The Siberian wild apple, *Malus baccata* (L.) Borkh., is an additional contributor to the genomes of cultivated European and Chinese apples

Xilong Chen^{*,1,2}, Amandine Cornille^{*,#,2}, Na An³, Libo Xing¹, Juanjuan Ma¹, Caiping Zhao¹, Yibin Wang¹, Mingyu Han¹, Dong Zhang^{#,1}

1. College of Horticulture, Yangling Subsidiary Center Project of the National Apple Improvement Center, Northwest A&F University, Yangling 712100, Shannxi, China

2. Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-Yvette, France

3. College of Life Sciences, Northwest A&F University, Yangling 712100, Shannxi, China.

*co-first authors

#co-directed the work

Corresponding authors:

Amandine Cornille, amandine.cornille@universite-paris-saclay.fr

Dong Zhang, afant@nwafu.edu.cn

Abstract

It is crucial to understand domestication to unravel the evolutionary processes that shape the divergence of populations. Differences in life-history traits have probably led to marked differences in the mode and speed of evolution between trees and annuals, particularly the extent of crop-wild gene flow during domestication. Apple is an iconic tree and major fruit crop grown worldwide. The contribution of wild apple species to the genetic makeup of the cultivated apple genome remains a topic of intense investigations. We used population genomics in combination with SNPs to investigate the contributions of the two known wild apple relatives, *Malus sylvestris* and *Malus sieversii*, and a supposed contributor, *Malus baccata*, to European and Chinese rootstock and dessert genomes, with a focus on the extent of wild-crop gene flow during apple domestication. We showed that the European dessert and rootstock apples form a specific gene pool, whereas the Chinese dessert and rootstock apples were a mixture of three wild gene pools. Coalescent-based inferences and gene flow estimates indicated that *M. baccata* is an additional contributor to the genome of both European and Chinese cultivated apples through wild-to-crop introgressions. We also confirmed previous results on the contribution of *M. sylvestris* to the cultivated apple genome, and provided insights into the origin of the apple rootstock. This study further demonstrates the role of gene flow during apple domestication, as seen in other woody perennials, and show that domestication of the apple tree involved several wild apple species.

39 Keywords: Introgression, perennials, clonal, fruit tree, apple, domestication, rootstock, gene

40 flow.

41

Introduction

It is crucial to understand domestication to unravel the evolutionary processes that shape the divergence of populations, varieties, and species (Gaut, Díez, & Morrell, 2015; Glémin & Bataillon, 2009; Meyer, DuVal, & Jensen, 2012). Several studies have investigated the evolutionary history of fruit crop perennials by using genome-wide data (Aravanopoulos, Ganopoulos, & Tsaftaris, 2015; Besnard, Terral, & Cornille, 2018; Cao et al., 2016; Chagné et al., 2014; Daccord, Celton, Linsmith, Becker, Choisine, Schijlen, van de Geest, Bianco, Micheletti, Velasco, Di Pierro, Gouzy, Rees, et al., 2017; Harrison & Harrison, 2011; Hazzouri et al., 2015; Huang et al., 2016; Michael & VanBuren, 2015; Teh et al., 2017; Verde et al., 2013). However, documentation of the domestication of fruit tree crops still lags behind that of annual crop plants (Burgarella et al., 2019; Gaut et al., 2015; Meyer et al., 2012; Miller & Gross, 2011; Neale, Martínez-García, De La Torre, Montanari, & Wei, 2017). Differences in life-history traits have probably led to marked differences in the mode and speed of evolution between trees and annuals (Gaut et al., 2015), particularly the extent of gene flow during domestication. Documentation of the domestication of fruit tree crops is important not only academically but also for food production. Because of long-term cultivation, fruit trees are susceptible to ever-changing environments due to climate changes and pathogen outbreaks. Sustainable fruit production is now believed to highly rely on untapped wild genetic diversity (Zhang & Batley, 2020; Zhang, Mittal, Leamy, Barazani, & Song, 2017).

Apple is an iconic tree and major fruit crop grown worldwide. It is also a model species for understanding the role of gene flow during the domestication of clonally propagated perennial crops (Cornille et al., 2019; Cornille, Giraud, Smulders, Roldán-Ruiz, & Gladieux, 2014; Peace et al., 2019; Spengler, 2019). The domestication of apple started in Central Asia, from the local Central crab apple *Malus sieversii* (Ledeb.) M. Roem. Then, apple cultivation spread towards the West, Europe and beyond, and the cultivation was assisted by the possibility of propagating interesting genotypes by grafting (Cornille et al., 2012). Thus, grafting has been an important part of apple evolution and breeding (Wang et al., 2019). With the arrival of the cultivated apple in Europe about 1,500 years ago, the local crab apple *Malus sylvestris* Mill. contributed substantially to the genome of some dessert and cider varieties through wild-to-crop introgressions (Cornille et al., 2012). Population genetic approaches in combination with microsatellite markers have demonstrated that wild-to-crop gene flow has played a major role in the evolutionary history of the cultivated apple in Europe (Cornille et al., 2012). However, while the history of the dessert and cider apples is now well documented in the western part of Eurasia (Bina et al., 2021; Cornille et al., 2014; Migicovsky et al., 2021; Peace et al., 2019; Spengler, 2019), the domestication history of apples in the East remains unclear. In addition, the history of the apple rootstock that allowed the spread of apple across the world needs to be elucidated. Understanding the origin of the apple rootstock can also provide insights into the role of crop-wild gene flow, as several rootstocks are supposed to be the result of wild-crop hybridization and

are often dwarf forms of local wild species (Volk, Cornille, Durel, & Gutierrez, 2021; Wang et al., 2019).

In China, apple cultivation has a tangled and complex history that spans a millennium and involves several wild and cultivated apple species, used for dessert or as rootstocks, that were crossed by ancient communities or bred recently. For more than 2,000 years, ancient Asian communities cultivated *Malus × domestica* ssp. *chinensis* Li Y.N. (called Mianpingguo, which means soft apple) (Gao et al., 2015; Yunong, 1999). *Malus × domestica* ssp. *chinensis* is supposed to originate from *M. sieversii* located in the northeastern part of China (Xinjiang Province), close to the borders of Kazakhstan and Kyrgyzstan. Then, the cultivated species was propagated along the Silk Routes across China where several local cultivars once existed (Gao et al., 2015). However, from the 18th century, most of the local cultivars disappeared, and, nowadays, only a few Chinese landraces remain, namely, Mianpingguo, Shaguo, Huahong, and Caiping. These four Chinese cultivars are supposed to belong to two species: *M. domestica* subsp. *chinensis* and *Malus asiatica*. Some hybrids of *M. domestica* ssp. *chinensis* and *M. asiatica* are cultivated in China. In addition, some modern apple varieties from Western countries, such as Golden Delicious, Gala, and Fuji, are also cultivated in China, and they were introduced from Europe and the United States less than a decade ago. A few other cultivated Chinese apple species are traditionally used as rootstocks and for ornamental purposes, and some of them originated from several hybridization events (Gao et al., 2015). *Malus prunifolia* (Chinese crabapple called Qiuzi) and *Malus robusta* (i.e., *Malus baccata* × *M. prunifolia* cross

called Balenghaitang) are used as rootstocks. *Malus micromalus*, *Malus halliana*, and *Malus spectabilis* are used as ornamental species. These rootstocks and ornamental species cannot be found in natural habitats in China, so they are classified as cultivated species (Yunong, 1999). The Chinese people have also used local wild apple species for millennia. Because of their ease of propagation through apomixis and their cold resistance, *M. baccata* and *Malus hupehensis* are supposed to be the progenitors of the Chinese rootstocks. Indeed, *Malus baccata*, the Siberian wild apple, is the main genetic resource (especially as a rootstock) for apple breeding programs in China, because of its excellent resistance to cold stress and apple scab (Chen et al., 2019; Gyax et al., 2004; Volk et al., 2015). *Malus hupehensis*, the tea crabapple or Hupeh crab, is an apomictic species. Its seedling rootstocks can be propagated directly through seedlings, instead of using cuttings or layering propagation (Wang et al., 2019; Yang, Duan, & Zhang, 2008). China also introduced European-bred dwarf rootstocks in the 19th century, but most apple production areas in China still use local cultivated (*M. prunifolia* and *M. robusta*) and wild (*M. hupehensis* and *M. baccata*) apples as rootstocks because of their good adaptability (Wang et al., 2019).

Apple domestication and breeding in China have therefore probably involved crop-crop and wild-crop gene flow and contributions of several wild species. However, all available information on apple domestication in China is derived from historical records and traditional knowledge. Only Duan et al. (2017) showed that *M. asiatica* and *M. prunifolia* originated from crosses between *M. sieversii* and *M. baccata*, but this result remains anecdotic. It is therefore still unclear whether cultivated Chinese dessert and rootstock apples originated from *M. sieversii*

and/or *M. baccata* or/and *M. hupehensis*, or both originated from *M. sieversii* followed by recurrent wild-to-crop gene flow from *M. baccata* or/and *M. hupehensis*. Besides, the genetic relationships among wild and cultivated European and Chinese apples have not yet been elucidated, and it is unclear whether substantial crop-to-wild gene flow occurred during apple domestication in China, as observed in Europe and the Caucasus (Cornille et al., 2012; Bina et al., 2021).

