

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

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

12

13 *co-first authors

14 #co-directed the work

15 Corresponding authors:

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

17 Dong Zhang, afant@nwafu.edu.cn

18

19 **Abstract**

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

38

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

40 flow.

41

42 **Introduction**

43 It is crucial to understand domestication to unravel the evolutionary processes that shape the
44 divergence of populations, varieties, and species (Gaut, Díez, & Morrell, 2015; Glémin &
45 Bataillon, 2009; Meyer, DuVal, & Jensen, 2012). Several studies have investigated the
46 evolutionary history of fruit crop perennials by using genome-wide data (Aravanopoulos,
47 Ganopoulos, & Tsiftaris, 2015; Besnard, Terral, & Cornille, 2018; Cao et al., 2016; Chagné et
48 al., 2014; Daccord, Celton, Linsmith, Becker, Choisine, Schijlen, van de Geest, Bianco,
49 Micheletti, Velasco, Di Pierro, Gouzy, Rees, et al., 2017; Harrison & Harrison, 2011; Hazzouri et
50 al., 2015; Huang et al., 2016; Michael & VanBuren, 2015; Teh et al., 2017; Verde et al., 2013).
51 However, documentation of the domestication of fruit tree crops still lags behind that of annual
52 crop plants (Burgarella et al., 2019; Gaut et al., 2015; Meyer et al., 2012; Miller & Gross, 2011;
53 Neale, Martínez-García, De La Torre, Montanari, & Wei, 2017). Differences in life-history traits
54 have probably led to marked differences in the mode and speed of evolution between trees and
55 annuals (Gaut et al., 2015), particularly the extent of gene flow during domestication.
56 Documentation of the domestication of fruit tree crops is important not only academically but
57 also for food production. Because of long-term cultivation, fruit trees are susceptible to ever-
58 changing environments due to climate changes and pathogen outbreaks. Sustainable fruit
59 production is now believed to highly rely on untapped wild genetic diversity (Zhang & Batley,
60 2020; Zhang, Mittal, Leamy, Barazani, & Song, 2017).

61 Apple is an iconic tree and major fruit crop grown worldwide. It is also a model species
62 for understanding the role of gene flow during the domestication of clonally propagated
63 perennial crops (Cornille et al., 2019; Cornille, Giraud, Smulders, Roldán-Ruiz, & Gladieux,
64 2014; Peace et al., 2019; Spengler, 2019). The domestication of apple started in Central Asia,
65 from the local Central crab apple *Malus sieversii* (Ledeb.) M. Roem. Then, apple cultivation
66 spread towards the West, Europe and beyond, and the cultivation was assisted by the possibility
67 of propagating interesting genotypes by grafting (Cornille et al., 2012). Thus, grafting has been
68 an important part of apple evolution and breeding (Wang et al., 2019). With the arrival of the
69 cultivated apple in Europe about 1,500 years ago, the local crab apple *Malus sylvestris* Mill.
70 contributed substantially to the genome of some dessert and cider varieties through wild-to-crop
71 introgressions (Cornille et al., 2012). Population genetic approaches in combination with
72 microsatellite markers have demonstrated that wild-to-crop gene flow has played a major role in
73 the evolutionary history of the cultivated apple in Europe (Cornille et al., 2012). However, while
74 the history of the dessert and cider apples is now well documented in the western part of Eurasia
75 (Bina et al., 2021; Cornille et al., 2014; Migicovsky et al., 2021; Peace et al., 2019; Spengler,
76 2019), the domestication history of apples in the East remains unclear. In addition, the history of
77 the apple rootstock that allowed the spread of apple across the world needs to be elucidated.
78 Understanding the origin of the apple rootstock can also provide insights into the role of crop-
79 wild gene flow, as several rootstocks are supposed to be the result of wild-crop hybridization and

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

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

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

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

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

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

138

139 **Materials and Methods**

140 **Plant samples**

141 For this study, fruit flesh samples of 71 *Malus* accessions, representing 48 wild and 23 cultivated
142 apple samples, were collected from repositories in China (Table S1). Additionally, short-read
143 DNA sequences of 97 *Malus* accessions (Duan *et al.*, 2017), namely, 53 wild apple samples and
144 44 apple cultivars, were downloaded (SRA code accession: SRP075497, Table S2). The two
145 datasets, which included 168 accessions with 66 apple cultivars (7 historical Chinese cultivars,
146 30 cultivars from Europe or Western countries, 25 rootstocks, and 4 ornamental cultivars) and
147 102 wild apples (43 *M. sieversii*, 18 *M. baccata*, 11 *M. sylvestris*, and 30 other *Malus*
148 accessions), were merged for subsequent analyses (Table S1).

