



HAL
open science

The East Asian wild apples, *Malus baccata* (L.) Borkh and *Malus hupehensis* (Pamp.) Rehder., are additional contributors to the genomes of cultivated European and Chinese varieties

Xilong Chen, Amandine Cornille, Na An, Libo Xing, Juanjuan Ma, Caiping Zhao,
Yibin Wang, Mingyu Han, Dong Zhang

► To cite this version:

Xilong Chen, Amandine Cornille, Na An, Libo Xing, Juanjuan Ma, et al.. The East Asian wild apples, *Malus baccata* (L.) Borkh and *Malus hupehensis* (Pamp.) Rehder., are additional contributors to the genomes of cultivated European and Chinese varieties. *Molecular Ecology*, In press, <10.1111/mec.16485>. <hal-03803831>

HAL Id: hal-03803831

<https://hal.science/hal-03803831v1>

Submitted on 6 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

1 The East Asian wild apples, *Malus baccata* (L.) Borkh and *Malus hupehensis* (Pamp.) Rehder.,
2 are additional contributors to the genomes of cultivated European and Chinese varieties

3

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

6

7 1. College of Horticulture, Yangling Sub-Center of National Center for Apple Improvement,
8 Northwest A&F University, Yangling 712100, Shaanxi, China

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

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

12

13 *co-first authors

14 #co-directed the work

15 Corresponding authors:

16 Amandine Cornille, amandine.cornille@cnrs.fr

17 Dong Zhang, afant@nwafu.edu.cn

18

19 **Abstract**

20 Domestication process in long-lived plant perennials differs dramatically from that of annuals,
21 with a huge extent of genetic exchanges between crop and wild populations. Though apple is
22 a major fruit crop grown worldwide, the contribution of wild apple species to the genetic
23 makeup of the cultivated apple genome remains a topic of intense investigations. We used
24 population genomics approaches to investigate the contributions of several wild apple species
25 to European and Chinese rootstock and dessert genomes, with a focus on the extent of wild-
26 crop gene flow. Population genetic structure inferences revealed that the East Asian wild
27 apples, *M. baccata* (L.) Borkh and *M. hupehensis* (Pamp.), form a single panmictic group, and
28 that the European dessert and rootstock apples form a specific gene pool whereas the Chinese
29 dessert and rootstock apples were a mixture of three wild gene pools, suggesting different
30 evolutionary histories of European and Chinese apple varieties. Coalescent-based inferences
31 and gene flow estimates indicated that *M. baccata*-*M. hupehensis* contributed to the genome of
32 both European and Chinese cultivated apples through wild-to-crop introgressions, and not as
33 an initial contributor as previously supposed. We also confirmed the contribution through wild-
34 to-crop introgressions of *M. sylvestris* Mill. to the cultivated apple genome. Apple tree
35 domestication is therefore one example in woody perennials that involved gene flow from
36 several wild species from multiple geographic area. This study provides an overview of the
37 complex protracted process of domestication in long-lived plant perennials, and is a starting
38 point for apple breeding programs.

39

40

41 **Keywords:** Introgression, perennials, clonal, fruit tree, apple, domestication, rootstock, gene
42 flow.

43

44 **Introduction**

45 It is crucial to understand domestication to unravel the evolutionary processes that shape the
46 divergence of populations (Gaut et al., 2015; Glémin & Bataillon, 2009; Meyer et al., 2012).
47 Several studies have investigated the evolutionary history of fruit crop perennials by using
48 genome-wide data (Aravanopoulos et al., 2015; Cao et al., 2016; Chagné et al., 2014; Daccord
49 et al., 2017; N. Harrison & Harrison, 2011; T. Harrison & Winfree, 2015; Huang et al., 2016;
50 Michael & VanBuren, 2015; Teh et al., 2017; Verde et al., 2013), however, documentation of
51 the domestication of fruit tree crops still lags behind that of annual crop plants (Burgarella et
52 al., 2019; Gaut et al., 2015; Meyer et al., 2012; Miller & Gross, 2011; Neale et al., 2017).
53 Differences in life-history traits have probably led to marked differences in the mode and speed
54 of evolution between trees and annuals (Gaut et al., 2015), particularly the extent of gene flow
55 during domestication. Documentation of the domestication of fruit tree crops is important not
56 only academically but also for food production. Because of long-term cultivation, fruit trees
57 are susceptible to ever-changing environments due to climate changes and pathogen outbreaks.
58 Sustainable fruit production is now believed to highly rely on untapped wild genetic diversity
59 (Zhang & Batley, 2020; Zhang et al., 2017).

60 Apple is an iconic tree and major fruit crop grown worldwide. It is also a model species
61 for understanding the role of gene flow during the domestication of clonally propagated
62 perennial crops (Cornille et al., 2014, 2019; Peace et al., 2019; Spengler, 2019). The
63 domestication of apple started in Central Asia, from the local Central Asian wild apple *Malus*
64 *sieversii* (Ledeb.) M. Roem. Then, apple cultivation spread towards the West, Europe and
65 beyond, and the cultivation was assisted by the possibility of propagating interesting genotypes
66 by grafting (Cornille, Gladioux, Smulders, Roldán-Ruiz, Laurens, Le Cam, Nersesyan, Clavel,
67 Olonova, Feugey, et al., 2012). Thus, grafting has been an important part of apple evolution
68 and breeding (Wang et al., 2019). With the arrival of the cultivated apple in Europe about 1,500

69 years ago, the local crab apple *Malus sylvestris* Mill. contributed substantially to the genome
70 of some dessert and cider varieties through wild-to-crop introgressions (Cornille, Gladieux,
71 Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012). Population genetic approaches in
72 combination with microsatellite markers have demonstrated that wild-to-crop gene flow has
73 played a major role in the evolutionary history of the cultivated apple in Europe (Cornille,
74 Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012; Duan et al., 2017; Sun et al.,
75 2020). However, while the history of the dessert and cider apples is now well documented in
76 the western part of Eurasia (Bina et al., 2021; Cornille et al., 2014; Migicovsky et al., 2021;
77 Peace et al., 2019; Spengler, 2019), the domestication history of apples in the East remains
78 unclear. In addition, the history of the apple rootstock that allowed the spread of apple across
79 the world needs to be elucidated. Understanding the origin of the apple rootstock can also
80 provide insights into the role of crop-wild gene flow during apple domestication, as several
81 rootstocks are supposed to be the result of wild-crop hybridization and are often dwarf forms
82 of local wild species (Volk et al., 2021; Wang et al., 2019).

83 In China, apple cultivation has a tangled and complex history that spans a millennium
84 and involves several wild and cultivated apple species, used for dessert or as rootstocks, that
85 were crossed by ancient communities or bred recently. For more than 2,000 years, ancient
86 Asian communities cultivated *Malus × domestica* ssp. *chinensis* Li Y.N. (called Mianpingguo,
87 which means soft apple) (Gao et al., 2015; Li, 1999). *Malus × domestica* ssp. *chinensis* is
88 supposed to originate from *M. sieversii* located in the northwestern part of China (Xinjiang
89 Province), close to the borders of Kazakhstan and Kyrgyzstan where *M. sieversii* also grows.
90 Then, the cultivated species was propagated along the Silk Routes across China where several
91 local cultivars once existed (Gao et al., 2015). However, from the 18th century, most of the
92 local cultivars disappeared, and, nowadays, only a few Chinese landraces remain, namely,
93 Mianpingguo, Shaguo, Huahong, and Caiping. These four Chinese cultivars are supposed to

94 belong to two species, *M. domestica* subsp. *chinensis* and *Malus asiatica*. The two species are
95 supposed to be closely related, or even be the same species (Li, 1999), but no genetic data has
96 confirmed that hypothesis. Some hybrids of *M. domestica* ssp. *chinensis* and *M. asiatica* are
97 cultivated in China. In addition, some modern apple varieties from Western countries, such as
98 Golden Delicious, Gala, and Jonagold, are also cultivated in China, and they were introduced
99 from Europe and the United States less than three hundred years ago. A few other cultivated
100 Chinese apple species are traditionally used as rootstocks and for ornamental purposes, and
101 some of them originated from several hybridization events (Gao et al., 2015). *Malus prunifolia*
102 (Chinese crabapple called Qiuzi) and *Malus robusta* (i.e., *Malus baccata* × *M. prunifolia* cross
103 called Balenghaitang) are used as rootstocks. *Malus micromalus*, *Malus halliana*, and *Malus*
104 *spectabilis* are used as ornamental species. These rootstocks and ornamental species cannot be
105 found in natural habitats in China, so they are classified as cultivated species (Li, 1999). The
106 Chinese people have also used local wild apple species for millennia. *Malus baccata*, the
107 Siberian wild apple, is the main genetic resource (especially as a rootstock) for apple breeding
108 programs in China, because of its excellent resistance to cold stress and apple scab (Chen et
109 al., 2019b; Gyax et al., 2004; Volk et al., 2021). *Malus hupehensis*, the tea crabapple or Hupeh
110 crab, is an indigenous wild crabapple from China, and is an apomictic species. Its seedling
111 rootstocks can be propagated directly through seedlings, instead of using cuttings or layering
112 propagation (Wang et al., 2019; Yang et al., 2008). Because of their ease of propagation
113 through apomixis and their cold resistance, *M. baccata* and *M. hupehensis* are supposed to be
114 the progenitors of the Chinese rootstocks. However, such contributions remain matter of debate
115 because of the lack of population genetic studies. The two species are sympatric in north and
116 southwest China (Li, 1999), and form a sister group in phylogenetic trees (Robinson, Harris,
117 & Juniper, 2001; Zhang et al., 2018). These phylogenetic studies only used one or two
118 representatives of each species from narrow geographical distributions; additional samples are

119 now needed. The contributions of the Chinese cultivated varieties, as initial progenitors or/and
120 by recent wild-to-crop gene flow, therefore remain to be elucidated. China also introduced
121 European-bred dwarf rootstocks in the 19th century, but most apple production areas in China
122 still use local cultivated (*M. prunifolia* and *M. robusta*) and wild (*M. hupehensis* and *M.*
123 *baccata*) apples as rootstocks because of their good adaptability to cold environment (Wang et
124 al., 2019).