Here, we investigated the history of domestication of European and Chinese dessert apples and rootstocks, with a focus on the extent of crop-wild gene flow during domestication. We combined previously published sequenced genomes (Duan *et al.*, 2017) with a new dataset that included wild and cultivated Chinese apples sequenced for their RNA. Using single-nucleotide polymorphisms (SNPs) called from several wild apple species, including the wild apple relatives of *M. domestica* (*i.e.*, *M. sieversii* and *M. sylvestris*), *M. baccata*, *M. hupehensis*, and other ancestral species from the genus *Malus*, as well as European and Chinese apple cultivars, we aimed to answer the following questions: (1) What is the population structure and differentiation of wild and cultivated apples from Eurasia? (2) Did gene flow occur between wild and cultivated apples during domestication? (3) What is the domestication history of Chinese and European dessert apple and rootstock cultivars and were they independently domesticated from different wild species?

Materials and Methods

Plant samples

For this study, fruit flesh samples of 71 *Malus* accessions, representing 48 wild and 23 cultivated apple samples, were collected from repositories in China (Table S1). Additionally, short-read DNA sequences of 97 *Malus* accessions (Duan *et al.*, 2017), namely, 53 wild apple samples and 44 apple cultivars, were downloaded (SRA code accession: SRP075497, Table S2). The two datasets, which included 168 accessions with 66 apple cultivars (7 historical Chinese cultivars, 30 cultivars from Europe or Western countries, 25 rootstocks, and 4 ornamental cultivars) and 102 wild apples (43 *M. sieversii*, 18 *M. baccata*, 11 *M. sylvestris*, and 30 other *Malus* accessions), were merged for subsequent analyses (Table S1).

RNA extraction and library construction and sequencing

The fruit flesh of the 71 collected samples was immediately transferred to liquid nitrogen and stored at -80°C until RNA extraction. The RNA was extracted using the SDS–phenol method (Hu *et al.*, 2002). The RNA quality was checked using an agarose gel, and RNA concentrations were estimated with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). After RNA extraction and DNase I treatment, mRNA was isolated from the total RNA by using magnetic oligo (dT) beads. The mRNA was mixed with fragmentation buffer and cleaved into short fragments and used as templates for cDNA synthesis. The short fragments were purified, resolved with EB buffer for end reparation and single adenine nucleotide addition and connected with adapters. After agarose gel electrophoresis, suitable

fragments were selected as templates for polymerase chain reaction (PCR) amplification. During the quality control steps, an Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR system (Agilent, US) were used for evaluating the quantity and quality of the sample libraries. Finally, the constructed libraries were sequenced on an Illumina HiSeq 2000 system (BGI, Shenzhen, China). Quality control of the raw data was performed with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), and 420.73 GB of clean data was obtained (Table S3).

Read mapping and SNP calling and filtering

Two different pipelines were used for the sequenced DNA and RNA reads (Figure S1). For the sequenced RNA accessions, reads were processed as follows: SNP calling was performed on the basis of the transcript sequence data with GATK v3.5 calling variants pipeline for RNAseq (McKenna et al., 2010) (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531192?id=3891>). Filtered reads were mapped to the new high-quality *M. domestica* reference genome GDDH13 v1.1 (Daccord *et al.*, 2017) with Hisat2 software (Kim, Langmead, & Salzberg, 2015). The SAM file was processed with Picard (*Picard toolkit*, 2019) in several steps, as described by ForgeMIA. The GATK module HaplotypeCaller was used for SNP calling, and low-quality SNPs were filtered out with the GATK module VariantFiltration. For short-read DNA sequences of the 97 accessions (Duan *et al.*, 2017), SNPs were called using the GATK v3.5 pipeline (McKenna et al., 2010). The reads were aligned to the GDHH13 v1.1 *M.*

domestica reference genome (Daccord *et al.*, 2017) with BWA by using the “mem” algorithm (Li, 2013). The redundant reads were removed using the MarkDuplicate module from Picard (“Picard Toolkit,” 2019). The GATK module HaplotypeCaller was used for SNP calling, and GenotypeGVCFs module was used to produce a raw set of joint SNPs. The raw joint SNPs were finally filtered with the GATK VariantFiltration module and default quality thresholds.

Suitability of the markers used for phylogenetic, population structure, and demographic inferences

The filtering steps used for the analyses described below are provided in Figure S1. The SNP files obtained from the RNA-seq and DNA-seq datasets were combined with the VCFtools vcf-merge function (Danecek *et al.*, 2011). The clone and duplicate samples were removed (*i.e.*, individuals with pairwise KING-robust kinship estimates among individuals > 0.354) (Manichaikul *et al.*, 2010)) with Plink 2.0 (Chang *et al.*, 2015). The SNPs with minimum minor allele frequency < 0.01 and minimum site coverage $= 0.9$ were filtered out with Plink (Purcell *et al.*, 2007). A total of 76,239 SNPs were obtained. For population structure and demographic inferences, linked and non-synonymous SNPs were filtered out to avoid any bias associated with potential SNPs linked to the selected genomic region. First, SNPs in linkage disequilibrium (LD) were removed; the LD for apples was $< 1,000$ bp (Duan *et al.*, 2017). The linked SNPs were filtered out with Plink (Purcell *et al.*, 2007) by using the following parameters: window = 1 kb, step = 1 SNP, and $r^2 = 0.2$. Second, the remaining SNPs were annotated with SnpEff 4.0

(<https://pcingola.github.io/SnpEff/>) (Cingolani et al., 2012), and only synonymous SNPs were retained; therefore, 36,200 unlinked and synonymous SNPs were used (Figure S1).

Population structure, genetic diversity, and differentiation among wild and cultivated apples

The population structure and admixture among *Malus* accessions were inferred using a model-based clustering method implemented in ADMIXTURE v.1.23 (Alexander, Novembre, & Lange, 2009). ADMIXTURE is based on the same statistical model as the one implemented in STRUCTURE (Pritchard, Stephens, & Donnelly, 2000), but it uses a fast numerical optimization algorithm for large SNP datasets to infer the proportion of ancestry of genotypes in K distinct predefined clusters. The number of ancestral populations, K , was varied from 1 to 10, with 20 repetitions per K and default settings. The consensus solution for each K value was generated with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The amount of additional information explained by increasing K was determined using the cross-validation procedure of ADMIXTURE (Alexander et al., 2009). However, the K identified with the cross-validation procedure often does not correspond to the finest biologically relevant population structure (Cornille et al., 2015; Kalinowski, 2011; Puechmaille, 2016). Therefore, the bar plots were visually checked, and the K value for which all clusters had well-assigned individuals was selected (*i.e.*, that no further well-delimited and relevant clusters could be identified for higher K).

values than the chosen one). All the bar plots were generated with the R package “Pophelper” (Francis, 2017)

The genetic variation and differentiation among the groups detected with ADMIXTURE were further investigated using three different methods. First, a principal component analysis (PCA) was performed with the `dudi.pca` function of the `ade4` package in R (Dray & Dufour, 2007). The first two principal components were plotted with the R package “ggplot2” (Wickham, 2009). Second, a neighbor-net tree based on Nei’s distance (Masatoshi Nei, 1978) among individuals was created. The neighbor net was visualized with Splitstree v4 (Huson & Bryant, 2006). For the PCA and neighbor-net tree, individuals assigned to a given cluster with a membership coefficient ≥ 0.55 were indicated using the color of each cluster, and admixed individuals (*i.e.*, individuals with a membership coefficient to any given cluster < 0.55) were colored gray. We chose this threshold because both cultivated and wild apples were highly admixed, and the use of a higher cut-off would have removed too many individuals (see Results). For each wild or cultivated population (*i.e.*, group of individuals with membership coefficient > 0.55 to a given cluster detected with ADMIXTURE), Nei’s diversity (π) (M Nei & Li, 1979), observed (H_O) and expected heterozygosity (H_E) (Masatoshi Nei, 1973), inbreeding coefficient F_{IS} , and genetic differentiation (F_{ST}) between populations were estimated with Stacks population program v2.52 (Rochette, Rivera-Colón, & Catchen, 2019).

Detection of gene flow

Two approaches were used to investigate the occurrence of gene flow among the populations (*i.e.*, group of individuals with a membership coefficient > 0.55 to a given genetic cluster detected with ADMIXTURE).

