM + I + G model settings, used to produce the figure 1, were generated with the help of the noiserductor.pl script (Goremykin et al. 2010) embedding PAUP* (available as supplementary material, Supplementary Material online, Goremykin et al. 2013).

Calibration for Estimating Divergence Times within *Malus*

Macrofossils assigned to Pyrinae were described in middle-to-late Eocene fossil floras from the north-western North America. Clarno Formation (~44 Ma) of central Oregon contains well-preserved silicified fruit of *Quintacava velosida*, sharing similarity with the Maloideae (Manchester 1994) and wood assigned to the Maloideae (Wheeler and Manchester 2002). Leaves classified as from *Malus* or *Pyrus* are part of the middle Eocene (about 45 Ma) flora of the Republic site in Washington (Wehr and Hopkins 1994). Thunder Mountain flora of central Idaho of the same geological age contains a leaf fossil described as "*Malus collardii*" (Axelrod 1998). Pollen assigned to *Malus* or *Pyrus* has been reported from the late Eocene Florissant locality in Colorado (Leopold and Clay-Poole 2001) estimated to be of 34.07 ± 0.10 Ma age. Fossils with similarity to *Amelanchier*, *Crataegus*, and *Photinia*, as well as some relatives of *Malus* and *Sorbus*, are known from the early middle Eocene (48–50 Ma) (Campbell et al. 2007; Wolfe and Wehr 1988).

Previous molecular dating for Pyraeae (including *Aronia*, *Malus*, *Amelanchier*, and *Crataegus*) assumed an age of 44 My for the group (Lo et al. 2009), as based on estimates of DeVore and Pigg (2007). We based our calibration on 45-My-old leaf *Malus* fossils (Wehr and Hopkins 1994; Axelrod 1998). Because of difficulty of distinguishing fossilized leaves of *Malus* from *Pyrus*, 45 My was assumed to be the approximate age of the common *Malus*/*Pyrus* lineage.

An alternative calibration corresponded to the minimum possible age for *Malus*, estimated by Lo and Donoghue (2012) as 20 My. This calibration point provides the minimum estimate for the divergence of apple species from a common progenitor.

Dating Divergence Times within *Malus*

Divergence times for the major lineages were estimated using the Bayesian method as implemented in BEAST program (Drummond and Rambaut 2007). The program was let to compute the tree topology and to optimize substitution model parameters under general definition of GTR + I + G substitution model (BEAST incorporates Hasegawa–Kishino–Yano and GTR models only). Two independent Markov chain Monte Carlo runs were performed for 10,000,000 generations, sampling every 100th generation. In both runs, uncorrelated lognormal relaxed-clock model was used, which allows rate variation across branches, and a Yule tree prior to model speciation. In one experiment, *Pyrus* was constrained to be the outgroup, and the root age was constrained by a normal distribution with a mean of 45 Ma and a standard deviation of 1. In the other dating experiment, *Pyrus* was constrained to be the outgroup, and the age of *Malus* was constrained by a normal distribution with a mean of 20 Ma and a standard deviation of 1.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Prof. Francesco Salamini for critical reading of the manuscript, several useful comments, insight, and support. The authors thank Dr Massimo Pindo for technical support of the study.

References

Arthofer W, Schueler S, Steiner FM, Schlick-Steiner BC. 2010. Chloroplast DNA-based studies in molecular ecology may be compromised by nuclear-encoded plastid sequence. *Mol Ecol.* 19:3853–3856.

Axelrod DI. 1998. The Eocene Thunder Mountain Flora of central Idaho. *Univ Calif Publ Geol Sci.* 142:1–61.

Bortiri E, Coleman-Derr D, Lazo GR, Anderson OD, Gu YQ. 2008. The complete chloroplast genome sequence of *Brachypodium distachyon*: sequence comparison and phylogenetic analysis of eight grass plastomes. *BMC Res Notes.* 1:61.

Campbell CS, Baldwin BG, Donoghue MJ, Wojciechowski MF. 1995. A phylogeny of the genera of Maloideae (Rosaceae): evidence from internal transcribed spacers of nuclear ribosomal DNA sequences and congruence with morphology. *Am J Bot.* 82:903–918.