125 Apple domestication and breeding in China have therefore probably involved crop-crop
126 and wild-crop gene flow and contributions of several wild species. However, all available
127 information on apple domestication in China is derived from historical records and traditional
128 knowledge, or phylogenetic studies. Only Duan *et al.* (2017) showed that *M. asiatica* and *M.*
129 *prunifolia* originated from crosses between *M. sieversii* and *M. baccata*, but this result remains
130 anecdotic. It is therefore still unclear whether cultivated Chinese dessert and rootstock apples
131 originated from *M. sieversii* and/or *M. baccata* or/and *M. hupehensis*, or both originated from
132 *M. sieversii* followed by recurrent wild-to-crop gene flow from *M. baccata* or/and *M.*
133 *hupehensis*. Besides, the genetic relationships among wild and cultivated European and
134 Chinese apples have not yet been elucidated, and it is unclear whether substantial crop-to-wild
135 gene flow occurred during apple domestication in China, as observed in Europe and the
136 Caucasus (Bina et al., 2021; Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al.,
137 2012).

138 Here, we investigated the history of domestication of European and Chinese dessert and
139 rootstock apples, with a focus on the extent of crop-wild gene flow during domestication. We
140 combined previously published sequenced genomes (Duan et al., 2017) with a new dataset that
141 included wild and cultivated Chinese apples sequenced for their mRNA. Using single-
142 nucleotide polymorphisms (SNPs) called from several wild apple species, including the wild
143 apple relatives of *M. domestica* (*i.e.*, *M. sieversii* and *M. sylvestris*), *M. baccata*, *M. hupehensis*,

144 and other ancestral species from the genus *Malus*, as well as European and Chinese apple
145 cultivars, we aimed to answer the following questions: (1) What is the population structure and
146 differentiation of wild and cultivated apples from Eurasia? (2) Did gene flow occur between
147 wild and cultivated apples during domestication? (3) What is the domestication history of
148 Chinese and European dessert apple and rootstock cultivars and were they independently
149 domesticated from different wild species?

150

151 **Materials and Methods**

152 **Plant samples**

153 For this study, fruit flesh samples of 71 *Malus* accessions, representing 48 wild and 23
154 cultivated apple samples, were collected from repositories in China (Table S1). Additionally,
155 short-read DNA sequences of 97 *Malus* accessions (Duan et al., 2017), namely, 53 wild apple
156 samples and 44 apple cultivars, were downloaded (SRA code accession: SRP075497, Table
157 S2). The two datasets, which included 168 accessions with 66 apple cultivars (7 historical
158 Chinese cultivars, 30 cultivars from Europe or Western countries, 25 rootstocks, and 4
159 ornamental cultivars) and 102 wild apples (43 *M. sieversii*, 18 *M. baccata*, 11 *M. sylvestris*,
160 and 30 other *Malus* accessions), were merged for subsequent analyses (Table S1). Any wild
161 species that naturally occurs in forest and are not used for human resource purposes were
162 referred to as wild. Chinese apple varieties were considered as rootstocks following Li et al.
163 (1999) classification and European apple varieties were assigned to rootstock and dessert
164 categories following Morgan & Richards (2003).

165

166 **RNA extraction and library construction and sequencing**

167 The fruit flesh of the 71 collected samples was immediately transferred to liquid nitrogen and
168 stored at -80°C until RNA extraction. The RNA was extracted using the SDS–phenol method

169 (Hu et al., 2002). The RNA quality was checked using an agarose gel, and RNA concentrations
170 were estimated with a NanoDrop 1000 spectrophotometer (NanoDrop Technologies,
171 Wilmington, DE, USA). After RNA extraction and DNase I treatment, mRNA was isolated
172 from the total RNA by using magnetic oligo (dT) beads. The mRNA was mixed with
173 fragmentation buffer and cleaved into short fragments and used as templates for cDNA
174 synthesis. The short fragments were purified, resolved with EB buffer for end reparation and
175 single adenine nucleotide addition and connected with adapters. After agarose gel
176 electrophoresis, suitable fragments were selected as templates for polymerase chain reaction
177 (PCR) amplification. During the quality control steps, an Agilent 2100 Bioanalyzer and ABI
178 StepOnePlus Real-Time PCR system (Agilent, US) were used for evaluating the quantity and
179 quality of the sample libraries. Finally, the constructed libraries were sequenced on an Illumina
180 HiSeq 2000 system (BGI, Shenzhen, China).

181

182 **Read mapping and SNP calling and filtering**

183 Two different pipelines were used for the sequenced DNA and RNA reads (Figure S1). For the
184 sequenced RNA accessions, reads were processed as follows: SNP calling was performed on
185 the basis of the transcript sequence data with GATK v3.5 calling variants pipeline for RNAseq
186 (<https://gatk.broadinstitute.org/hc/en-us/articles/360035531192?id=3891>) (McKenna et al.,
187 2010). Fastp v0.19.4 (Chen et al., 2018) were used for quality and adapter trimming and
188 ribosomal removing before reads mapping (Table S3). Cleaned mRNA reads were mapped to
189 the high-quality *M. domestica* reference genome GDDH13 v1.1 (Daccord et al., 2017) with
190 HISAT2 v2.0.4 software (Kim et al., 2015). The SAM file was processed with MarkDuplicates
191 from Picard v1.13 ('Picard Toolkit', 2019) to markup duplicates. Hard-clip to any sequences
192 overhanging into the intronic regions was applied using SplitNCigarReads from GATK with
193 the following arguments: -rf ReassignOneMappingQuality, -U ALLOW_N_CIGAR_READS.

194 The GATK module HaplotypeCaller was used for SNP calling, and clustered SNPs, and low-
195 quality SNPs were filtered out with the GATK module VariantFiltration using specific
196 arguments: -window 35 -cluster 3 -filterName FS -filter "FS > 30.0" -filterName QD -filter
197 "QD < 2.0". For short-read DNA sequences of the 97 accessions (Duan et al., 2017), SNPs
198 were called following the GATK best practices using the GATK v3.5 (McKenna et al., 2010).
199 After trimming low-quality reads and adapter sequence with fastp v0.19.4 (Chen, Zhou, Chen,
200 & Gu, 2018; Chen et al., 2019), cleaned DNA reads were aligned to the GDHH13 v1.1 *M.*
201 *domestica* reference genome (Daccord et al., 2017) with BWA 0.7.15 by using the “mem”
202 algorithm (H. Li, 2013). The redundant reads were removed using the MarkDuplicate module
203 from Picard v1.13 (‘Picard Toolkit’, 2019). The GATK module HaplotypeCaller was used for
204 SNP calling, and GenotypeGVCFs module was used to produce a raw set of joint SNPs. The
205 raw joint SNPs were finally filtered with the GATK VariantFiltration module and default
206 quality thresholds. Scripts are available on ForgeMIA
207 (https://forgemia.inra.fr/amandine.cornille/rnaseq_apple_evolutionary_history).

208

209 **Suitability of the markers used for phylogenetic, population structure, and demographic** 210 **inferences**

211 The filtering steps used for the analyses described below are provided in Figure S1. The SNP
212 files obtained from the RNA-seq and DNA-seq datasets were combined with the VCFtools
213 v0.1.16 vcf-merge function (Danecek et al., 2011). The highly genetically related and duplicate
214 individuals were removed (*i.e.*, individuals with pairwise KING-robust kinship estimates
215 among individuals > 0.354) (Manichaikul et al., 2010) with Plink v2.0 (Chang et al., 2015).
216 The SNPs with minimum minor allele frequency < 0.01 and minimum site coverage = 0.9 were
217 filtered out with Plink v1.07 (Purcell et al., 2007). A total of 76,239 SNPs were obtained. For
218 population structure and demographic inferences, linked and non-synonymous SNPs were

219 filtered out to avoid any bias associated with potential SNPs linked to the selected genomic
220 region. First, SNPs in linkage disequilibrium (LD) were removed; the LD for apples was
221 <1,000 bp (Duan et al., 2017). The linked SNPs were filtered out with Plink v1.07 (Purcell et
222 al., 2007) by using the following parameters: window = 1 kb, step = 1 SNP, and $r^2 = 0.2$.
223 Second, the remaining SNPs were annotated with SnpEff v3.6 (Cingolani et al., 2012), and
224 only synonymous SNPs were retained; therefore, 36,200 unlinked and synonymous SNPs were
225 used (Figure S1).

226

227 **Population structure, genetic diversity, and differentiation among wild and cultivated** 228 **apples**

229 The population structure and admixture among *Malus* accessions were inferred using a model-
230 based clustering method implemented in ADMIXTURE v1.23 (Alexander et al., 2009).
231 ADMIXTURE is based on the same statistical model as the one implemented in STRUCTURE
232 (Pritchard et al., 2000), but it uses a fast numerical optimization algorithm for large SNP
233 datasets to infer the proportion of ancestry of genotypes in K distinct predefined clusters. The
234 number of ancestral populations, K , was varied from 1 to 10, with 20 repetitions per K and
235 default settings. The consensus solution for each K value was generated with CLUMPAK
236 (Kopelman et al., 2015). The amount of additional information explained by increasing K was
237 determined using the cross-validation procedure of ADMIXTURE v1.23 (Alexander et al.,
238 2009). However, the K identified with the cross-validation procedure often does not correspond
239 to the finest population structure (Cornille et al., 2015; Kalinowski, 2011; Puechmaille, 2016).
240 Therefore, the bar plots were visually checked, and the K value for which all clusters had well-
241 assigned individuals was selected (*i.e.*, that no further well-delimited and relevant clusters
242 could be identified for higher K values than the chosen one). All the bar plots were generated
243 with the R package “Pophelper” (Francis, 2017).