149

150 **RNA extraction and library construction and sequencing**

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

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

167

168 **Read mapping and SNP calling and filtering**

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

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

185

186 **Suitability of the markers used for phylogenetic, population structure, and demographic** 187 **inferences**

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

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

202

203 **Population structure, genetic diversity, and differentiation among wild and cultivated**
204 **apples**

205 The population structure and admixture among *Malus* accessions were inferred using a model-
206 based clustering method implemented in ADMIXTURE v.1.23 (Alexander, Novembre, & Lange,
207 2009). ADMIXTURE is based on the same statistical model as the one implemented in
208 STRUCTURE (Pritchard, Stephens, & Donnelly, 2000), but it uses a fast numerical optimization
209 algorithm for large SNP datasets to infer the proportion of ancestry of genotypes in K distinct
210 predefined clusters. The number of ancestral populations, K , was varied from 1 to 10, with 20
211 repetitions per K and default settings. The consensus solution for each K value was generated
212 with CLUMPAK (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The amount of
213 additional information explained by increasing K was determined using the cross-validation
214 procedure of ADMIXTURE (Alexander et al., 2009). However, the K identified with the cross-
215 validation procedure often does not correspond to the finest biologically relevant population
216 structure (Cornille et al., 2015; Kalinowski, 2011; Puechmaille, 2016). Therefore, the bar plots
217 were visually checked, and the K value for which all clusters had well-assigned individuals was
218 selected (*i.e.*, that no further well-delimited and relevant clusters could be identified for higher K

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

221 The genetic variation and differentiation among the groups detected with ADMIXTURE
222 were further investigated using three different methods. First, a principal component analysis
223 (PCA) was performed with the `dudi.pca` function of the `ade4` package in R (Dray & Dufour,
224 2007). The first two principal components were plotted with the R package “ggplot2” (Wickham,
225 2009). Second, a neighbor-net tree based on Nei’s distance (Masatoshi Nei, 1978) among
226 individuals was created. The neighbor net was visualized with Splitstree v4 (Huson & Bryant,
227 2006). For the PCA and neighbor-net tree, individuals assigned to a given cluster with a
228 membership coefficient ≥ 0.55 were indicated using the color of each cluster, and admixed
229 individuals (*i.e.*, individuals with a membership coefficient to any given cluster < 0.55) were
230 colored gray. We chose this threshold because both cultivated and wild apples were highly
231 admixed, and the use of a higher cut-off would have removed too many individuals (see Results).
232 For each wild or cultivated population (*i.e.*, group of individuals with membership coefficient $>$
233 0.55 to a given cluster detected with ADMIXTURE), Nei’s diversity (π) (M Nei & Li, 1979),
234 observed (H_O) and expected heterozygosity (H_E) (Masatoshi Nei, 1973), inbreeding coefficient
235 F_{IS} , and genetic differentiation (F_{ST}) between populations were estimated with Stacks population
236 program v2.52 (Rochette, Rivera-Colón, & Catchen, 2019).

237

238 **Detection of gene flow**

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

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

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

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

262

263 **Approximate Bayesian computation**

264 The history of apple domestication was reconstructed using the approximate Bayesian
265 computation (ABC) framework in combination with a coalescent-based inference simulator,
266 fastsimcoal2 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013). Our aims were to
267 infer (1) whether gene flow occurred among certain crop and wild populations during apple
268 domestication and (2) the European and Chinese cultivated apples originated from which wild
269 apple population(s). Previous studies (Cornille, Giraud, Smulders, Roldán-Ruiz, & Gladieux,
270 2014; Cornille et al., 2012; Duan et al., 2017; Migicovsky, Gardner, Richards, Thomas Chao, et
271 al., 2021) have reported that the European dessert apple (*M. domestica*) diverged from *M.*
272 *sieversii*. Therefore, the origin of the European dessert apple was not tested, and it was assumed
273 that it diverged from *M. sieversii* in our models. The scenarios were established according to the
274 results obtained from ADMIXTURE, PCA, and neighbor-net tree as well as the results of *D-*
275 *statistics* and TreeMix. The populations were defined as those detected with ADMIXTURE
276 analyses (*i.e.*, group of individuals assigned with a membership coefficient > 0.55 to a given
277 genetic group).

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

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

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

298 population i (S_i , $i = \{1, 2, 3, 4, 5\}$), mean number of pairwise differences of population i (π_i , i
299 $= \{1, 2, 3, 4, 5\}$), mean number of differences between pairs of populations (π_{i_j} , $i, j = \{1, 2, 3,$
300 $4, 5\}$, $i \neq j$), and pairwise F_{ST} ($F_{ST_{i_j}}$, $i, j = \{1, 2, 3, 4, 5\}$, $i \neq j$). Summary statistics were also
301 added on the basis of the joint frequency spectrum between all pairs of populations (Tellier et al.,
302 2011; Wakeley & Hey, 1997) computed with a home-made script: sites polymorphic in
303 population i , but monomorphic in population j , and vice-versa ($S_{x1_{i_j}}$ and $S_{x2_{i_j}}$, respectively);
304 number of shared polymorphic sites between population i and population j (S_{si_j}); and number of
305 sites showing fixed differences between populations i and j ($S_{f_{i_j}}$).