Campbell CS, Evans RC, Morgan DR, Dickinson TA, Arsenault MP. 2007. Phylogeny of subtribe Pyrinae (formerly the Maloideae,

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

Rosaceae): limited resolution of a complex evolutionary history. *Plant Syst Evol.* 266:119–145.
Campbell CS, Wojciechowski MF, Baldwin BG, Alice LA, Donoghue MJ. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). *Mol Biol Evol.* 14: 81–90.
Cecil E. 1910. A history of gardening in England. London: John Murray.
Coart E, Van Glabeke S, De Loose M, Larsen AS, Roldan-Ruiz I. 2006. Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). *Mol Ecol.* 15:2171–2182.
Cornille A, Gladieux P, Smulders MJM, et al. (14 co-authors). 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* 8(5):e1002703.
Crosby JA, Janick J, Pecknold PC, Korban SS, O'Connor PA, Ries SM, Goffreda J, Voordeckers A. 1992. Breeding apples for scab resistance: 1945-1990. *Fruit Var J.* 46:145–166.
DeVore ML, Pigg KB. 2007. A brief review of the fossil history of the family Rosaceae with a focus on the Eocene Okanogan Highlands of eastern Washington State, USA, and British Columbia, Canada. *Plant Syst Evol.* 266:45–57.
Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
Fawcett JA, Maere S, Van de Peer Y. 2009. Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc Nat Acad Sci U S A.* 106:5737–5742.
Feng T-T, Zhou Z-Q, Tang J-M, Cheng M-H, Zhou S-L. 2007. ITS sequence variation supports the hybrid origin of *Malus toringoides* Hughes. *Can J Bot.* 85:659–666.
Gillham NW, Boynton JE, Harris EH. 1991. Transmission of plastid genes. *Cell Cult Somatic Cell Genet Plants.* 7A:55–92.
Goremykin V, Hirsch-Ernst K, Wölfl S, Hellwig F. 2003. Chloroplast genome of a "basal" angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. *Plant Syst Evol.* 242:119–135.
Goremykin V, Nikiforova S, Bininda-Emonds ORP. 2010. Automated removal of noisy data in phylogenomic analysis. *J Mol Evol.* 71: 319–331.
Goremykin VV, Lockhart PJ, Viola R, Velasco R. 2012. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *Plant J.* 71:615–626.
Goremykin VV, Nikiforova SV, Biggs PJ, Zhong B, De Lange P, Martin W, Woetzel S, Atherton RA, McLenachan T, Lockhart PJ. 2013. The evolutionary root of flowering plants. *Syst Biol.* 62:51–62.
Harris SA, Ingram T. 1991. Chloroplast DNA and biosystematics: the effect of intraspecific diversity and plastid transmission. *Taxon* 40: 393–412.
Harris SA, Robinson JP, Juniper BE. 2002. Genetic clues to the origin of the apple. *Trends Genet.* 18:426–430.
Harrison N, Harrison RJ. 2011. On the evolutionary history of domesticated apple. *Nat Genet.* 43:1043–1044.
Hokanson SC, Lamboy WF, Szewc-McFadden AK, McFerson JR. 2001. Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. *Euphytica* 118:281–294.
Hu Y, Zhang Q, Rao G, Sodmergen. 2008. Occurrence of plastids in the sperm cells of Caprifoliaceae: biparental plastid inheritance in angiosperms is unilaterally derived from maternal inheritance. *Plant Cell Physiol.* 49:958–968.
Juniper BE. 2007. The mysterious origin of the sweet apple. On its way to a grocery counter near you, this delicious fruit traversed continents and mastered coevolution. *Am Sci.* 95:44–51.
Juniper BE, Mabberley DJ. 2006. The story of the apple. Portland (OR): Timber Press Inc.
Kuroiwa T. 1991. The replication, differentiation, and inheritance of plastids with emphasis on the concept of organelle nuclei. *Int Rev Cytol.* 128:1–62.
Leopold EB, Clay-Poole S. 2001. Florissant leaf and pollen floras of Colorado compared: climatic considerations. In: Evanoff E, Gregory-Wodzicki KM, Johnson KR, editors. Fossil flora and stratigraphy of the florissant formation, Colorado. Proceedings of the Denver Museum of Nature & Science, Series 4 No 1. p. 17–69.
Li Q-Y, Guo W, Liao W-B, Macklin JA, Li J-H. 2012. Generic limits of Pyrinae: insights from nuclear ribosomal DNA sequences. *Bot Stud.* 53:151–164.
Lo EYY, Donoghue MJ. 2012. Expanded phylogenetic and dating analyses of the apples and their relatives (Pyreae, Rosaceae). *Mol Phylogenet Evol.* 63:230–243.
Lo EYY, Stefanovic S, Christensen KI, Dickinson TA. 2009. Evidence for genetic association between East Asian and Western North American *Crataegus* L. (Rosaceae) and rapid divergence of the Eastern North American lineages based on multiple DNA sequences. *Mol Phylogenet Evol.* 51:157–168.
Lo EYY, Stefanovic S, Dickinson TA. 2007. Molecular reappraisal of relationships between *Crataegus* and *Mespilus* (Rosaceae, Pyreae)—two genera or one? *Syst Bot.* 32:596–616.
Luby JJ. 2003. Taxonomic classification and brief history. In: Ferree DC and Warrington IJ, editors. Apples: botany, production and uses. Cambridge (MA): CABI Publishing. p. 1–14.
Manchester SR. 1994. Fruits and seeds of the middle Eocene nut beds flora, Clarno formation, north central Oregon. *Palaeontogr Am.* 58: 1–205.
Micheletti D, Troggio M, Salamini F, Viola R, Velasco R, Salvi S. 2011. On the evolutionary history of domesticated apple. *Nat Genet.* 43: 1044–1045.
Miller A, Gross BL. 2011. From forest to field: perennial fruit crops domestication. *Am J Bot.* 98:1389–1414.
Oh SH, Potter D. 2003. Phylogenetic utility of the second intron of LEAFY in *Neillia* and *Stephanandra* (Rosaceae) and implications for the origin of *Stephanandra*. *Mol Phylogenet Evol.* 29: 203–215.
Parks M, Cronn R, Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol.* 7:84.
Pickersgill B. 2007. Domestication of plants in the Americas: insights from Mendelian and molecular genetics. *Ann Bot.* 100:925–940.
Ponomarenko VV. 1983. History of the origin and evolution of the apple *Malus*. *Trudy po prikladnoi botanike, genetike i selektsii* 76:10–18 (in Russian, English abstract).
Qian GZ, Lui LF, Tang GG. 2006. A new selection of *Malus* (Rosaceae) from China. *Ann Bot Fenn.* 43:68–73.
Robinson JP, Harris SA, Juniper BE. 2001. Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Syst Evol.* 226:35–58.
Routson KJ, Volk GM, Richards CM, Smith SE, Nabhan GP, de Echeverria VW. 2012. Genetic variation and distribution of pacific crabapple. *J Am Soc Hort Sci.* 137:325–332.
Sandlers R. 2010. The apple book. London: Frances Lincoln Limited.
Savolainen V, Corbaz R, Moncousin C, Spichiger R, Manen JF. 1995. Chloroplast DNA variation and parentage analysis in 55 apples. *Theor Appl Genet.* 90:1138–1141.
Tuskan GA, DiFazio S, Jansson S, et al. (110 co-authors). 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313:1596–1604.
Van Eseltine GP. 1933. Notes on the species of apples. I. The American crabapples. *Agr Exp Sta New York Tech Bull.* 208:1–22.
Vavilov NI. 1930. Wild progenitors of the fruit trees of Turkestan and the Caucasus and the problem of the origin of fruit trees. Proceedings of the 9th International Horticultural Congress; 1930 August: London. p. 271–286.
Velasco R, Zharkikh A, Affourtit J, et al. (86 co-authors). 2010. The genome of the domesticated apple (*Malus domestica* Borkh.). *Nat Genet.* 42:833–839.
Way RD, Aldwinckle HS, Lamb RC, Rejman A, Sansavini S, Shen T, Watkins R, Westwood MN, Yoshida Y. 1990. Apples (*Malus*). In: Moore JN, Ballington R, editors. Genetic resources of temperate fruit and nut crops. Leuven (Belgium): International Society for Horticultural Science. p. 1–62.