244 The genetic variation and differentiation among the groups detected with
245 ADMIXTURE were further investigated using three different methods. First, a principal
246 component analysis (PCA) was performed with the `dudi.pca` function of the “ade4” package in
247 R (Dray & Dufour, 2007). The first two principal components were plotted with the R package
248 “ggplot2” (Wickham, 2009). Second, a neighbor-net tree based on Nei’s distance (Nei, 1978)
249 among individuals was created. The neighbor net was visualized with Splitstree v4.12.3 (Huson
250 & Bryant, 2006). For the PCA and neighbor-net, individuals assigned to a given cluster with a
251 membership coefficient ≥ 0.55 were indicated using the color of each cluster, and admixed
252 individuals (*i.e.*, individuals with a membership coefficient to any given cluster < 0.55) were
253 colored gray. We chose this threshold because both cultivated and wild apples were highly
254 admixed, and the use of a higher cut-off would have removed too many individuals (see
255 Results). For each wild or cultivated population (*i.e.*, group of individuals with membership
256 coefficient > 0.55 to a given cluster detected with ADMIXTURE), Nei’s diversity (π) (Nei &
257 Li, 1979), observed (H_O) and expected heterozygosity (H_E) (Nei, 1978), inbreeding coefficient
258 F_{IS} , and genetic differentiation (F_{ST}) between populations were estimated with Stacks v2.52
259 population program (Rochette et al., 2019).

260

261 **Detection of gene flow**

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

265 First, TreeMix v1.12 (Pickrell & Pritchard, 2012) with the ipyrad v0.9.78 (Eaton &
266 Overcast, 2020) was used to analyze the relationships among the populations and the potential
267 influence of gene flow. TreeMix uses genome-wide allele frequency data to estimate the
268 maximum likelihood tree of populations. Then, on the basis of populations with the poorest fit

269 to the tree model, the program infers the presence, magnitude, and the direction of migration
270 events between them (Pickrell & Pritchard, 2012). Trees were rooted with the *Malus* species
271 from North America, which are known to be the most divergent and ancestral *Malus* group
272 (Harris et al., 2002 and see Results). Zero to nine migration numbers were tested, and the run
273 for which the likelihood of trees reached an initial plateau (Aguirre-Liguori et al., 2019;
274 Brandenburg et al., 2017) and the magnitude of migration rates decreased was retained.

275 Second, Dtrios of Dsuite v0.3 (Malinsky et al., 2021) was used to perform four-taxon
276 D-statistic tests on the final unlinked synonymous SNP dataset. *D-statistics* is also called
277 Patterson's *D* (ABBA-BABA) statistic (Durand et al., 2011; Green et al., 2010), and it checks
278 whether gene flow occurred among populations (*i.e.*, group of individuals assigned with a
279 membership coefficient > 0.55 to a given genetic group) with ADMIXTURE and 36,200
280 synonymous unlinked SNPs. Dsuite calculates *D-statistics* not only for one specific quartet but
281 also, comprehensively, for sets of populations in a VCF file and keeps the outgroup fixed (here
282 the North American wild apples, see Results). The significance of *D-statistics* was assessed
283 using Jackknife (Green et al., 2010) on 20 blocks. Results for which the *Z-score* was higher
284 than 3 and *p-value* < 0.01 were retained.

285

286 **Approximate Bayesian computation**

287 The history of apple domestication was reconstructed using the approximate Bayesian
288 computation (ABC) framework in combination with a coalescent-based inference simulator,
289 fastsimcoal2 v2.5.2.21 (Excoffier et al., 2013). Our aims were to infer (1) whether gene flow
290 occurred among certain crop and wild populations during apple domestication and (2) the
291 European and Chinese cultivated apples originated from which wild apple population(s).
292 Previous studies (Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012;
293 Cornille et al., 2014; Duan et al., 2017; Migicovsky et al., 2021) have reported that the

294 European dessert apple (*M. domestica*) diverged from *M. sieversii*. Therefore, the origin of the
295 European dessert apple was not tested, and it was assumed that it diverged from *M. sieversii* in
296 our models. The scenarios were established according to the results obtained from
297 ADMIXTURE, PCA, and neighbor-net tree as well as the results of *D-statistics* and TreeMix.
298 The populations were defined as those detected with ADMIXTURE analyses (*i.e.*, group of
299 individuals assigned with a membership coefficient > 0.55 to a given genetic group).

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

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

317 For all models, 36,200 unlinked SNPs were simulated, with 8,000 simulations per
318 scenario. For each simulation, the following summary statistics were computed with

319 arlsumstats v3.5 (Excoffier & Lischer, 2010): the number of sites with segregating substitutions
320 of population i ($S_{\cdot i}$, $i = \{1, 2, 3, 4, 5\}$), mean number of pairwise differences of population i
321 ($\pi_{\cdot i}$, $i = \{1, 2, 3, 4, 5\}$), mean number of differences between pairs of populations ($\pi_{i \cdot j}$, $i, j =$
322 $\{1, 2, 3, 4, 5\}$, $i \neq j$), and pairwise F_{ST} ($F_{ST \cdot i \cdot j}$, $i, j = \{1, 2, 3, 4, 5\}$, $i \neq j$). Summary statistics
323 were also added on the basis of the joint frequency spectrum between all pairs of populations
324 (Tellier et al., 2011) computed with a home-made script: sites polymorphic in population i , but
325 monomorphic in population j , and vice-versa ($S_{x1 \cdot i \cdot j}$ and $S_{x2 \cdot i \cdot j}$, respectively); number of shared
326 polymorphic sites between population i and population j ($S_{si \cdot j}$); and number of sites showing
327 fixed differences between populations i and j ($S_{fi \cdot j}$).

328 The “abcrf” v1.7.0 R statistical package (Pudlo et al., 2016) was used to perform the
329 ABC-RF analysis, which provides a classification vote that represents the number of times a
330 scenario is selected as the best one among n trees in the constructed random forest. For each
331 ABC set, the scenario, or a group of scenarios, with the highest number of classification votes
332 was selected as the best scenario, or best group of scenarios, among a total of 500 classification
333 trees (Breiman, 2001). The posterior probabilities and prior error rates (*i.e.*, the probability of
334 choosing a wrong group of scenarios when drawing model index and parameter values from
335 the priors of the best scenario) over 10 replicate analyses (Estoup et al., 2018) were computed
336 for each ABC step. The simulated models were also visually checked to be compatible with
337 the observed dataset by projecting the simulated and observed datasets onto the first two linear
338 discriminant analysis (LDA) axes (Pudlo et al., 2016) and whether the observed dataset was
339 within the clouds of simulated datasets. The parameter inferences were then calculated using
340 the final selected model, according to the three-set ABC procedure. The ABC-RF approach
341 includes the model checking step performed *a posteriori* in previous ABC methods.

342 Once the most likely apple domestication history was obtained, posterior distributions
343 were estimated for each parameter and 1,000 pseudo-observed datasets were re-simulated using

344 the 95% confidence interval of the posterior estimates to evaluate the goodness of fit of the
345 final chosen demographic model with “abc” R package (Csilléry et al., 2012).

346

347 **Results**

348 **The European and Chinese cultivars have different ancestries**

349 We removed 20 individuals of the initial dataset ($N = 168$) because they had unknown passport
350 information ($N = 1$) or were detected as a clone or duplicate ($N = 19$, Table S1). The first run
351 of ADMIXTURE, in which the 20 individuals were removed ($N = 148$), revealed 20 wild apple
352 trees described as *M. sieversii* or *M. sylvestris* or *M. baccata* admixed with the cultivated gene
353 pool (Figures S4 and S5) or misidentified on the field (*i.e.*, fully assigned to *M. domestica* gene
354 pool), as previously observed (Cornille et al., 2013, 2014, 2015; Feurtey et al., 2017). We also
355 detected 13 wild individuals, described as not one of the three wild apple species studied here,
356 admixed with multiple gene pools (Table S1, Figures S4 and S5). In this study, we focused on
357 the history of the cultivated apple, so we wanted “pure” wild apple reference samples of the
358 three wild apple relatives, *i.e.*, *M. baccata*, *M. sieversii*, and *M. sylvestris*. Therefore, we filtered
359 out the 33 wild admixed or misidentified individuals detected with the first ADMIXTURE
360 analysis (Figures S6, S7, Table S1). We retained *M. hupehensis*, as the samples belonging to
361 this species grouped with *M. baccata* at a high membership (Figure S6); this population was
362 referred as to *M. baccata*-*M. hupehensis*. Population genetic structure inferences for the whole
363 dataset ($N = 148$) vs. the pruned dataset for those admixed individuals ($N = 115$) were consistent
364 (Figures S4 and 1a). However, ADMIXTURE revealed a clearer population genetic structure
365 for the dataset pruned for wild admixed individuals. Therefore, we retained the dataset pruned
366 for wild admixed or misidentified individuals for further analyses.