First, TreeMix (Pickrell & Pritchard, 2012) with the ipyrad tool (Eaton & Overcast, 2020) was used to analyze the relationships among the populations and the potential influence of gene flow. TreeMix uses genome-wide allele frequency data to estimate the maximum likelihood tree of populations. Then, on the basis of populations with the poorest fit to the tree model, the program infers the presence and magnitude of migration events between them (Pickrell & Pritchard, 2012). Trees were rooted with the *Malus* species from North America, which are known to be the most divergent and ancestral *Malus* group (Harris *et al.*, 2002 and see Results). Zero to 9 migration numbers were tested, and the run for which the likelihood of trees reached an initial plateau (Aguirre-Liguori *et al.*, 2019; Brandenburg *et al.*, 2017) and the magnitude of migration rates decreased was retained.

Second, Dtrios of Dsuite (Malinsky, Matschiner, & Svardal, 2021) was used to perform four-taxon *D-statistic* tests on the final unlinked synonymous SNP dataset. *D-statistics* is also called Patterson's *D* (ABBA-BABA) statistic (Durand, Patterson, Reich, & Slatkin, 2011; Green *et al.*, 2010), and it checks whether gene flow occurred among populations (*i.e.*, group of individuals assigned with a membership coefficient > 0.55 to a given genetic group) with ADMIXTURE and 36,200 synonymous unlinked SNPs. Dsuite calculates *D-statistics* not only for one specific quartet but also, comprehensively, for sets of populations in a VCF file and

keeps the outgroup fixed (here the North American wild apples, see Results). The significance of *D-statistics* was assessed using Jackknife (Green et al., 2010) on 20 blocks. Results for which the Z-score was higher than 3 and *P-value* < 0.01 were retained.

Approximate Bayesian computation

The history of apple domestication was reconstructed using the approximate Bayesian computation (ABC) framework in combination with a coalescent-based inference simulator, fastsimcoal2 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013). Our aims were to infer (1) whether gene flow occurred among certain crop and wild populations during apple domestication and (2) the European and Chinese cultivated apples originated from which wild apple population(s). Previous studies (Cornille, Giraud, Smulders, Roldán-Ruiz, & Gladieux, 2014; Cornille et al., 2012; Duan et al., 2017; Migicovsky, Gardner, Richards, Thomas Chao, et al., 2021) have reported that the European dessert apple (*M. domestica*) diverged from *M. sieversii*. Therefore, the origin of the European dessert apple was not tested, and it was assumed that it diverged from *M. sieversii* in our models. The scenarios were established according to the results obtained from ADMIXTURE, PCA, and neighbor-net tree as well as the results of *D-statistics* and TreeMix. The populations were defined as those detected with ADMIXTURE analyses (*i.e.*, group of individuals assigned with a membership coefficient > 0.55 to a given genetic group).

A newly developed ABC method was used for model selection and parameter estimation according to a machine learning tool called “Random Forest” (ABC-RF). This approach allowed us to disentangle complex demographic models (Pudlo et al., 2016) by comparing groups of scenarios with a specific type of evolutionary event to other groups with different types of evolutionary events (instead of considering all scenarios separately) (Estoup, Raynal, Verdu, & Marin, 2018) in what we will hereafter call “ABC rounds.” Such a grouping approach in the scenario choice is more powerful than testing all scenarios individually to disentangle the main evolutionary events that characterize speciation (Estoup et al., 2018).

Two nested sets of ABC analyses were run, and, within each set, several rounds were run (Figure S2). Such a nested ABC approach avoids the comparison of too complex models with numerous populations and parameters (Estoup et al., 2018). The first ABC set tested the history of domestication of the European apple rootstock (set 1, Figure S2). Once the history of the cultivated European apples was inferred, a second set of ABC analyses was run to infer the domestication history of the cultivated Chinese dessert and rootstock apples (set 2, Figure S3). For the two ABC sets, the modalities of gene flow among crop and wild populations were defined based on the TreeMix analyses. Prior distributions for each parameter are provided in Table S4.

For all models, 36,200 unlinked SNPs were simulated, with 8,000 simulations per scenario. For each simulation, the following summary statistics were computed with arlsumstats v 3.5 (Excoffier & Lischer, 2010): the number of sites with segregating substitutions of

population i (S_i , $i = \{1, 2, 3, 4, 5\}$), mean number of pairwise differences of population i (π_i , $i = \{1, 2, 3, 4, 5\}$), mean number of differences between pairs of populations (π_{i-j} , $i, j = \{1, 2, 3, 4, 5\}$, $i \neq j$), and pairwise F_{ST} ($F_{ST_{i-j}}$, $i, j = \{1, 2, 3, 4, 5\}$, $i \neq j$). Summary statistics were also added on the basis of the joint frequency spectrum between all pairs of populations (Tellier et al., 2011; Wakeley & Hey, 1997) computed with a home-made script: sites polymorphic in population i , but monomorphic in population j , and vice-versa ($S_{x1_{i-j}}$ and $S_{x2_{i-j}}$, respectively); number of shared polymorphic sites between population i and population j (S_{si-j}); and number of sites showing fixed differences between populations i and j (S_{fi-j}).

The *abcrf* v.1.7.0 R statistical package (Pudlo et al., 2016) was used to perform the ABC-RF analysis, which provides a classification vote that represents the number of times a scenario is selected as the best one among n trees in the constructed random forest. For each ABC set, the scenario, or a group of scenarios, with the highest number of classification votes was selected as the best scenario, or best group of scenarios, among a total of 500 classification trees (Breiman, 2001). The posterior probabilities and prior error rates (*i.e.*, the probability of choosing a wrong group of scenarios when drawing model index and parameter values from the priors of the best scenario) over 10 replicate analyses (Estoup et al., 2018) were computed for each ABC step. The simulated models were also visually checked to be compatible with the observed dataset by projecting the simulated and observed datasets onto the first two linear discriminant analysis (LDA) axes (Pudlo et al., 2016) and whether the observed dataset was within the clouds of simulated datasets. The parameter inferences were then calculated using the final selected model,

according to the three-set ABC procedure. The ABC-RF approach includes the model checking step performed *a posteriori* in previous ABC methods.

Once the most likely apple domestication history was obtained, posterior distributions were estimated for each parameter and 1,000 pseudo-observed datasets were re-simulated using the 95% confidence interval of the posterior estimates to evaluate the goodness of fit of the final chosen demographic model with *abc* R package (Csilléry, François, & Blum, 2012).

Results

The European and Chinese cultivars have different ancestries

We removed 20 individuals of the initial dataset ($N = 168$) because they had unknown passport information ($N = 1$) or were detected as a clone or duplicate ($N = 19$, Table S1). The first run of ADMIXTURE, in which the 20 individuals were removed ($N = 148$), revealed 20 wild apple trees described as *M. sieversii* or *M. sylvestris* or *M. baccata* admixed with the cultivated gene pool (Figures S4 and S5) or misidentified on the field (*i.e.*, fully assigned to *M. domestica* gene pool), as previously observed (Cornille, Gladieux, & Giraud, 2013; Cornille et al., 2015, 2014; Feurtey, Cornille, Shykoff, Snirc, & Giraud, 2017). We also detected 13 wild individuals, described as not one of the three wild apple species studied here, admixed with multiple gene pools (Table S1, Figures S4 and S5). In this study, we focused on the history of the cultivated apple, so we wanted “pure” wild apple reference samples of the three wild apple relatives, *i.e.*, *M. baccata*, *M. sieversii*, and *M. sylvestris*. Therefore, we filtered out the 33 wild admixed or

misidentified individuals detected with the first ADMIXTURE analysis (Figure S6, Table 1). We retained *M. hupehensis*, as the samples belonging to this species grouped with *M. baccata* at a high membership (Figure S6). Population genetic structure inferences for the whole dataset ($N = 148$) vs. the pruned dataset for those admixed individuals ($N = 115$) were consistent (Figures S4 and 1a). However, ADMIXTURE revealed a clearer population genetic structure for the dataset pruned for wild admixed individuals. Therefore, we retained the dataset pruned for wild admixed or misidentified individuals for further analyses.

ADMIXTURE revealed five main genetic groups (Figure 1a). Cross-validation error monotonically decreased up to $K = 8$ (Figures S7 and S8). At $K > 5$, further substructure was observed, but new clusters were represented by only admixed individuals (Figure S7). Besides, at $K = 5$, the population structure was consistent with the morphological classification of the wild species (Cornille et al., 2019, 2014, 2012; Harris et al., 2002; Robinson, Harris, & Juniper, 2001), with *M. sylvestris* (orange), *M. sieversii* (purple), *M. baccata*/*M. hupehensis* (blue), and North American wild apples (green) each forming a distinct group. At $K = 5$, the European apple cultivars (*i.e.*, *M. domestica*, including European dessert and rootstock cultivars) formed a distinct genetic group highly admixed with *M. sylvestris* (Table 1), as shown by Cornille et al. (2012), but also with the *M. baccata*/*M. hupehensis* (blue) genetic group (Table 1). For $K = 5$, the Chinese dessert and rootstocks as well as ornamental Chinese apple species did not form a specific genetic group, but a genetic mixture of the blue (*M. baccata*/*M. hupehensis*) and purple

(*M. sieversii*) gene pools (Table 1) as well as, to a lesser extent, with the orange genetic group (*M. sylvestris*).