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024

Wehr WC, Hopkins DO. 1994. The Eocene orchards and gardens of Republic, Washington. *Wash Geol.* 22:27–35.
Wheeler EA, Manchester SR. 2002. Woods of the Eocene nut beds flora, Clarno Formation, Oregon, USA. In: IAWA journal supplement 3. International Association of Wood Anatomists. The Netherlands: National Herbarium Nederland.
Williams AH. 1982. Chemical evidence from the flavonoids relevant to the classification of *Malus* species. *Bot J Linn Soc.* 84:31–39.
Wissemann V, Ritz CM. 2005. The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS-1 and *atp*B-*rbc*L intergenic spacer (IGS) versus conventional taxonomy. *Bot J Linn Soc.* 147: 275–290.
Wolfe JA, Wehr W. 1988. Rosaceous *Chamaebatiaria*-like foliage from the paleogene of western North America. *Aliso* 12:177–200.
Zhang Q, Li J, Zhao Y, Korban SS, Han Y. 2012. Evaluation of genetic diversity in Chinese wild apple species along with apple cultivars using SSR markers. *Plant Mol Biol Rep.* 30:539–546.
Zhong B, Deusch O, Goremykin VV, Penny D, Biggs PJ, Atherton RA, Nikiforova SV, Lockhart PJ. 2011. Systematic error in seed plant phylogenomics. *Genome Biol Evol.* 3:1340–1348.
Zohary D, Hopf M. 1994. Domestication of plants in the Old World. Oxford: Clarendon Press.
Zohary D, Spiegelroy P. 1975. Beginnings of fruit growing in the Old World. *Science* 187:319–327.

Downloaded from https://academic.oup.com/mbe/article/30/8/1751/1014963 by guest on 19 April 2024