367 ADMIXTURE revealed five main genetic groups (Figure 1a). Cross-validation error
368 monotonically decreased up to $K = 8$ (Figures S8 and S9). At $K > 5$, further substructure was

369 observed, but new clusters were represented by only admixed individuals (Figure S8). Besides,
370 at $K = 5$, the population structure was consistent with the morphological classification of the
371 wild species (Cornille et al., 2014, 2019; Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens,
372 Cam, et al., 2012; Harris et al., 2002; Robinson et al., 2001), with *M. sylvestris* (orange), *M.*
373 *sieversii* (purple), *M. baccata* - *M. hupehensis* (blue), and North American wild apples (green)
374 each forming a distinct group. At $K = 5$, the European apple cultivars (*i.e.*, *M. domestica*,
375 including European dessert and rootstock cultivars) formed a distinct genetic group highly
376 admixed with *M. sylvestris* (Table 1), as shown by Cornille et al. (2012), but also with the *M.*
377 *baccata* - *M. hupehensis* (blue) genetic group (Table 1). For $K = 5$, the Chinese dessert and
378 rootstocks as well as ornamental Chinese apple species did not form a specific genetic group,
379 but a genetic mixture of the blue (*M. baccata* - *M. hupehensis*) and purple (*M. sieversii*) gene
380 pools (Table 1) as well as, to a lesser extent, with the orange genetic group (*M. sylvestris*).

381 Then, we assessed the genetic variations among wild and cultivated apple populations
382 detected with ADMIXTURE. We assigned wild individuals with a membership coefficient $>$
383 0.55 to a given cluster in the corresponding population (Table 1). Because of the high level of
384 admixture in cultivated apples, we retained all cultivars and defined five cultivated apple
385 groups based on the ADMIXTURE results as well as their historical uses and geographic
386 origins (Tables 1 and S1). The use and geographic origins of the cultivars were documented on
387 the basis of the studies performed by Morgan & Richards (1993, 2003) and Urrestarazu *et al.*
388 (2016) for the European-Western cultivars and Zhou (1999) for the Chinese cultivars. The
389 North American and Siberian wild apples were the most genetically differentiated population
390 (Figure 1b, c, d). *M. sieversii* individuals formed a distinct group (Figure 1b), close to that of
391 the European dessert apple. The European cultivars (*M. domestica*) were closer to *M. sylvestris*
392 than to the known progenitor *M. sieversii* (Figure 1b, c, d), as detected previously (Cornille,
393 Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012; Duan et al., 2017; Migicovsky

394 et al., 2021). *M. sylvestris* was nested with *M. domestica* in the neighbor-net tree (Figure 1b),
395 but it formed a distinct, but close, gene pool to that of *M. domestica* in the PCA (Figure 1c).
396 Most of the European rootstocks were admixed with the three wild apple species (Figure 1b,
397 c). The Chinese dessert and rootstock formed a bushy structure intermingled with *M. baccata*
398 and *M. hupehensis*. F_{ST} estimates further confirmed that the European cultivars were
399 genetically close to *M. sylvestris* and *M. sieversii*, and the Chinese cultivars were genetically
400 close to *M. baccata*-*M. hupehensis* (Figure 1d). Genetic diversity estimates for the nine
401 populations are provided in Figure S10 and Table S5. Genetic diversity was significantly lower
402 in the wild apple than in the cultivated apple gene pools, except for the *M. baccata*-*M.*
403 *hupehensis* group; this can be explained by the high level of admixed individuals in the
404 cultivated populations (Table 1, Figure 1).

405 Therefore, ADMIXTURE, neighbor-net tree, PCA, and F_{ST} suggest that the cultivated
406 European and Chinese apples have different ancestries, which raise questions about their origin
407 and whether crop-wild gene flow occurred. The European dessert and rootstock cultivars
408 grouped together (*M. domestica*) and formed a specific gene pool; this has been previously
409 reported for the dessert apples (Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et
410 al., 2012; Migicovsky et al., 2021). The European dessert and rootstock cultivars showed
411 different levels of admixture with the wild species: the European dessert cultivars were mostly
412 admixed with *M. sylvestris* (and *M. sylvestris* was nested in *M. domestica*), whereas the
413 European rootstock cultivars were mainly admixed with *M. baccata* - *M. hupehensis* and *M.*
414 *sieversii*. The high level of admixture of the European rootstock with *M. baccata* - *M.*
415 *hupehensis* and *M. sieversii* raises questions about the origin of the European rootstock
416 cultivars: their distinct genetic differentiation from *M. baccata* - *M. hupehensis* and clustering
417 with *M. domestica* (which we know originated from *M. sieversii*) may suggest that *M. baccata*-
418 *M. hupehensis* contributed to the European rootstock through recent crop-to-wild

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

436

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

438 TreeMix (Figure 2 and S11) indicated the occurrence of gene flow from the European wild
439 apple to the European dessert and rootstock, but not from *M. sieversii*. TreeMix also indicated
440 gene flow from *M. sieversii* to the Chinese apple rootstock and from *M. baccata-M. hupehensis*
441 to the Chinese dessert. The D-suite inferences were congruent with TreeMix results, with
442 significant excess of sharing of derived alleles among almost all crop and wild population pairs
443 (Table S6). In addition, TreeMix analyses suggested that the cultivated European apples

444 formed a monophyletic group close to *M. sieversii*; in contrast, the Chinese cultivars formed a
445 polyphyletic group, with the Chinese dessert apple being a sister group with *M. sieversii* and
446 Chinese rootstock, a sister group with *M. baccata-M. hupehensis*. However, TreeMix can be
447 biased by its starting tree, especially when many populations are admixed (Lipson, 2020;
448 Lipson et al., 2013). Therefore, we investigated the domestication history of the Chinese and
449 European cultivars by using coalescent-based approaches, which allow us to disentangle
450 between crop-wild gene flow and ancestral polymorphisms.

451

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

454 For ABC inferences, we excluded the North American wild apple population, as it was the
455 most genetically distant and did not contribute to the cultivated apple genome in the analyses
456 (Figure 1). We also excluded the Chinese ornamental apple because it included only four
457 individuals. Our aim was to analyze (1) the extent of crop-wild gene flow and (2) infer the
458 divergence history of the cultivated apple populations (excluding the European dessert, as its
459 history is already known) (Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al.,
460 2012; Duan et al., 2017; Harris et al., 2002; Migicovsky et al., 2021; Sun et al., 2020; Velasco
461 et al., 2010).

462 First, we inferred the origin of the cultivated European rootstock apples (Figures S2
463 and Figure 3) with respect to the three wild apple populations and European dessert. We
464 assumed that three wild apples diverged from an unknown ancestral population and the
465 European dessert apple *M. domestica* diverged from *M. sieversii*. Our results above (population
466 structure and gene flow estimates) suggest that *M. baccata-M. hupehensis* did contribute to the
467 European rootstock through recent crop-to-wild introgressions and the European rootstock
468 cultivars originated from *M. sieversii*. For the first set of ABC (Figure S2), we tested whether

469 the European rootstock diverged from *M. domestica* or *M. sieversii* (Figure S2) and assumed
470 four different modalities of gene flow: no gene flow, gene flow between European dessert and
471 *M. sylvestris* only (Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012),
472 gene flow between the European rootstock and *M. sylvestris* and *M. baccata-M. hupehensis*
473 only, and a combination of the first two modalities of gene flow (Figure S2). The last modality
474 was defined on the basis of an additional TreeMix analysis that focused on only the populations
475 used for ABC set 1 (Figure S12). Therefore, we simulated eight scenarios for the first ABC set
476 (Figure S2). For the second set of ABC analysis (set 2, Figure S3), we inferred the
477 domestication history of the Chinese dessert and rootstock apple on the basis of the most likely
478 group of scenarios selected from the first ABC set. We tested whether the Chinese desert and
479 rootstock cultivated apples diverged from (i) *M. domestica*, (ii) *M. baccata-M. hupehensis*, or
480 (iii) *M. sieversii*. We defined 12 different scenarios of divergence of the Chinese cultivars and
481 assumed two modalities of gene flow: gene flow as inferred from TreeMix (Figure 2) and no
482 gene flow between the Chinese rootstock and *M. baccata-M. hupehensis* and *M. sieversii* + no
483 gene flow between the Chinese dessert with *M. baccata-M. hupehensis* and *M. sieversii* (Figure
484 S3). In total, we simulated 24 scenarios. For each ABC round, the observed summary statistics
485 were within the cloud of simulated summary statistics, which did not overlap across the model
486 groups (Figure S13) and indicated that these tests could discriminate among various competing
487 scenarios (except for the fourth round of the second ABC set, Figure 3).

488 For the first round of the first ABC set (Figures 3a and S2), the classification votes were
489 the highest 10 times out of 10 for the group of scenarios that assumed gene flow among the
490 European dessert (*M. domestica*) and *M. sylvestris*, European rootstock and *M. sylvestris*, and
491 *M. baccata-M. hupehensis* (i.e., gene flow modality 4 in Figure S3 and 3, 479.50 votes out of
492 the 500 RF trees, posterior probability $P = 1.00$, prior error rate = 0.00, Table S7, Figure S13a,
493 Figure 3a). In the second round, the classification votes were the highest 10 times out of 10 for

494 the scenarios that assumed the European rootstock from *M. sieversii* (422 votes out of the 500
495 RF trees, posterior probability $P = 0.87$, prior error rate = 10.02%, Table S8, Figure S13b).
496 Thus, the first ABC set indicated the European rootstock originated from *M. sieversii* with
497 subsequent gene flow from *M. baccata-M. hupehensis* and *M. sylvestris*.

498 For the first round of the second ABC set, the classification votes were the highest 10
499 times out of 10 for the scenarios assuming gene flow between the Chinese rootstock and dessert
500 with *M. baccata-M. hupehensis* (*i.e.*, gene flow modality 2, 98.9% RF trees voting for this
501 group, posterior probability $P = 1.00$, prior error rate = 0.00%, Table S9 and Figure S14a). For
502 the second round of the second ABC set, the classification votes were the highest 10 times out
503 of 10 for the scenarios that assumed the two Chinese cultivated apples were domesticated
504 independently from different wild species (67.8% RF trees voting for this group, posterior
505 probability $P = 0.88$, prior error rate = 9.61%, Table S10 and Figure S14b). For the third round,
506 the classification votes were the highest 10 times out of 10 for the scenarios that assumed the
507 Chinese rootstock and dessert originated from *M. sieversii* or *M. domestica* (56% RF trees
508 voting for this group of scenarios, posterior probability $P = 1.00$, prior error rate = 0.00%, Table
509 S11 and Figure S14c). For the fourth round, the classification votes were the highest 10 times
510 out of 10 for the scenarios that assumed the Chinese dessert apple originated from *M. domestica*
511 (57.6% votes out of the 500 RF trees, posterior probability $P = 0.78$, prior error rate = 15.27%,
512 Table S12, Figure S14d). However, origin of the Chinese dessert from *M. domestica* needs to
513 be considered with caution, as we were not able to fit the observed data within the LDA (Figure
514 S14d).