Then, we assessed the genetic variations among wild and cultivated apple populations detected with ADMIXTURE. We assigned wild individuals with a membership coefficient > 0.55 to a given cluster in the corresponding population (Table 1). Because of the high level of admixture in cultivated apples, we retained all cultivars and defined five cultivated apple groups based on the ADMIXTURE results as well as their historical uses and geographic origins (Tables 1 and S1). The use and geographic origins of the cultivars were documented on the basis of the studies performed by Morgan & Richards (1993, 2003) and Urrestarazu *et al.* (2016) for the European-Western cultivars and Zhi-Qin (1999) for the Chinese cultivars. The North American and Siberian wild apples were the most genetically differentiated population (Figure 1b, c, d). *Malus sieversii* individuals formed a distinct group (Figure 1b), close to that of the European dessert apple. The European cultivars (*M. domestica*) were closer to *M. sylvestris* than to the known progenitor *M. sieversii* (Figure 1b, c, d), as detected previously (Cornille *et al.*, 2012; Duan *et al.*, 2017; Migicovsky, Gardner, Richards, Thomas Chao, *et al.*, 2021). *Malus sylvestris* was nested with *M. domestica* in the neighbor-net tree (Figure 1b), but it formed a distinct, but close, gene pool to that of *M. domestica* in the PCA (Figure 1c). Most of the European rootstocks were admixed with the three wild apple species (Figure 1b, c). The Chinese dessert and rootstock formed a bushy structure intermingled with *M. baccata* and *M. hupehensis*. F_{ST} estimates further confirmed that the European cultivars were genetically close to *M. sylvestris* and *M. sieversii*,

and the Chinese cultivars were genetically close to *M. baccata*-*M. hupehensis* (Figure 1d). Genetic diversity estimates for the nine populations are provided in Figure S9 and Table S5. Genetic diversity was significantly lower in the wild apple than in the cultivated apple gene pools, except for the *M. baccata*/*M. hupehensis* group; this can be explained by the high level of admixed individuals in the cultivated populations (Table 1, Figure 1).

Therefore, ADMIXTURE, neighbor-net tree, PCA, and F_{ST} suggest that the cultivated European and Chinese apples have different ancestries, which raise questions about their origin and whether crop-wild gene flow occurred. The European dessert and rootstock cultivars grouped together (*M. domestica*) and formed a specific gene pool; this has been previously reported for the dessert apples (Cornille et al., 2012; Migicovsky, Gardner, Richards, Thomas Chao, et al., 2021). The European dessert and rootstock cultivars showed different levels of admixture with the wild species: the European dessert cultivars were mostly admixed with *M. sylvestris* (and *M. sylvestris* was nested in *M. domestica*), whereas the European rootstock cultivars were mainly admixed with *M. baccata* and *M. sieversii*. The high level of admixture of the European rootstock with *M. baccata* and *M. sieversii* raises questions about the origin of the European rootstock cultivars: their distinct genetic differentiation from *M. baccata* and clustering with *M. domestica* (which we know originated from *M. sieversii*) may suggest that *M. baccata* contributed to the European rootstock through recent crop-to-wild introgressions and the European rootstocks originated from *M. sieversii*. In contrast to the European cultivars (*M. domestica*), the Chinese cultivars, both dessert and rootstock, did not form a specific gene pool,

but they were a mixture of *M. sieversii* or *M. baccata* gene pools. The high level of admixture of the Chinese cultivars (rootstock or dessert) with *M. baccata* and *M. sieversii* and even, sometimes, their full membership to genetic clusters of two wild apple gene pools, suggests that wild apple trees are grown in orchards for consumption without any strong domestication process in China and/or substantial wild-crop gene flow occurred. In addition, the observation that the Chinese cultivars are nested and genetically close to *M. baccata* (in contrast, *M. domestica* was grouped with *M. sieversii* and *M. sylvestris*) suggests that *M. baccata* may be the wild progenitor of the Chinese cultivars; domesticated populations are expected to be nested within their source population because they recently diverged from a subset of individuals within the source population (Matsuoka et al., 2002). Although our analyses showed that *M. sylvestris* is nested within *M. domestica*, *M. sylvestris* is known to not be a progenitor of the cultivated apple; it is a secondary contributor through crop-to-wild introgressions. Therefore, we investigated the contributions of each wild species by crop-to-wild gene flow or/and an initial domestication event.

Substantial wild-to-crop gene flow from the three wild apple relatives

TreeMix (Figures 2 and S10) indicated the occurrence of gene flow from the European wild apple to the European dessert and rootstock, but not from *M. sieversii*. TreeMix also indicated gene flow from *M. sieversii* to the Chinese apple rootstock and from *M. baccata* to the Chinese dessert. The D-suite inferences were congruent with TreeMix results, with significant excess of

sharing of derived alleles among almost all crop and wild population pairs (Table S6). In addition, TreeMix analyses suggested that the cultivated European apples formed a monophyletic group close to *M. sieversii*; in contrast, the Chinese cultivars formed a polyphyletic group, with the Chinese dessert apple being a sister group with *M. sieversii* and Chinese rootstock, a sister group with *M. baccata*. However, TreeMix can be biased by its starting tree, especially when many populations are admixed (Lipson, 2020; Lipson et al., 2013). Therefore, we investigated the domestication history of the Chinese and European cultivars by using coalescent-based approaches, which allow us to disentangle between crop-wild gene flow and ancestral polymorphisms.

Domestication of the European and Chinese apples from *M. sieversii*-*M. domestica* followed by substantial gene flow from the Siberian and European wild apples

For ABC inferences, we excluded the North American wild apple population, as it was the most genetically distant and did not contribute to the cultivated apple genome in the analyses (Figure 1). We also excluded the Chinese ornamental apple because it included only four individuals. Our aim was to analyze (1) the extent of crop-wild gene flow and (2) infer the divergence history of the cultivated apple populations (excluding the European dessert, as its history is already known) (Cornille et al., 2012; Duan et al., 2017; Harris et al., 2002; Migicovsky, Gardner, Richards, Thomas Chao, et al., 2021; Sun et al., 2020; Velasco et al., 2010).

First, we inferred the origin of the cultivated European rootstock apples (Figures S2 and 3) with respect to the three wild apple populations and European dessert. We assumed that three wild apples diverged from an unknown ancestral population and the European dessert apple *M. domestica* diverged from *M. sieversii*. Our results above (population structure and gene flow estimates) suggest that *M. baccata*-*M. hupehensis* did contribute to the European rootstock through recent crop-to-wild introgressions and the European rootstock cultivars originated from *M. sieversii*. For the first set of ABC (Figure S2), we tested whether the European rootstock diverged from *M. domestica* or *M. sieversii* (Figure S2) and assumed four different modalities of gene flow: no gene flow, gene flow between European dessert and *M. sylvestris* only (Cornille et al., 2012), gene flow between the European rootstock and *M. sylvestris* and *M. baccata*/*M. hupehensis* only, and a combination of the first two modalities of gene flow (Figure S2). The last modality was defined on the basis of an additional TreeMix analysis that focused on only the populations used for ABC set 1 (Figure S11). Therefore, we simulated eight scenarios for the first ABC set (Figure S2). For the second set of ABC analysis (set 2, Figure S3), we inferred the domestication history of the Chinese dessert and rootstock apple on the basis of the most likely group of scenarios selected from the first ABC set. We tested whether the Chinese dessert and rootstock cultivated apples diverged from (i) *M. domestica*, (ii) *M. baccata*-*M. hupehensis*, or (iii) *M. sieversii*. We defined 12 different scenarios of divergence of the Chinese cultivars and assumed two modalities of gene flow: gene flow as inferred from TreeMix (Figure 2) and no gene flow between the Chinese rootstock and *M. baccata*-*M. hupehensis* and *M. sieversii* + no

gene flow between the Chinese dessert with *M. baccata*-*M. hupehensis* and *M. sieversii* (Figure S3). In total, we simulated 24 scenarios. For each ABC round, the observed summary statistics were within the cloud of simulated summary statistics, which did not overlap across the model groups (Figures S9, S10) and indicated that these tests could discriminate among various competing scenarios (except for the fourth round of the second ABC set, Figure 3).

For the first round of the first ABC set (Figures 3a and S2), the classification votes were the highest 10 times out of 10 for the group of scenarios that assumed gene flow among the European dessert (*M. domestica*) and *M. sylvestris*, European rootstock and *M. sylvestris*, and *M. baccata*/*M. hupehensis* (i.e., gene flow modality 4 in Figures S3 and 3, 479.50 votes out of the 500 RF trees, posterior probability $P = 1.00$, prior error rate = 0.00, Table S7, Figure S12a, Figure 3a). In the second round, the classification votes were the highest 10 times out of 10 for the scenarios that assumed the European rootstock from *M. sieversii* (422 votes out of the 500 RF trees, posterior probability $P = 0.87$, prior error rate = 10.02%, Table S8, Figure S12b). Thus, the first ABC set indicated the European rootstock originated from *M. sieversii* with subsequent gene flow from *M. baccata* and *M. sylvestris*.