306 The *abcrf* v.1.7.0 R statistical package (Pudlo et al., 2016) was used to perform the ABC-
307 RF analysis, which provides a classification vote that represents the number of times a scenario
308 is selected as the best one among n trees in the constructed random forest. For each ABC set, the
309 scenario, or a group of scenarios, with the highest number of classification votes was selected as
310 the best scenario, or best group of scenarios, among a total of 500 classification trees (Breiman,
311 2001). The posterior probabilities and prior error rates (*i.e.*, the probability of choosing a wrong
312 group of scenarios when drawing model index and parameter values from the priors of the best
313 scenario) over 10 replicate analyses (Estoup et al., 2018) were computed for each ABC step. The
314 simulated models were also visually checked to be compatible with the observed dataset by
315 projecting the simulated and observed datasets onto the first two linear discriminant analysis
316 (LDA) axes (Pudlo et al., 2016) and whether the observed dataset was within the clouds of
317 simulated datasets. The parameter inferences were then calculated using the final selected model,

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

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

324

325 **Results**

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

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

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

345 ADMIXTURE revealed five main genetic groups (Figure 1a). Cross-validation error
346 monotonically decreased up to $K = 8$ (Figures S7 and S8). At $K > 5$, further substructure was
347 observed, but new clusters were represented by only admixed individuals (Figure S7). Besides, at
348 $K = 5$, the population structure was consistent with the morphological classification of the wild
349 species (Cornille et al., 2019, 2014, 2012; Harris et al., 2002; Robinson, Harris, & Juniper,
350 2001), with *M. sylvestris* (orange), *M. sieversii* (purple), *M. baccata*/*M. hupehensis* (blue), and
351 North American wild apples (green) each forming a distinct group. At $K = 5$, the European apple
352 cultivars (*i.e.*, *M. domestica*, including European dessert and rootstock cultivars) formed a
353 distinct genetic group highly admixed with *M. sylvestris* (Table 1), as shown by Cornille et al.
354 (2012), but also with the *M. baccata*/*M. hupehensis* (blue) genetic group (Table 1). For $K = 5$, the
355 Chinese dessert and rootstocks as well as ornamental Chinese apple species did not form a
356 specific genetic group, but a genetic mixture of the blue (*M. baccata*/*M. hupehensis*) and purple

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

359 Then, we assessed the genetic variations among wild and cultivated apple populations
360 detected with ADMIXTURE. We assigned wild individuals with a membership coefficient > 0.55
361 to a given cluster in the corresponding population (Table 1). Because of the high level of
362 admixture in cultivated apples, we retained all cultivars and defined five cultivated apple groups
363 based on the ADMIXTURE results as well as their historical uses and geographic origins (Tables
364 1 and S1). The use and geographic origins of the cultivars were documented on the basis of the
365 studies performed by Morgan & Richards (1993, 2003) and Urrestarazu *et al.* (2016) for the
366 European-Western cultivars and Zhi-Qin (1999) for the Chinese cultivars. The North American
367 and Siberian wild apples were the most genetically differentiated population (Figure 1b, c, d).
368 *Malus sieversii* individuals formed a distinct group (Figure 1b), close to that of the European
369 dessert apple. The European cultivars (*M. domestica*) were closer to *M. sylvestris* than to the
370 known progenitor *M. sieversii* (Figure 1b, c, d), as detected previously (Cornille *et al.*, 2012;
371 Duan *et al.*, 2017; Migicovsky, Gardner, Richards, Thomas Chao, *et al.*, 2021). *Malus sylvestris*
372 was nested with *M. domestica* in the neighbor-net tree (Figure 1b), but it formed a distinct, but
373 close, gene pool to that of *M. domestica* in the PCA (Figure 1c). Most of the European rootstocks
374 were admixed with the three wild apple species (Figure 1b, c). The Chinese dessert and rootstock
375 formed a bushy structure intermingled with *M. baccata* and *M. hupehensis*. F_{ST} estimates further
376 confirmed that the European cultivars were genetically close to *M. sylvestris* and *M. sieversii*,