515 The final scenario indicated that the European rootstock diverged from *M. sieversii* and
516 the Chinese cultivars did not diverge from *M. baccata-M. hupehensis*, but their origin, either
517 from *M. sieversii* or *M. domestica*, remains unclear (Figure 3b). The ABC further confirmed
518 TreeMix gene flow estimates: *M. sylvestris* and *M. baccata-M. hupehensis* contributed to the

519 European and Chinese dessert and rootstock gene pools (Figure 3b). The model parameters are
520 listed in Table S13, and the posterior estimates are presented in Table S13. However, it should
521 be noted that the credibility intervals were large, and the parameter estimates need to be
522 carefully considered. The goodness of fit of the two models (Figures S15 and S16) confirmed
523 our model choice ($P = 0.069$ and $P = 0.07$, respectively).

524

525 **Discussion**

526 In this study, we revealed that the East Asian wild apples, *M. baccata* and *M. hupehensis*, form
527 a panmictic group that has contributed to the cultivated European and Chinese dessert and
528 rootstock apples through wild-to-crop introgression, as has been described in apple in Europe
529 (Cornille, Gladieux, Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012). We have also
530 provided insight into the origin of the apple rootstock. We showed that the cultivated European
531 dessert and rootstock apples grouped together and formed a specific gene pool, whereas the
532 Chinese dessert and rootstock apples were a mixture of the three wild apple gene pools, mainly
533 with *M. baccata*-*M. hupehensis* and *M. sieversii*. The coalescent-based inferences indicate that
534 both European and Chinese rootstocks diverged from *M. sieversii* (or *M. domestica* for the
535 cultivated Chinese apple gene pool), with subsequent gene flow from the wild species *M.*
536 *sylvestris* and *M. baccata*-*M. hupehensis*. We also confirmed previous results for the
537 contribution of *M. sylvestris* to the cultivated dessert apple gene pool (Cornille, Gladieux,
538 Smulders, Roldán-Ruiz, Laurens, Cam, et al., 2012). This study confirmed that domestication
539 of the apple tree involved several wild apple species and that crop-wild species hybridization
540 had a key role in fruit tree domestication. Substantial hybridization between domestic and wild
541 forms have also been described in other fruit trees (Groppi et al., 2021; Liu et al., 2019; Wu et
542 al., 2018), but the apple tree is the model system with the most documentation. Our results
543 support the view that domestication of woody perennials, and crops in general, was probably a

544 protracted and diffuse process that involved multiple geographically disparate species (Allaby
545 et al., 2008; Purugganan, 2019).

546

547 **The European and Chinese apple rootstocks did not originate from the two East Asian**
548 **wild apples**

549 Although grafting has been an important part of perennial woody crop evolution and breeding
550 (Vavilov, 1926; Warschefsky et al., 2016; Zohary & Spiegel-Roy, 1975), surprisingly little is
551 known about the origin of rootstocks. In apples, most of the studies focused on the history of
552 dessert and cider cultivars, mostly from the Western Hemisphere. The Central Asian wild
553 apple, *M. sieversii*, is now known to be the wild ancestor of the cultivated dessert and cider
554 apple, *M. domestica* (Cornille et al., 2014, 2019; Cornille, Gladieux, Smulders, Roldán-Ruiz,
555 Laurens, Cam, et al., 2012; Daccord et al., 2017; Duan et al., 2017; Harris et al., 2002;
556 Migicovsky et al., 2021). However, the origin of apple rootstock has not yet been elucidated.
557 Here, we provide new insights into the domestication history of apple rootstock and show that
558 *M. sieversii* is the progenitor of the European apple rootstock and either *M. sieversii* or *M.*
559 *domestica* is the progenitor of the Chinese apple. The mixed origin of Chinese dessert cultivars,
560 the shared ancestral polymorphism between *M. domestica* and *M. sieversii*, and the current low
561 number of Chinese cultivars, may explain our inability to infer the origin of the Chinese dessert
562 apple with the ABC framework. Indeed, only four main Chinese cultivars are still cultivated in
563 China; they are supposed to belong to two species, *M. domestica* subsp. *chinensis* and *M.*
564 *asiatica*. Note as well that only few main rootstocks are used in China (Wang et al., 2019). In
565 our study, the sampling size of the Chinese cultivated apple was limited because of its
566 cultivation history in this country. Thus, despite the large use of *M. baccata* and *M. hupehensis*
567 (Wang et al., 2019) as breeding sources in China and for rootstock in Europe, our results
568 suggest that they are not the progenitors of the rootstock varieties. Rather, our results suggest

569 that *M. baccata*-*M. hupehensis* did contribute to apple domestication through wild-to-crop
570 gene flow, as has been described in apple in Europe (Cornille, Gladieux, Smulders, Roldán-
571 Ruiz, Laurens, Cam, et al., 2012).

572

573

574 **Major contribution of East Asian wild apples to the cultivated apple genome through**
575 **crop-to-wild introgression**

576 There is much debate on the contribution of other wild species along the Silk Route to the
577 genetic makeup of the cultivated apple genome. Recently, the Caucasian crab apple has been
578 shown to be an additional contributor to the cultivated apple genome (Bina et al., 2021). Our
579 results demonstrate that interspecific hybridization has been a driving force in the evolution of
580 apple. We confirm that the European wild crabapple *M. sylvestris* has been a major secondary
581 contributor to the diversity of apples. We also show that the East Asian wild apples, which
582 form a panmictic group, are contributors to the diversity of the cultivated apple. Furthermore,
583 we show the substantial gene flow from *M. baccata*-*M. hupehensis* to the European apple
584 rootstock and Chinese dessert apple and rootstock. This is not surprising because *M. baccata*
585 and *M. hupehensis* are widely used in the high-latitude apple-producing areas of China as a
586 rootstock and breeding resource because of its disease resistance and cold tolerance (X. Chen
587 et al., 2019a; Zhou, 1999). In the Western hemisphere, *M. baccata*-*M. hupehensis* may be used
588 as parents for some crosses. Interspecific hybridization has also been observed in date palms
589 (Flowers et al., 2019), grapes (Myles et al., 2011), almonds (Delplancke et al., 2011), and
590 apricots (Groppi et al., 2021; Liu et al., 2019; Q. Zhang et al., 2021). Substantial crop-wild
591 gene flow is mostly due to the self-incompatibility system of fruit trees that favors the selection
592 of the best phenotypes grown from open-pollinated seeds and appears to be a key element in
593 perennial crop evolution (Cornille et al., 2014; Gaut et al., 2015).

594 Coalescent-based methods combined with ABC-RF therefore provided a powerful
595 statistical framework for inferring the contributions of the wild species to the Chinese and
596 European cultivated gene pools. However, a broader view of the domestication history of the
597 apple tree in East Asia will require additional sampling of *M. hupehensis*. The initial aim of the
598 study was to investigate the contribution of *M. baccata* to the cultivated apple rootstock gene
599 pool, but our new results open avenues for investigating the contributions of *M. hupehensis*
600 which naturally occurs in China. In addition, to keep the number of scenarios tractable, we
601 fixed the divergence histories of the wild apple populations starting from an ancestral
602 population. Despite the high power of ABC-RF framework, a thorough investigation of the
603 evolutionary histories of the wild apples, which would require additional samples of each wild
604 species across their distributions using additional SNP markers, will likely provide additional
605 insight into the apple domestication and diversification histories.

606

607 **Concluding remarks and perspectives**

608 The domestication of fruit trees stands in stark contrast to that of annuals, especially with
609 respect to the extent of wild-crop gene flow. Substantial crop-wild gene flow makes the
610 resolution of relationships between crop and wild fruit trees challenging. In addition, the
611 problem with unravelling the domestication history of fruit trees is magnified by their recent
612 domestication in terms of number of generations and thus shared ancestral polymorphism
613 among the crop and wild fruit trees (Cornille et al., 2014; Gaut et al., 2015; Miller & Gross,
614 2011). We used genomic data to distinguish between crop-wild shared ancestral polymorphism
615 and recent gene flow and showed that the history of apples is a rare well-documented example
616 of the evolution of a domesticated crop over a long period and involved at least four wild
617 species. Introgression of from wild relatives is a likely source of crop adaptation (Burgarella et
618 al., 2019). The contribution of several species to the cultivated apple genome raises questions

619 about their role in the adaptation of cultivated apples during domestication worldwide.
620 Genome-wide association showed that traits associated with fruit quality and texture were
621 selected during apple domestication (Duan et al., 2017; Migicovsky et al., 2021). However, the
622 genomic landscape of wild-to-crop introgressions has not been investigated on the basis of
623 genome-wide data (Daccord et al., 2017; Duan et al., 2017; Migicovsky & Myles, 2017; Sun
624 et al., 2020; Velasco et al., 2010). Further investigations of the genomic architecture of wild-
625 crop gene flow, and its adaptive role during domestication, are required to understand the role
626 of gene flow during the divergence and adaptation of woody perennials to new environments.