For the first round of the second ABC set, the classification votes were the highest 10 times out of 10 for the scenarios assuming gene flow between the Chinese rootstock and dessert with *M. baccata* (i.e., gene flow modality 2, 98.9% RF trees voting for this group, posterior probability $P = 1.00$, prior error rate = 0.00%, Table S9 and Figure S13a). For the second round of the second ABC set, the classification votes were the highest 10 times out of 10 for the

scenarios that assumed the two Chinese cultivated apples were domesticated independently from different wild species (67.8% RF trees voting for this group, posterior probability $P = 0.88$, prior error rate = 9.61%, Table S10 and Figure S10b). For the third round, the classification votes were the highest 10 times out of 10 for the scenarios that assumed the Chinese rootstock and dessert originated from *M. sieversii* or *M. domestica* (56% RF trees voting for this group of scenarios, posterior probability $P = 1.00$, prior error rate = 0.00%, Table S11 and Figure S13c). For the fourth round, the classification votes were the highest 10 times out of 10 for the scenarios that assumed the Chinese dessert apple originated from *M. domestica* (57.6% votes out of the 500 RF trees, posterior probability $P = 0.78$, prior error rate = 15.27%, Table S12, Figure S13d). However, origin of the Chinese dessert from *M. domestica* needs to be considered with caution, as we were not able to fit the observed data within the LDA (Figure S13d).

The final scenario indicated that the European rootstock diverged from *M. sieversii* and the Chinese cultivars did not diverge from *M. baccata*, but their origin, either from *M. sieversii* or *M. domestica*, remains unclear (Figure 3b). The ABC further confirmed TreeMix gene flow estimates: *M. sylvestris* and *M. baccata* contributed to the European and Chinese dessert and rootstock gene pools (Figure 3b). The model parameters are listed in Table S13, and the posterior estimates are presented in Table S13. However, it should be noted that the credibility intervals were large and the parameter estimates need to be carefully considered. The goodness of fit of the two models (Figures S15 and S16) confirmed our model choice ($P = 0.069$ and $P = 0.07$, respectively).

496

497 **Discussion**

498 In this study, we used population genomic approaches in combination with SNPs to investigate
 499 the relative contributions of each wild apple relatives, *M. sylvestris* and *M. sieversii*, and a
 500 supposed contributor, *M. baccata*, to the genomes of cultivated European and Chinese dessert
 501 and rootstock apples. We showed that the cultivated European dessert and rootstock apples
 502 grouped together and formed a specific gene pool, whereas the Chinese dessert and rootstock
 503 apples were a mixture of the three wild apple gene pools, mainly with *M. baccata* and *M.*
 504 *sieversii*. The coalescent-based inferences indicate that both European and Chinese rootstocks
 505 diverged from *M. sieversii* (or *M. domestica* for the cultivated Chinese apple gene pool), with
 506 subsequent gene flow from the wild species *M. sylvestris* and *M. baccata*. We also confirmed
 507 previous results for the contribution of *M. sylvestris* to the cultivated dessert apple gene pool
 508 (Cornille et al., 2012). Therefore, our results show that *M. baccata* is an additional contributor to
 509 the cultivated apple genome, and we have also provided insight into the origin of the apple
 510 rootstock. This study confirmed that domestication of the apple tree involved several wild apple
 511 species and that crop-wild species hybridization had a key role in fruit tree domestication.
 512 Substantial hybridization between domestic and wild forms have also been described in other
 513 fruit trees (Groppi, Liu, Cornille, Decroocq, & Decroocq, 2021; Liu et al., 2019; Wu et al.,
 514 2018), but the apple tree is the model system with the most documentation. Our results support
 515 the view that domestication of woody perennials, and crops in general, was probably a protracted

and diffuse process that involved multiple geographically disparate species (Allaby, Fuller, & Brown, 2008; Purugganan, 2019).

The European and Chinese apple rootstocks did not originate from *M. baccata*

Although grafting has been an important part of perennial woody crop evolution and breeding (Vavilov, 1926; Warschafsky et al., 2016; Zohary & Spiegel-Roy, 1975), surprisingly little is known about the origin of rootstocks. In apples, most of the studies focused on the history of dessert and cider cultivars, mostly from the Western Hemisphere. The Central Asian wild apple, *M. sieversii*, is now known to be the wild ancestor of the cultivated dessert and cider apple, *M. domestica* (Cornille et al., 2012; Cornille et al., 2019; Cornille, Giraud, Smulders, Roldán-Ruiz, & Gladieux, 2014; Daccord et al., 2017; Duan et al., 2017; Harris, Robinson, & Juniper, 2002; Migicovsky, Gardner, Richards, Thomas Chao et al., 2021; Peace et al., 2019). However, the origin of apple rootstock has not yet been elucidated. Here, we provide new insights into the domestication history of apple rootstock and show that *M. sieversii* is the progenitor of the European apple rootstock and either *M. sieversii* or *M. domestica* is the progenitor of the Chinese apple. The mixed origin of Chinese dessert cultivars, the shared ancestral polymorphism between *M. domestica* and *M. sieversii*, and the current low number of Chinese cultivars, may explain our inability to infer the origin of the Chinese dessert apple with ABC. Indeed, only four main Chinese cultivars are still cultivated in China; they are supposed to belong to two species, *M. domestica* subsp. *chinensis* and *M. asiatica*. Note as well that only few main rootstocks are used

in China (Wang et al., 2019). In our study, the sampling size of the Chinese cultivated apple was limited because of its cultivation history in this country.

Thus, despite the large use of *M. baccata* and *M. hupehensis* (Wang et al., 2019) as breeding sources in China and for rootstock in Europe, *M. baccata* is not the initial progenitor of the rootstock varieties. Rather, our results suggest that *M. baccata* did contribute to apple domestication through wild-to-crop gene flow.

Major contribution of *M. baccata* to the cultivated apple genome through crop-to-wild introgression

There is much debate on the contribution of other wild species along the Silk Route to the genetic makeup of the cultivated apple genome. Recently, the Caucasian crab apple has been shown to be an additional contributor to the cultivated apple genome (Bina et al., 2021). Our results demonstrate that interspecific hybridization has been a driving force in the evolution of apple. We confirm that the wild European crabapple *M. sylvestris* has been a major secondary contributor to the diversity of apples. We also show that the Siberian wild apple is a contributor to the diversity of the cultivated apple. Furthermore, we show the substantial gene flow from *M. baccata* to the European apple rootstock and Chinese dessert apple and rootstock. This is not surprising because *M. baccata* is widely used in the high-latitude apple-producing areas of China as a rootstock and breeding resource because of its disease resistance and cold tolerance (Chen et al., 2019; Zhi-Qin, 1999). In the Western hemisphere, *M. baccata* may be used as a parent for

some crosses. Interspecific hybridization has also been observed in date palms (Flowers et al., 2019), grapes (Myles et al., 2011), almonds (Delplancke et al., 2011), and apricots (Groppi et al., 2021; Liu et al., 2019; Q. Zhang et al., 2018). Substantial crop-wild gene flow is mostly due to the self-incompatibility system of fruit trees that favors the selection of the best phenotypes grown from open-pollinated seeds and appears to be a key element in perennial crop evolution (Cornille et al., 2014; Gaut et al., 2015).

Concluding remarks and perspectives

The domestication of fruit trees stands in stark contrast to that of annuals, especially with respect to the extent of wild-crop gene flow. Substantial crop-wild gene flow makes the resolution of relationships between crop and wild fruit trees challenging. In addition, the problem with unravelling the domestication history of fruit trees is magnified by their recent domestication in terms of number of generations and thus shared ancestral polymorphism among the crop and wild fruit trees (Cornille et al., 2014; Gaut et al., 2015; Miller & Gross, 2011). We used genomic data to distinguish between crop-wild shared ancestral polymorphism and recent gene flow and showed that the history of apples is a rare well-documented example of the evolution of a domesticated crop over a long period and involved at least four wild species. Introgression of from wild relatives is a likely source of crop adaptation (Burgarella et al., 2018). The contribution of several species to the cultivated apple genome raises questions about their role in the adaptation of cultivated apples during domestication worldwide. Genome-wide association

showed that traits associated with fruit quality and texture were selected during apple domestication (Duan et al., 2017; Migicovsky, Gardner, Richards, Thomas Chao et al. 2021). However, the genomic landscape of wild-to-crop introgressions has not been investigated on the basis of genome-wide data (Daccord et al., 2017; Duan et al., 2017; Migicovsky & Myles, 2017; Sun et al., 2020; Velasco et al., 2010). Further investigations of the genomic architecture of wild-crop gene flow, and its adaptive role during domestication, are required to understand the role of gene flow during the divergence and adaptation of woody perennials to new environments.

Acknowledgements

We thank the Franco-Chinese Campus France program « Decouverte » 2017, the *Jeunes Talents France-Chine* 2019 program, and the ATIP-Avenir CNRS Inserm grant for funding. We also thank Adrien Falce, Olivier Langella and Benoit Johannet for help and support on the INRAE-Génétique Quantitative et Evolution-Le Moulon lab cluster. We thank the INRAE MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing help and support, in particular Véronique Martin, Eric Montaubon and Valentin Loux. This work was financially supported by the National Key Research and Development Program of China (2018YFD1000101, 2019YFD1000803); the Key Technology R & D and Industrialization Demonstration Science and Technology for the Transformation and Upgrading of Apple Industries in Shaanxi Province-Apple Anvil Combinations Evaluation and Seedlings Industrialization (2020zdzx03-01-04); The

China Apple Research System (CARS 27), Tang Scholar by Cyrus Tang Foundation and Northwest A&F University.

Data Availability

The RNAseq raw data are available at NCBI sequence read archive (SRA) under BioProject accession PRJNA763361 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA763361?reviewer=5i8m9p6b4qi4ucg4kf0qe45v5j&sort_by=-accession&page=2). Vcf used for analyses is available at the Zenodo (<https://zenodo.org/record/5513618>). Scripts are available at https://forgemia.inra.fr/amandine.cornille/rnaseq_apple_evolutionary_history/.

Author Contributions

DZ, AC, and XC conceived and designed the experiments; AC, DZ and MH obtained the funding; XC, NA, LX, YW sampled the material; XC performed the molecular analyses; JM and CJ performed the RNA extraction; XC and AC analyzed the data. The manuscript was written by AC and XC, with essential inputs from other co-authors.

Tables

Table 1. Mean proportions of nine cultivated and wild apple populations assigned to five genetic clusters inferred with ADMIXTURE at $K = 5$ ($N = 115$, 36,200 SNPs). We divided the cultivated apple samples into five populations according to ADMIXTURE inferences and their origin (Europe or China) and use (dessert or rootstock). We assigned only wild populations with membership coefficient > 0.55 to a given genetic group. Thus, samples were partitioned into nine populations, namely, four wild and five cultivated apple populations.

Crop/Wild	Population	N	Cluster				
			Green	Blue	Orange	Dark blue	Red
Pure wild	North American	5	0.97	0.02	0.01	0.00	0.00
	<i>M. baccata</i> / <i>M. hupehensis</i>	16	0.00	0.86	0.01	0.04	0.10
	<i>M. sylvestris</i>	7	0.00	0.00	0.85	0.00	0.14
	<i>M. sieversii</i>	29	0.00	0.00	0.04	0.92	0.04
Crop	European dessert (<i>M.</i>	26	0.00	0.01	0.09	0.09	0.81

	<i>domestica</i>)						
	European rootstock	16	0.00	0.40	0.01	0.45	0.13
	Chinese dessert	5	0.00	0.47	0.00	0.36	0.17
	Chinese ornamental	4	0.00	0.61	0.01	0.30	0.08
	Chinese rootstock	7	0.03	0.27	0.03	0.15	0.52
Total		115					

623 N: number of individuals.

624

Figures

Figure 1. Genetic structure, variation, differentiation, and relationships of wild and cultivated European and Chinese apples. a. Population structure inferred with ADMIXTURE at $K = 5$ ($N = 115$, 36,200 unlinked synonymous SNPs); each individual is represented by a vertical bar partitioned into K segments that represent the proportions of ancestry of its genome in K clusters. Colors represent the inferred ancestry from K ancestral genetic clusters. b. Neighbor net tree depicting the relationships among the wild and cultivated apple populations identified with ADMIXTURE for $K = 5$. Colors correspond to the genetic groups inferred at $K = 5$, and admixed samples (*i.e.*, individuals with membership coefficient < 0.55 to any given gene pool) are in grey. c. Principal component analysis (PCA) representing the genetic variation among wild and cultivated apples, with the total variance explained by the first two components (29.72% and 16.11%, respectively). d. Pairwise genetic differentiation (F_{ST}) among populations identified with ADMIXTURE at $K = 5$ (excluding wild individuals with membership coefficient < 0.55 to a given gene pool).

Figure 2. Gene flow inferred with TreeMix ($N = 114$, 36,200 SNPs; Chinese ornamentals were removed because of low number of individuals, $N = 4$) among the wild and cultivated apple populations detected with ADMIXTURE at $K = 5$. The North American wild apples were used as the outgroup. The best number of migration events was six (Figure S10). Arrows of different colors represent migration events.

645

646 **Figure 3. Domestication histories of European and Chinese dessert and rootstock apples**
 647 **inferred using the approximate Bayesian computation random forest framework (ABC-RF,**
 648 **36,200 SNPs)** a. Most likely scenario of domestication of the European rootstock inferred with
 649 ABC-RF set 1 ($N = 98$). b. Most likely scenario of domestication of the European and Chinese
 650 apples (set 2, $N = 111$). For each scenario, associated statistics (posterior probability, prior error
 651 rates, and percentage of votes) and parameter estimates for the final model in set 2. The grey
 652 arrow indicates that the Chinese rootstock and dessert cultivars may have originated from *M.*
 653 *sieversii* or *M. domestica*.

654

References

- Aguirre-Liguori, J. A., Gaut, B. S., Jaramillo-Correa, J. P., Tenaillon, M. I., Montes-Hernández, S., García-Oliva, F., ... Eguiarte, L. E. (2019). Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). *Molecular Ecology*, 28(11), 2814–2830. doi: 10.1111/mec.15098
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664. doi: 10.1101/gr.094052.109
- Allaby, R. G., Fuller, D. Q., & Brown, T. A. (2008). The genetic expectations of a protracted model for the origins of domesticated crops. *PNAS*, 105(37), 13982–13986. doi: 10.1073/pnas.0803780105
- Aravanopoulos, F. A., Ganopoulos, I., & Tsaftaris, A. (2015). Chapter Four—Population and Conservation Genomics in Forest and Fruit Trees. In C. Plomion & A.-F. Adam-Blondon (Eds.), *Advances in Botanical Research* (pp. 125–155). Academic Press. doi: 10.1016/bs.abr.2015.04.001
- Besnard, G., Terral, J.-F., & Cornille, A. (2018). On the origins and domestication of the olive: A review and perspectives. *Annals of Botany*, 121(3), 385–403. doi: 10.1093/aob/mcx145
- Bina, H., Yousefzadeh, H., Venon, A., Remoué, C., Rousselet, A., Falque, M., ... Cornille, A. (2021). Evidence of an additional center of apple domestication in Iran, with

674 contributions from the Caucasian crab apple. *BioRxiv*, 2021.03.28.437401. doi:
675 10.1101/2021.03.28.437401

676 Brandenburg, J.-T., Mary-Huard, T., Rigaille, G., Hearne, S. J., Corti, H., Joets, J., ... Tenailon,
677 M. I. (2017). Independent introductions and admixtures have contributed to adaptation of
678 European maize and its American counterparts. *PLOS Genetics*, 13(3), e1006666. doi:
679 10.1371/journal.pgen.1006666

680 Breiman, L. (2001). Random Forests. *Machine Language*, 45(1), 5–32. doi:
681 10.1023/A:1010933404324

682 Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., ... Berthouly-
683 Salazar, C. (2019). Adaptive Introgression: An Untapped Evolutionary Mechanism for
684 Crop Adaptation. *Frontiers in Plant Science*, 0. doi: 10.3389/fpls.2019.00004

685 Burgarella, C., Barnaud, A., Kane, N. A., Jankowsky, F., Scarcelli, N., Billot, C., ... Berthouly-
686 Salazar, C. (2018). Adaptive introgression: An untapped evolutionary mechanism for crop
687 adaptation. *BioRxiv*. doi: 10.1101/379966

688 Cao, K., Zhou, Z., Wang, Q., Guo, J., Zhao, P., Zhu, G., ... Wang, L. (2016). Genome-wide
689 association study of 12 agronomic traits in peach. *Nature Communications*, 7(1), 13246.
690 doi: 10.1038/ncomms13246

691 Chagné, D., Crowhurst, R. N., Pindo, M., Thrimawithana, A., Deng, C., Ireland, H., ... Velasco,
692 R. (2014). The Draft Genome Sequence of European Pear (*Pyrus communis* L. ‘Bartlett’).
693 *PLOS ONE*, 9(4), e92644. doi: 10.1371/journal.pone.0092644