627

628 **Acknowledgements**

629 We thank the Franco-Chinese Campus France program « Decouverte » 2017, the Jeunes
630 Talents France-Chine 2019 program, and the ATIP-Avenir CNRS Inserm grant for funding.
631 We also thank Adrien Falce, Olivier Langella and Benoit Johannet for help and support on the
632 INRAE-Génétique Quantitative et Evolution-Le Moulon lab cluster. We thank the INRAE
633 MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing help and support,
634 in particular Véronique Martin, Eric Montaubon and Valentin Loux. This work was financially
635 supported by the National Key Research and Development Program of China
636 (2018YFD1000101, 2019YFD1000803); the Key Technology R & D and Industrialization
637 Demonstration Science and Technology for the Transformation and Upgrading of Apple
638 Industries in Shaanxi Province-Apple Anvil Combinations Evaluation and Seedlings
639 Industrialization (2020zdzx03-01-04); The China Apple Research System (CARS 27), Tang
640 Scholar by Cyrus Tang Foundation and Northwest A&F University.

641

642 **Data Availability**

643 The RNA-seq raw data are available at NCBI sequence read archive (SRA) under BioProject
644 accession PRJNA763361. VCF file used for analyses is available at the Zenodo (<https://zenodo.org/record/5513618>). Scripts are available at [https://forgemia.inra.fr/amandine.cornille/rnas](https://forgemia.inra.fr/amandine.cornille/rnas_eq_apple_evolutionary_history)
645 [eq_apple_evolutionary_history](https://forgemia.inra.fr/amandine.cornille/rnas_eq_apple_evolutionary_history).
646

647

648 **Author Contributions**

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

653

654 **Tables**

655

656 **Table 1. Mean proportions of nine cultivated and wild apple populations assigned to five**
657 **genetic clusters inferred with ADMIXTURE at $K = 5$ ($N = 115, 36,200$ SNPs).** We divided
658 the cultivated apple samples into five populations according to ADMIXTURE inferences and
659 their origin (Europe or China) and use (dessert or rootstock). We assigned only wild
660 populations with membership coefficient > 0.55 to a given genetic group. Thus, samples were
661 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-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. domestica</i>)	26	0.00	0.01	0.09	0.09	0.81
	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					

662 N: number of individuals.

663

664 **Figures**

665

666 **Figure 1. Genetic structure, variation, differentiation, and relationships of wild and**
667 **cultivated European and Chinese apples.** a. Population structure inferred with
668 ADMIXTURE at $K = 5$ ($N = 115$, 36,200 unlinked synonymous SNPs); each individual is
669 represented by a vertical bar partitioned into K segments that represent the proportions of
670 ancestry of its genome in K clusters. Colors represent the inferred ancestry from K ancestral
671 genetic clusters. b. Neighbor net tree depicting the relationships among the wild and cultivated
672 apple populations identified with ADMIXTURE for $K = 5$. Colors correspond to the genetic
673 groups inferred at $K = 5$, and admixed samples (*i.e.*, individuals with membership coefficient
674 < 0.55 to any given gene pool) are in grey. c. Principal component analysis (PCA) representing
675 the genetic variation among wild and cultivated apples, with the total variance explained by
676 the first two components (29.72% and 16.11%, respectively). d. Pairwise genetic
677 differentiation (F_{ST}) among populations identified with ADMIXTURE at $K = 5$ (excluding wild
678 individuals with membership coefficient < 0.55 to a given gene pool).

679

680 **Figure 2. Gene flow inferred with TreeMix ($N = 114$, 36,200 SNPs; Chinese ornamentals**
681 **were removed because of low number of individuals, $N = 4$) among the wild and cultivated**
682 **apple populations detected with ADMIXTURE at $K = 5$.** The North American wild apples
683 were used as the outgroup. The best number of migration events was six (Figure S11). Arrows
684 of different colors represent migration events. Width of the arrow indicates the migration
685 weight.

686

687

688 **Figure 3. Domestication histories of European and Chinese dessert and rootstock apples**
689 **inferred using the approximate Bayesian computation random forest framework (ABC-**
690 **RF, 36,200 SNPs)** a. Most likely scenario of domestication of the European rootstock inferred
691 with ABC-RF set 1 ($N = 98$). b. Most likely scenario of domestication of the European and
692 Chinese apples (set 2, $N = 111$). For each scenario, associated statistics (posterior probability,
693 prior error rates, and percentage of votes) and parameter estimates for the final model in set 2.
694 The grey arrow indicates that the Chinese rootstock and dessert cultivars may have originated
695 from *M. sieversii* or *M. domestica*.
696

697 **References**

- 698 Aguirre-Liguori, J. A., Gaut, B. S., Jaramillo-Correa, J. P., Tenailon, M. I., Montes-
699 Hernández, S., García-Oliva, F., Hearne, S. J., & Eguiarte, L. E. (2019). Divergence
700 with gene flow is driven by local adaptation to temperature and soil phosphorus
701 concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*).
702 *Molecular Ecology*, 28(11), 2814–2830. <https://doi.org/10.1111/mec.15098>
- 703 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of
704 ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664.
705 <https://doi.org/10.1101/gr.094052.109>
- 706 Allaby, R. G., Fuller, D. Q., & Brown, T. A. (2008). The genetic expectations of a protracted
707 model for the origins of domesticated crops. *PNAS*, 105(37), 13982–13986.
708 <https://doi.org/10.1073/pnas.0803780105>
- 709 Aravanopoulos, F. A., Ganopoulos, I., & Tsiftaris, A. (2015). Chapter Four—Population and
710 Conservation Genomics in Forest and Fruit Trees. In C. Plomion & A.-F. Adam-
711 Blondon (Eds.), *Land Plants—Trees* (Vol. 74, pp. 125–155). Academic Press.
712 <https://doi.org/10.1016/bs.abr.2015.04.001>
- 713 Bina, H., Yousefzadeh, H., Venon, A., Remoué, C., Rousselet, A., Falque, M., Faramarzi, S.,
714 Chen, X., Samanchina, J., Gill, D., Kabaeva, A., Giraud, T., Hossainpour, B.,
715 Abdollahi, H., Gabrielyan, I., Nersesyan, A., & Cornille, A. (2021). Evidence of an
716 additional center of apple domestication in Iran, with contributions from the
717 Caucasian crab apple. *BioRxiv*, 2021.03.28.437401.
718 <https://doi.org/10.1101/2021.03.28.437401>
- 719 Brandenburg, J.-T., Mary-Huard, T., Rigaille, G., Hearne, S. J., Corti, H., Joets, J., Vitte, C.,
720 Charcosset, A., Nicolas, S. D., & Tenailon, M. I. (2017). Independent introductions
721 and admixtures have contributed to adaptation of European maize and its American

722 counterparts. *PLOS Genetics*, 13(3), e1006666.
723 <https://doi.org/10.1371/journal.pgen.1006666>

724 Breiman, L. (2001). Random Forests. *Machine Language*, 45(1), 5–32.
725 <https://doi.org/10.1023/A:1010933404324>

726 Burgarella, C., Barnaud, A., Kane, N. A., Jankowski, F., Scarcelli, N., Billot, C., Vigouroux,
727 Y., & Berthouly-Salazar, C. (2019). Adaptive Introgression: An Untapped
728 Evolutionary Mechanism for Crop Adaptation. *Frontiers in Plant Science*, 10.
729 <https://www.frontiersin.org/article/10.3389/fpls.2019.00004>

730 Cao, K., Zhou, Z., Wang, Q., Guo, J., Zhao, P., Zhu, G., Fang, W., Chen, C., Wang, X.,
731 Wang, X., Tian, Z., & Wang, L. (2016). Genome-wide association study of 12
732 agronomic traits in peach. *Nature Communications*, 7, 13246.

733 Chagné, D., Crowhurst, R. N., Pindo, M., Thrimawithana, A., Deng, C., Ireland, H., Fiers,
734 M., Dzierzon, H., Cestaro, A., Fontana, P., Bianco, L., Lu, A., Storey, R., Knäbel, M.,
735 Saeed, M., Montanari, S., Kim, Y. K., Nicolini, D., Larger, S., ... Velasco, R. (2014).
736 The Draft Genome Sequence of European Pear (*Pyrus communis* L. ‘Bartlett’). *PLoS*
737 *ONE*, 9(4), e92644. <https://doi.org/10.1371/journal.pone.0092644>

738 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015).
739 Second-generation PLINK: Rising to the challenge of larger and richer datasets.
740 *GigaScience*, 4. <https://doi.org/10.1186/s13742-015-0047-8>

741 Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ
742 preprocessor. *Bioinformatics*, 34(17), i884–i890.
743 <https://doi.org/10.1093/bioinformatics/bty560>

744 Chen, X., Li, S., Zhang, D., Han, M., Jin, X., Zhao, C., Wang, S., Xing, L., Ma, J., Ji, J., &
745 An, N. (2019a). Sequencing of a Wild Apple (*Malus baccata*) Genome Unravels the
746 Differences Between Cultivated and Wild Apple Species Regarding Disease

747 Resistance and Cold Tolerance. *G3: Genes, Genomes, Genetics*, 9(7), 2051–2060.
748 <https://doi.org/10.1534/g3.119.400245>

749 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., &
750 Ruden, D. M. (2012). A program for annotating and predicting the effects of single
751 nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster*
752 strain w¹¹¹⁸; iso-2; iso-3. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>

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

757 Cornille, A., Feurtey, A., Gélin, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., & Giraud,
758 T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild fruit
759 tree: A basis for conservation and breeding programs in apples. *Evolutionary*
760 *Applications*, 8(4), 373–384. <https://doi.org/10.1111/eva.12250>

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

764 Cornille, A., Gladieux, P., & Giraud, T. (2013). Crop-to-wild gene flow and spatial genetic
765 structure in the wild closest relatives of the cultivated apple. *Evolutionary*
766 *Applications*, 6(5), 737–748.