- 694 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015).
695 Second-generation PLINK: Rising to the challenge of larger and richer datasets.
696 *GigaScience*, 4. doi: 10.1186/s13742-015-0047-8
- 697 Chen, X., Li, S., Zhang, D., Han, M., Jin, X., Zhao, C., ... An, N. (2019). Sequencing of a Wild
698 Apple (*Malus baccata*) Genome Unravels the Differences Between Cultivated and Wild
699 Apple Species Regarding Disease Resistance and Cold Tolerance. *G3: Genes, Genomes,*
700 *Genetics*, 9(7), 2051–2060. doi: 10.1534/g3.119.400245
- 701 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012).
702 A program for annotating and predicting the effects of single nucleotide polymorphisms,
703 SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w¹¹¹⁸□; iso-2; iso-3. *Fly*,
704 6(2), 80–92. doi: 10.4161/fly.19695
- 705 Cornille, A., Gladieux, P., & Giraud, T. (2013). Crop-to-wild gene flow and spatial genetic
706 structure in the wild closest relatives of the cultivated apple. *Evolutionary Applications*,
707 6(5), 737–748. doi: 10/f44gn5
- 708 Cornille, A., Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., ...
709 Giraud, T. (2012). New insight into the history of domesticated apple: Secondary
710 contribution of the European wild apple to the genome of cultivated varieties. *PLoS*
711 *Genet*, 8(5), e1002703.
- 712 Cornille, A., Feurtey, A., Gélín, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud, T.
713 (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: A

basis for conservation and breeding programs in apples. *Evolutionary Applications*, 8(4), 373–384. doi: 10.1111/eva.12250

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

Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., ... Roldán-Ruiz, I. (2019). A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science*, 24(8), 770–782. doi: <https://doi.org/10.1016/j.tplants.2019.05.007>

Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: An R package for approximate Bayesian computation (ABC): *R package: abc. Methods in Ecology and Evolution*, 3(3), 475–479. doi: 10.1111/j.2041-210X.2011.00179.x

Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., ... Bucher, E. (2017). High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nature Genetics*, 49(7), 1099–1106. doi: 10.1038/ng.3886

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. doi: 10.1093/bioinformatics/btr330

Delplancke, M., Alvarez, N., Espíndola, A., Joly, H., Benoit, L., Brouck, E., & Arrigo, N. (2011). Gene flow among wild and domesticated almond species: Insights from chloroplast and nuclear markers. *Evolutionary Applications*, 5(4), 317–329.

- Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for Ecologists. *Journal of Statistical Software*, 022(i04). Retrieved from <https://ideas.repec.org/a/jss/jstsof/v022i04.html>
- Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., ... Chen, X. (2017). Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications*, 8(1), 249. doi: 10.1038/s41467-017-00336-7
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture between Closely Related Populations. *Molecular Biology and Evolution*, 28(8), 2239–2252. doi: 10.1093/molbev/msr048
- Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, 36(8), 2592–2594. doi: 10.1093/bioinformatics/btz966
- Estoup, A. A., Raynal, L., Verdu, P., & Marin, J.-M. (2018). Model choice using Approximate Bayesian Computation and Random Forests: Analyses based on model grouping to make inferences about the genetic history of Pygmy human populations. *Journal de La Societe Française de Statistique*, 159(3), 167–190.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust Demographic Inference from Genomic and SNP Data. *PLOS Genetics*, 9(10), e1003905. doi: 10.1371/journal.pgen.1003905

- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- Feurtey, A., Cornille, A., Shykoff, J. A., Snirc, A., & Giraud, T. (2017). Crop to wild gene flow and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation strategy of the wild apple in Europe. *Evolutionary Applications*, 10(2), 180–188. PMC (PMC5253423). doi: 10.1111/eva.12441
- Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., ... Purugganan, M. D. (2019). Cross-species hybridization and the origin of North African date palms. *Proceedings of the National Academy of Sciences*, 116(5), 1651–1658. doi: 10.1073/pnas.1817453116
- Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17(1), 27–32. doi: 10.1111/1755-0998.12509
- Gao, Y., Liu, F., Wang, K., Wang, D., Gong, X., Liu, L., ... Volk, G. M. (2015). Genetic diversity of *Malus* cultivars and wild relatives in the Chinese National Repository of Apple Germplasm Resources. *Tree Genetics & Genomes*, 11(5), 106. doi: 10.1007/s11295-015-0913-7
- Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics*, 31(12), 709–719. doi: 10.1016/j.tig.2015.10.002

- 772 Glémin, S., & Bataillon, T. (2009). A comparative view of the evolution of grasses under
773 domestication. *New Phytologist*, 183(2), 273–290. doi: [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-8137.2009.02884.x)
774 8137.2009.02884.x
- 775 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Pääbo, S. (2010).
776 A Draft Sequence of the Neandertal Genome. *Science*, 328(5979), 710–722. doi:
777 10.1126/science.1188021
- 778 Groppi, A., Liu, S., Cornille, A., Decroocq, S., & Decroocq, D. (2021). Population genomics of
779 apricots unravels domestication history and adaptive events. *Nature Communications*, 12,
780 3956.
- 781 Gygax, M., Gianfranceschi, L., Liebhard, R., Kellerhals, M., Gessler, C., & Patocchi, A. (2004).
782 Molecular markers linked to the apple scab resistance gene Vbj derived from *Malus*
783 *baccata* jackii. *Theoretical and Applied Genetics*, 109(8), 1702–1709. doi:
784 10.1007/s00122-004-1803-9
- 785 Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple.
786 *Trends in Genetics*, 18(8), 426–430. doi: 10.1016/S0168-9525(02)02689-6
- 787 Harrison, N., & Harrison, R. J. (2011). On the evolutionary history of the domesticated apple.
788 *Nature Genetics*, 43(11), 1043–1044. doi: 10.1038/ng.935
- 789 Hazzouri, K. M., Flowers, J. M., Visser, H. J., Khierallah, H. S. M., Rosas, U., Pham, G. M., ...
790 Purugganan, M. D. (2015). Whole genome re-sequencing of date palms yields insights

into diversification of a fruit tree crop. *Nature Communications*, 6(1), 8824. doi: 10.1038/ncomms9824

Hu, C. G., Honda, C., Kita, M., Zhang, Z., Tsuda, T., & Moriguchi, T. (2002). A simple protocol for RNA isolation from fruit trees containing high levels of polysaccharides and polyphenol compounds. *Plant Molecular Biology Reporter*, 20(1), 69–69. doi: 10.1007/BF02801935

Huang, J., Zhang, C., Zhao, X., Fei, Z., Wan, K., Zhang, Z., ... Li, X. (2016). The Jujube Genome Provides Insights into Genome Evolution and the Domestication of Sweetness/Acidity Taste in Fruit Trees. *PLOS Genetics*, 12(12), e1006433. doi: 10.1371/journal.pgen.1006433

Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution*, 23(2), 254–267. doi: 10.1093/molbev/msj030

Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, 106(4), 625–632.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4), 357–360. doi: 10.1038/nmeth.3317

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure

810 inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. doi: 10.1111/1755-
811 0998.12387

812 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
813 *ArXiv:1303.3997 [q-Bio]*. Retrieved from <http://arxiv.org/abs/1303.3997>

814 Lipson, M. (2020). Applying f4-statistics and admixture graphs: Theory and examples.
815 *Molecular Ecology Resources*, 20(6), 1658–1667. doi: 10.1111/1755-0998.13230

816 Lipson, M., Loh, P.-R., Levin, A., Reich, D., Patterson, N., & Berger, B. (2013). Efficient
817 Moment-Based Inference of Admixture Parameters and Sources of Gene Flow. *Molecular*
818 *Biology and Evolution*, 30(8), 1788–1802. doi: 10.1093/molbev/mst099

819 Liu, S., Cornille, A., Decroocq, S., Tricon, D., Chague, A., Eyquard, J., ... Decroocq, V. (2019).
820 The complex evolutionary history of apricots: Species divergence, gene flow and
821 multiple domestication events. In *Molecular Ecology*. doi: 10.1111/mec.15296

822 Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite—Fast D-statistics and related
823 admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595. doi:
824 10.1111/1755-0998.13265

825 Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., & Chen, W.-M. (2010).
826 Robust relationship inference in genome-wide association studies. *Bioinformatics*
827 *(Oxford, England)*, 26(22), 2867–2873. doi: 10.1093/bioinformatics/btq559

828 Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez G, J., Buckler, E., & Doebley, J. (2002).
829 A single domestication for maize shown by multilocus microsatellite genotyping. *PNAS*,
830 99(9), 6080–6084. doi: 10.1073/pnas.052125199

831 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo,
832 M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing
833 next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. doi:
834 10.1101/gr.107524.110

835 Meyer, R. S., Duval, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication:
836 An historical review and quantitative analysis of 203 global food crops. *New Phytologist*,
837 196(1), 29–48. doi: 10.1111/j.1469-8137.2012.04253.x

838 Michael, T. P., & VanBuren, R. (2015). Progress, challenges and the future of crop genomes.
839 *Current Opinion in Plant Biology*, 24, 71–81. doi: 10.1016/j.pbi.2015.02.002

840 Migicovsky, Z., Gardner, K. M., Richards, C., Chao, C. T., Schwaninger, H. R., Fazio, G., ...
841 Myles, S. (2021). Genomic consequences of apple improvement. *Horticulture Research*,
842 8(1), 1–13. doi: 10.1038/s41438-020-00441-7