767 Cornille, A., Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Cam, B. L.,
768 Nersesyán, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X.-G.,
769 Tenaillon, M. I., & Giraud, T. (2012). New Insight into the History of Domesticated
770 Apple: Secondary Contribution of the European Wild Apple to the Genome of

771 Cultivated Varieties. *PLOS Genetics*, 8(5), e1002703.
772 <https://doi.org/10.1371/journal.pgen.1002703>

773 Cornille, A., Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Le Cam, B.,
774 Nersesyan, A., Clavel, J., Olonova, M., Feugey, L., Gabrielyan, I., Zhang, X.-G.,
775 Tenailon, M. I., & Giraud, T. (2012). New insight into the history of domesticated
776 apple: Secondary contribution of the European wild apple to the genome of cultivated
777 varieties. *PLoS Genet*, 8(5), e1002703.

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

781 Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest,
782 H., Bianco, L., Micheletti, D., Velasco, R., Di Pierro, E. A., Gouzy, J., Rees, D. J. G.,
783 Guérif, P., Muranty, H., Durel, C.-E., Laurens, F., Lespinasse, Y., Gaillard, S., ...
784 Bucher, E. (2017). High-quality *de novo* assembly of the apple genome and
785 methylome dynamics of early fruit development. *Nature Genetics*, 49(7), 1099–1106.
786 <https://doi.org/10.1038/ng.3886>

787 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker,
788 R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & Group, 1000
789 Genomes Project Analysis. (2011). The variant call format and VCFtools.
790 *Bioinformatics*, 27(15), 2156–2158.

791 Delplancke, M., Alvarez, N., Espíndola, A., Joly, H., Benoit, L., Brouck, E., & Arrigo, N.
792 (2011). Gene flow among wild and domesticated almond species: Insights from
793 chloroplast and nuclear markers. *Evolutionary Applications*, 5(4), 317-329.

794 Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for
795 Ecologists. *Journal of Statistical Software; Vol 1, Issue 4 (2007)*.
796 <https://doi.org/10.18637/jss.v022.i04>

797 Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., Wang, X., Jiao, C., Legall, N., Mao, L.,
798 Wan, S., Wang, K., He, T., Feng, S., Zhang, Z., Mao, Z., Shen, X., Chen, X., Jiang,
799 Y., ... Chen, X. (2017). Genome re-sequencing reveals the history of apple and
800 supports a two-stage model for fruit enlargement. *Nature Communications*, 8(1), 249.
801 <https://doi.org/10.1038/s41467-017-00336-7>

802 Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture
803 between Closely Related Populations. *Molecular Biology and Evolution*, 28(8), 2239–
804 2252. <https://doi.org/10.1093/molbev/msr048>

805 Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq
806 datasets. *Bioinformatics*, 36(8), 2592–2594.
807 <https://doi.org/10.1093/bioinformatics/btz966>

808 Estoup, A., Raynal, L., Verdu, P., & Marin, J.-M. (2018). Model choice using Approximate
809 Bayesian Computation and Random Forests: Analyses based on model grouping to
810 make inferences about the genetic history of Pygmy human populations. *Journal de*
811 *La Société Française de Statistique*, 159(3), 167–190.

812 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust
813 Demographic Inference from Genomic and SNP Data. *PLOS Genetics*, 9(10),
814 e1003905. <https://doi.org/10.1371/journal.pgen.1003905>

815 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to
816 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
817 *Resources*, 10(3), 564–567.

818 Feurtey, A., Cornille, A., Shykoff, J. A., Snirc, A., & Giraud, T. (2017). Crop-to-wild gene
819 flow and its fitness consequences for a wild fruit tree: Towards a comprehensive
820 conservation strategy of the wild apple in Europe. *Evolutionary Applications*, *10*(2),
821 180–188. PMC. <https://doi.org/10.1111/eva.12441>

822 Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A.,
823 Ferrand, S., Khierallah, H. S. M., Fuller, D. Q., Aberlenc, F., Fournaraki, C., &
824 Purugganan, M. D. (2019). Cross-species hybridization and the origin of North
825 African date palms. *Proceedings of the National Academy of Sciences*, *116*(5), 1651–
826 1658. <https://doi.org/10.1073/pnas.1817453116>

827 Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize
828 population structure. *Molecular Ecology Resources*, *17*(1), 27–32.
829 <https://doi.org/10.1111/1755-0998.12509>

830 Gao, Y., Gao, S., Xiong, C., Yu, G., Chang, J., Ye, Z., & Yang, C. (2015). Comprehensive
831 analysis and expression profile of the homeodomain leucine zipper IV transcription
832 factor family in tomato. *Plant Physiology and Biochemistry*, *96*, 141–153.
833 <https://doi.org/10.1016/j.plaphy.2015.07.025>

834 Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of
835 Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719.
836 <https://doi.org/10.1016/j.tig.2015.10.002>

837 Glémin, S., & Bataillon, T. (2009). A comparative view of the evolution of grasses under
838 domestication. *New Phytologist*, *183*(2), 273–290. [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-8137.2009.02884.x)
839 [8137.2009.02884.x](https://doi.org/10.1111/j.1469-8137.2009.02884.x)

840 Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N.,
841 Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S.,
842 Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A.,

843 ... Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*,
844 328(5979), 710–722. <https://doi.org/10.1126/science.1188021>

845 Groppi, A., Liu, S., Cornille, A., Decroocq, S., & Decroocq, D. (2021). Population genomics
846 of apricots unravels domestication history and adaptive events. *Nature*
847 *Communications*, 12, 3956.

848 Gygax, M., Gianfranceschi, L., Liebhard, R., Kellerhals, M., Gessler, C., & Patocchi, A.
849 (2004). Molecular markers linked to the apple scab resistance gene Vbj derived from
850 *Malus baccata* jackii. *Theoretical and Applied Genetics*, 109(8), 1702–1709.
851 <https://doi.org/10.1007/s00122-004-1803-9>

852 Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the
853 apple. *Trends Genet*, 18(8), 426–430.

854 Harrison, N., & Harrison, R. J. (2011). On the evolutionary history of the domesticated apple.
855 *Nat Genet*, 43(11), 1043–1044.

856 Harrison, T., & Winfree, R. (2015). Urban drivers of plant-pollinator interactions. *Functional*
857 *Ecology*, 29(7), 879–888. <https://doi.org/10.1111/1365-2435.12486>

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

862 Huang, J., Zhang, C., Zhao, X., Fei, Z., Wan, K., Zhang, Z., Pang, X., Yin, X., Bai, Y., Sun,
863 X., Gao, L., Li, R., Zhang, J., & Li, X. (2016). The Jujube Genome Provides Insights
864 into Genome Evolution and the Domestication of Sweetness/Acidity Taste in Fruit
865 Trees. *PLOS Genetics*, 12(12), 1–20. <https://doi.org/10.1371/journal.pgen.1006433>

- 866 Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary
867 studies. *Molecular Biology and Evolution*, 23(2), 254–267.
868 <https://doi.org/10.1093/molbev/msj030>
- 869 Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the
870 main genetic clusters within species: Simulations and implications for human
871 population structure. *Heredity*, 106(4), 625–632.
- 872 Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low
873 memory requirements. *Nature Methods*, 12(4), 357–360.
874 <https://doi.org/10.1038/nmeth.3317>
- 875 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015).
876 Clumpak: A program for identifying clustering modes and packaging population
877 structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191.
878 <https://doi.org/10.1111/1755-0998.12387>
- 879 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-
880 MEM. *ArXiv:1303.3997 [q-Bio]*. <http://arxiv.org/abs/1303.3997>
- 881 Li, Y. (1999). An Investigation and Studies on the Origin and Evolution of *Malus domestica*
882 Borkh. In the World. *Acta Horticulturae Sinica*, 4.
- 883 Lipson, M. (2020). Applying f4-statistics and admixture graphs: Theory and examples.
884 *Molecular Ecology Resources*, 20(6), 1658–1667. [https://doi.org/10.1111/1755-](https://doi.org/10.1111/1755-0998.13230)
885 [0998.13230](https://doi.org/10.1111/1755-0998.13230)
- 886 Lipson, M., Loh, P.-R., Levin, A., Reich, D., Patterson, N., & Berger, B. (2013). Efficient
887 Moment-Based Inference of Admixture Parameters and Sources of Gene Flow.
888 *Molecular Biology and Evolution*, 30(8), 1788–1802.
889 <https://doi.org/10.1093/molbev/mst099>

890 Liu, S., Cornille, A., Decroocq, S., Tricon, D., Chague, A., Eyquard, J.-P., Liu, W.-S.,
891 Giraud, T., & Decroocq, V. (2019). The complex evolutionary history of apricots:
892 Species divergence, gene flow and multiple domestication events. *Molecular Ecology*,
893 28(24), 5299–5314. <https://doi.org/10.1111/mec.15296>

894 Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite—Fast D-statistics and related
895 admixture evidence from VCF files. *Molecular Ecology Resources*, 21(2), 584–595.
896 <https://doi.org/10.1111/1755-0998.13265>

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

900 Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez G, J., Buckler, E., & Doebley, J.
901 (2002). A single domestication for maize shown by multilocus microsatellite
902 genotyping. *PNAS*, 99(9), 6080–6084. <https://doi.org/10.1073/pnas.052125199>

903 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., & Kernytsky, A. (2010).
904 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-
905 generation DNA sequencing data. *Genome Res*, 20.
906 <https://doi.org/10.1101/gr.107524.110>

907 Meyer, R. S., Duval, A. E., & Jensen, H. R. (2012). Patterns and processes in crop
908 domestication: An historical review and quantitative analysis of 203 global food
909 crops. *New Phytologist*, 196(1), 29–48. [https://doi.org/10.1111/j.1469-](https://doi.org/10.1111/j.1469-8137.2012.04253.x)
910 [8137.2012.04253.x](https://doi.org/10.1111/j.1469-8137.2012.04253.x)

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

914 Migicovsky, Z., Gardner, K. M., Richards, C., Thomas Chao, C., Schwaninger, H. R., Fazio,
915 G., Zhong, G.-Y., & Myles, S. (2021). Genomic consequences of apple improvement.
916 *Horticulture Research*, 8(1), 1–13. <https://doi.org/10.1038/s41438-020-00441-7>

917 Migicovsky, Z., & Myles, S. (2017). Exploiting Wild Relatives for Genomics-assisted
918 Breeding of Perennial Crops. *Frontiers in Plant Science*, 8, 460.
919 <https://doi.org/10.3389/fpls.2017.00460>

920 Miller, A., & Gross, B. L. (2011). From forest to field: Perennial fruit crops domestication.
921 *American Journal of Botany*, 98, 1389–1414.