843 Migicovsky, Z., & Myles, S. (2017). Exploiting Wild Relatives for Genomics-assisted Breeding
844 of Perennial Crops. *Frontiers in Plant Science*, 8, 460. doi: 10.3389/fpls.2017.00460

845 Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication.
846 *American Journal of Botany*, 98(9), 1389–1414. doi: 10.3732/ajb.1000522

- 847 Morgan, J., & Richards, A. (1993). *The Book of Apples*. Brogdale Horticultural Trust, Ebury
848 Press.
- 849 Morgan, J., & Richards, A. (2003). *The New Book of Apples*. Brogdale Horticultural Trust, Ebury
850 Press.
- 851 Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., ... Buckler, E. S.
852 (2011). Genetic structure and domestication history of the grape. *PNAS*, 108(9), 3530–
853 3535. doi: 10.1073/pnas.1009363108
- 854 Neale, D. B., Martínez-García, P. J., De La Torre, A. R., Montanari, S., & Wei, X.-X. (2017).
855 Novel Insights into Tree Biology and Genome Evolution as Revealed Through Genomics.
856 *Annual Review of Plant Biology*, 68(1), 457–483. doi: 10.1146/annurev-arplant-042916-
857 041049
- 858 Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of
859 restriction endonucleases. *Proceedings of the National Academy of Sciences of the United*
860 *States of America*, 76(10), 5269–5273.
- 861 Nei, Masatoshi. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of*
862 *the National Academy of Sciences*, 70(12), 3321–3323. doi: 10.1073/pnas.70.12.3321
- 863 Nei, Masatoshi. (1978). Estimation of average heterozygosity and genetic distance from a small
864 number of individuals. *Genetics*, 89(3), 583–590.

Peace, C. P., Bianco, L., Troggio, M., van de Weg, E., Howard, N. P., Cornille, A., ...
Vanderzande, S. (2019). Apple whole genome sequences: Recent advances and new
prospects. *Horticulture Research*, 6(1), 59. doi: 10.1038/s41438-019-0141-7

Picard toolkit. (2019). In *Broad Institute, GitHub repository*. Broad Institute. Retrieved from
<http://broadinstitute.github.io/picard/>

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from
Genome-Wide Allele Frequency Data. *PLOS Genetics*, 8(11), e1002967. doi:
10.1371/journal.pgen.1002967

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using
Multilocus Genotype Data. *Genetics*, 155(2), 945–959. doi: 10.1093/genetics/155.2.945

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable
ABC model choice via random forests. *Bioinformatics*, 32(6), 859–866. doi:
10.1093/bioinformatics/btv684

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population
structure when sampling is uneven: Subsampling and new estimators alleviate the
problem. *Molecular Ecology Resources*, 16(3), 608–627. doi: 10.1111/1755-0998.12512

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P.
C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based
Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559–575.

- 884 Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication.
885 *Current Biology*, 29(14), R705–R714. doi: 10.1016/j.cub.2019.05.053
- 886 Robinson, J. P., Harris, S. A., & Juniper, B. E. (2001). Taxonomy of the genus *Malus* Mill.
887 (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant*
888 *Systematics and Evolution*, 226(1), 35–58. doi: 10.1007/s006060170072
- 889 Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for
890 paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*,
891 28(21), 4737–4754. doi: 10.1111/mec.15253
- 892 Spengler, R. N. (2019). Origins of the Apple: The Role of Megafaunal Mutualism in the
893 Domestication of *Malus* and Rosaceous Trees. *Frontiers in Plant Science*, 10. doi:
894 10.3389/fpls.2019.00617
- 895 Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., ... Fei, Z. (2020). Phased
896 diploid genome assemblies and pan-genomes provide insights into the genetic history of
897 apple domestication. *Nature Genetics*, 52(12), 1423–1432. doi: 10.1038/s41588-020-
898 00723-9
- 899 Teh, B. T., Lim, K., Yong, C. H., Ng, C. C. Y., Rao, S. R., Rajasegaran, V., ... Tan, P. (2017). The
900 draft genome of tropical fruit durian (*Durio zibethinus*). *Nature Genetics*, 49(11), 1633–
901 1641. doi: 10.1038/ng.3972
- 902 Tellier, A., Pfaffelhuber, P., Haubold, B., Naduvilezhath, L., Rose, L. E., Städler, T., ... Metzler,
903 D. (2011). Estimating Parameters of Speciation Models Based on Refined Summaries of

904 the Joint Site-Frequency Spectrum. *PLOS ONE*, 6(5), e18155. doi:
905 10.1371/journal.pone.0018155

906 Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., ... Durel, C.-E.
907 (2016). Analysis of the genetic diversity and structure across a wide range of germplasm
908 reveals prominent gene flow in apple at the European level. *BMC Plant Biology*, 16(1),
909 1–20. doi: 10.1186/s12870-016-0818-0

910 Vavilov, N. I. (1926). Studies on the origin of cultivated plants. *Trudy Byuro. Prikl. Bot.*, 16,
911 139–245.

912 Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., ... Pruss, D.
913 (2010). The genome of the domesticated apple (*Malus x domestica* Borkh.). *Nat Genet*,
914 42. doi: 10.1038/ng.654

915 Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., ... Rokhsar, D. S. (2013).
916 The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of
917 genetic diversity, domestication and genome evolution. *Nature Genetics*, 45(5), 487–494.
918 doi: 10.1038/ng.2586

919 Volk, G. M., Chao, C. T., Norelli, J., Brown, S. K., Fazio, G., Peace, C., ... Bretting, P. (2015).
920 The vulnerability of US apple (*Malus*) genetic resources. *Genetic Resources and Crop*
921 *Evolution*, 62(5), 765–794. doi: 10.1007/s10722-014-0194-2

922 Volk, G. M., Cornille, A., Durel, C.-E., & Gutierrez, B. (2021). Botany, Taxonomy, and Origins
 923 of the Apple. In S. S. Korban (Ed.), *The Apple Genome* (pp. 19–32). Cham: Springer
 924 International Publishing. doi: 10.1007/978-3-030-74682-7_2

925 Wakeley, J., & Hey, J. (1997). Estimating Ancestral Population Parameters. *Genetics*, 145(3),
 926 847–855.

927 Wang, Y., Li, W., Xu, X., Qiu, C., Wu, T., Wei, Q., ... Han, Z. (2019). Progress of Apple
 928 Rootstock Breeding and Its Use. *Horticultural Plant Journal*, 5(5), 183–191. doi:
 929 10.1016/j.hpj.2019.06.001

930 Warschefskey, E. J., Klein, L. L., Frank, M. H., Chitwood, D. H., Londo, J. P., von Wettberg, E. J.
 931 B., & Miller, A. J. (2016). Rootstocks: Diversity, Domestication, and Impacts on Shoot
 932 Phenotypes. *Trends in Plant Science*, 21(5), 418–437. doi: 10.1016/j.tplants.2015.11.008

933 Wedger, M. J., Schumann, A. C., & Gross, B. L. (2021). Candidate genes and signatures of
 934 directional selection on fruit quality traits during apple domestication. *American Journal*
 935 *of Botany*, 108(4), 616–627. doi: 10.1002/ajb2.1636

936 Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
 937 doi: 10.1007/978-0-387-98141-3

938 Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., ... Zhang, S. (2018). Diversification and
 939 independent domestication of Asian and European pears. *Genome Biology*, 19(1), 77. doi:
 940 10.1186/s13059-018-1452-y

- Yang, H., Duan, K., & Zhang, W. (2008). Biology and physiology of *Malus hupehensis* for the apogamic plant resource. *Acta Horticulturae*, 769, 441–447.
- Yunong, L. (1999). An Investigation and Studies on the Origin and Evolution of *Malus domestica* Borkh. In the World. *Acta Horticulturae Sinica*, 4.
- Zhang, F., & Batley, J. (2020). Exploring the application of wild species for crop improvement in a changing climate. *Biotic Interactions. AGRI 2019*, 56, 218–222. doi: 10.1016/j.pbi.2019.12.013
- Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B.-H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, 10(1), 5–24. doi: <https://doi.org/10.1111/eva.12434>
- Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., ... Cheng, T. (2018). The genetic architecture of floral traits in the woody plant *Prunus mume*. *Nature Communications*, 9(1), 1702. doi: 10.1038/s41467-018-04093-z
- Zhi-Qin, Z. (1999). The apple genetic resources in China: The wild species and their distributions, informative characteristics and utilisation. *Genetic Resources and Crop Evolution*, 46(6), 599–609. doi: 10.1023/A:1008747709534
- Zohary, D., & Spiegel-Roy, P. (1975). Beginnings of fruit growing in the Old World. *Science*, 319–327.

a.

b.

c.

d.

a.

ABC-RF set 1: Domestication history of the European cultivated apple

b.

ABC-RF set 2: History of domestication of the Chinese dessert and rootstock