922 Morgan, J., & Richards, A. (1993). *The Book of Apples*. Brogdale Horticultural Trust, Ebury
923 Press.

924 Morgan, J., & Richards, A. (2003). *The New Book of Apples*. Brogdale Horticultural Trust,
925 Ebury Press.

926 Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B.,
927 Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C. D., & Buckler, E. S. (2011).
928 Genetic structure and domestication history of the grape. *PNAS*, 108(9), 3530–3535.
929 <https://doi.org/10.1073/pnas.1009363108>

930 Neale, D. B., Martínez-García, P. J., De La Torre, A. R., Montanari, S., & Wei, X.-X. (2017).
931 Novel Insights into Tree Biology and Genome Evolution as Revealed Through
932 Genomics. *Annual Review of Plant Biology*, 68(1), 457–483.
933 <https://doi.org/10.1146/annurev-arplant-042916-041049>

934 Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small
935 number of individuals. *Genetics*, 89(3), 583–590.

936 Nei, M., & Li, W. H. (1979). Mathematical model for studying genetic variation in terms of
937 restriction endonucleases. *Proceedings of the National Academy of Sciences of the*
938 *United States of America*, 76(10), 5269–5273.

939 Peace, C. P., Bianco, L., Troggio, M., van de Weg, E., Howard, N. P., Cornille, A., Durel, C.-
940 E., Myles, S., Migicovsky, Z., Schaffer, R. J., Costes, E., Fazio, G., Yamane, H., van
941 Nocker, S., Gottschalk, C., Costa, F., Chagné, D., Zhang, X., Patocchi, A., ...
942 Vanderzande, S. (2019). Apple whole genome sequences: Recent advances and new
943 prospects. *Horticulture Research*, 6(1), 59. [https://doi.org/10.1038/s41438-019-0141-](https://doi.org/10.1038/s41438-019-0141-7)
944 7

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

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

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

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

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

960 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller,
961 J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool
962 Set for Whole-Genome Association and Population-Based Linkage Analyses.
963 *American Journal of Human Genetics*, 81(3), 559–575.

964 Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication.
965 *Current Biology*, 29(14), R705–R714. <https://doi.org/10.1016/j.cub.2019.05.053>

966 Robinson, J. P., Harris, S. A., & Juniper, B. E. (2001). Taxonomy of the genus *Malus* Mill.
967 (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh. *Plant Syst.*
968 *Evol.*, 226(1), 35–58.

969 Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods
970 for paired-end sequencing improve RADseq-based population genomics. *Molecular*
971 *Ecology*, 28(21), 4737–4754. <https://doi.org/10.1111/mec.15253>

972 Spengler, R. N. (2019). Origins of the Apple: The Role of Megafaunal Mutualism in the
973 Domestication of *Malus* and Rosaceous Trees. *Frontiers in Plant Science*, 10.
974 <https://doi.org/10.3389/fpls.2019.00617>

975 Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., Khan, A., Ban, S., Xu, K.,
976 Cheng, L., Zhong, G.-Y., & Fei, Z. (2020). Phased diploid genome assemblies and
977 pan-genomes provide insights into the genetic history of apple domestication. *Nature*
978 *Genetics*, 52(12), 1423–1432. <https://doi.org/10.1038/s41588-020-00723-9>

979 Teh, B. T., Lim, K., Yong, C. H., Ng, C. C. Y., Rao, S. R., Rajasegaran, V., Lim, W. K., Ong,
980 C. K., Chan, K., Cheng, V. K. Y., Soh, P. S., Swarup, S., Rozen, S. G., Nagarajan, N.,
981 & Tan, P. (2017). The draft genome of tropical fruit durian (*Durio zibethinus*). *Nature*
982 *Genetics*, 49, 1633.

983 Tellier, A., Pfaffelhuber, P., Haubold, B., Naduvilezhath, L., Rose, L. E., Städler, T.,
984 Stephan, W., & Metzler, D. (2011). Estimating Parameters of Speciation Models
985 Based on Refined Summaries of the Joint Site-Frequency Spectrum. *PLoS ONE*, 6(5),
986 e18155. <https://doi.org/10.1371/journal.pone.0018155>

987 Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Poncet, C.,
988 Lateur, M., Houben, P., Ordidge, M., Fernandez-Fernandez, F., Evans, K. M.,

- 989 Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Miranda, C.,
990 Gassmann, J., Kellerhals, M., ... Durel, C.-E. (2016). Analysis of the genetic diversity
991 and structure across a wide range of germplasm reveals prominent gene flow in apple
992 at the European level. *BMC Plant Biology*, *16*(1), 1–20.
993 <https://doi.org/10.1186/s12870-016-0818-0>
- 994 Vavilov, N. I. (1926). Studies on the origin of cultivated plants. *Trudy Byuro. Prikl. Bot.*, *16*,
995 139–245.
- 996 Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Fontana,
997 P., Bhatnagar, S. K., Troggio, M., & Pruss, D. (2010). The genome of the
998 domesticated apple (*Malus x domestica* Borkh.). *Nat Genet*, *42*.
999 <https://doi.org/10.1038/ng.654>
- 1000 Verde, I., Abbott, A. G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T.,
1001 Dettori, M. T., Grimwood, J., Cattonaro, F., Zuccolo, A., Rossini, L., Jenkins, J.,
1002 Vendramin, E., Meisel, L. A., Decroocq, V., Sosinski, B., Prochnik, S., Mitros, T., ...
1003 Rokhsar, D. S. (2013). The high-quality draft genome of peach (*Prunus persica*)
1004 identifies unique patterns of genetic diversity, domestication and genome evolution.
1005 *Nat Genet*, *45*(5), 487–494.
- 1006 Volk, G. M., Cornille, A., Durel, C.-E., & Gutierrez, B. (2021). Botany, Taxonomy, and
1007 Origins of the Apple. In S. S. Korban (Ed.), *The Apple Genome* (pp. 19–32). Springer
1008 International Publishing. https://doi.org/10.1007/978-3-030-74682-7_2
- 1009 Wang, Y., Li, W., Xu, X., Qiu, C., Wu, T., Wei, Q., Ma, F., & Han, Z. (2019). Progress of
1010 Apple Rootstock Breeding and Its Use. *Horticultural Plant Journal*, *5*(5), 183–191.
1011 <https://doi.org/10.1016/j.hpj.2019.06.001>
- 1012 Warschefsky, E. J., Klein, L. L., Frank, M. H., Chitwood, D. H., Londo, J. P., von Wettberg,
1013 E. J. B., & Miller, A. J. (2016). Rootstocks: Diversity, Domestication, and Impacts on

1014 Shoot Phenotypes. *Trends in Plant Science*, 21(5), 418–437.
1015 <https://doi.org/10.1016/j.tplants.2015.11.008>

1016 Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.
1017 <https://doi.org/10.1007/978-0-387-98141-3>

1018 Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., Ming, R., Tai, S., Khan, A. M.,
1019 Postman, J. D., Gu, C., Yin, H., Zheng, D., Qi, K., Li, Y., Wang, R., Deng, C. H.,
1020 Kumar, S., Chagné, D., ... Zhang, S. (2018). Diversification and independent
1021 domestication of Asian and European pears. *Genome Biology*, 19(1), 77.
1022 <https://doi.org/10.1186/s13059-018-1452-y>

1023 Yang, H., Duan, K., & Zhang, W. (2008). Biology and physiology of *Malus hupehensis* for
1024 the apogamic plant resource. *Acta Horticulturae*, 769, 441–447.

1025 Zhang, F., & Batley, J. (2020). Exploring the application of wild species for crop
1026 improvement in a changing climate. *Current Opinion in Plant Biology*, 56, 218–222.
1027 <https://doi.org/10.1016/j.pbi.2019.12.013>

1028 Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B.-H. (2017). Back into the wild—
1029 Apply untapped genetic diversity of wild relatives for crop improvement.
1030 *Evolutionary Applications*, 10(1), 5–24. <https://doi.org/10.1111/eva.12434>

1031 Zhang, Q., Zhang, D., Yu, K., Ji, J., Liu, N., Zhang, Y., Xu, M., Zhang, Y.-J., Ma, X., Liu, S.,
1032 Sun, W.-H., Yu, X., Hu, W., Lan, S.-R., Liu, Z.-J., & Liu, W. (2021). Frequent
1033 germplasm exchanges drive the high genetic diversity of Chinese-cultivated common
1034 apricot germplasm. *Horticulture Research*, 8(1), 1–11.
1035 <https://doi.org/10.1038/s41438-021-00650-8>

1036 Zhou, Z. (1999). The apple genetic resources in China: The wild species and their
1037 distributions, informative characteristics and utilisation. *Genetic Resources and Crop
1038 Evolution*, 46(6), 599–609. <https://doi.org/10.1023/A:1008747709534>

1039 Zohary, D., & Spiegel-Roy, P. (1975). Beginnings of fruit growing in the Old World.

1040 *Science*, 319–327.

1041