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

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

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

411

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

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

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

426

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

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

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

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

461 For the first round of the first ABC set (Figures 3a and S2), the classification votes were
462 the highest 10 times out of 10 for the group of scenarios that assumed gene flow among the
463 European dessert (*M. domestica*) and *M. sylvestris*, European rootstock and *M. sylvestris*, and *M.*
464 *baccata*/*M. hupehensis* (i.e., gene flow modality 4 in Figures S3 and 3, 479.50 votes out of the
465 500 RF trees, posterior probability $P = 1.00$, prior error rate = 0.00, Table S7, Figure S12a,
466 Figure 3a). In the second round, the classification votes were the highest 10 times out of 10 for
467 the scenarios that assumed the European rootstock from *M. sieversii* (422 votes out of the 500
468 RF trees, posterior probability $P = 0.87$, prior error rate = 10.02%, Table S8, Figure S12b). Thus,
469 the first ABC set indicated the European rootstock originated from *M. sieversii* with subsequent
470 gene flow from *M. baccata* and *M. sylvestris*.

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

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

487 The final scenario indicated that the European rootstock diverged from *M. sieversii* and
488 the Chinese cultivars did not diverge from *M. baccata*, but their origin, either from *M. sieversii*
489 or *M. domestica*, remains unclear (Figure 3b). The ABC further confirmed TreeMix gene flow
490 estimates: *M. sylvestris* and *M. baccata* contributed to the European and Chinese dessert and
491 rootstock gene pools (Figure 3b). The model parameters are listed in Table S13, and the posterior
492 estimates are presented in Table S13. However, it should be noted that the credibility intervals
493 were large and the parameter estimates need to be carefully considered. The goodness of fit of
494 the two models (Figures S15 and S16) confirmed our model choice ($P = 0.069$ and $P = 0.07$,
495 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

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

518

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

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

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

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

542

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

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

556 some crosses. Interspecific hybridization has also been observed in date palms (Flowers et al.,
557 2019), grapes (Myles et al., 2011), almonds (Delplancke et al., 2011), and apricots (Groppi et al.,
558 2021; Liu et al., 2019; Q. Zhang et al., 2018). Substantial crop-wild gene flow is mostly due to
559 the self-incompatibility system of fruit trees that favors the selection of the best phenotypes
560 grown from open-pollinated seeds and appears to be a key element in perennial crop evolution
561 (Cornille et al., 2014; Gaut et al., 2015).

562

563 **Concluding remarks and perspectives**

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

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

583

584 **Acknowledgements**

585 We thank the Franco-Chinese Campus France program « Decouverte » 2017, the *Jeunes Talents*
586 *France-Chine* 2019 program, and the ATIP-Avenir CNRS Inserm grant for funding. We also
587 thank Adrien Falce, Olivier Langella and Benoit Johannet for help and support on the INRAE-
588 Génétique Quantitative et Evolution-Le Moulon lab cluster. We thank the INRAE MIGALE
589 bioinformatics platform (<http://migale.jouy.inra.fr>) for providing help and support, in particular
590 Véronique Martin, Eric Montaubon and Valentin Loux. This work was financially supported by
591 the National Key Research and Development Program of China (2018YFD1000101,
592 2019YFD1000803); the Key Technology R & D and Industrialization Demonstration Science
593 and Technology for the Transformation and Upgrading of Apple Industries in Shaanxi Province-
594 Apple Anvil Combinations Evaluation and Seedlings Industrialization (2020zdzx03-01-04); The

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

597

598 **Data Availability**

599 The RNAseq raw data are available at NCBI sequence read archive (SRA) under BioProject
600 accession PRJNA763361

601 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA763361?reviewer=5i8m9p6b4qi4ucg4kf0qe45v>

602 [5j&sort_by=-accession&page=2](https://dataview.ncbi.nlm.nih.gov/object/PRJNA763361?reviewer=5i8m9p6b4qi4ucg4kf0qe45v)). Vcf used for analyses is available at the Zenodo

603 (<https://zenodo.org/record/5513618>). Scripts are available at

604 https://forgemia.inra.fr/amandine.cornille/rnaseq_apple_evolutionary_history/.

605

606 **Author Contributions**

607 DZ, AC, and XC conceived and designed the experiments; AC, DZ and MH obtained the

608 funding; XC, NA, LX, YW sampled the material; XC performed the molecular analyses; JM and

609 CJ performed the RNA extraction; XC and AC analyzed the data. The manuscript was written by

610 AC and XC, with essential inputs from other co-authors.

611

612

613

614 **Tables**

615

616 **Table 1. Mean proportions of nine cultivated and wild apple populations assigned to five**

617 **genetic clusters inferred with ADMIXTURE at $K = 5$ ($N = 115$, 36,200 SNPs).** We divided the

618 cultivated apple samples into five populations according to ADMIXTURE inferences and their

619 origin (Europe or China) and use (dessert or rootstock). We assigned only wild populations with

620 membership coefficient > 0.55 to a given genetic group. Thus, samples were partitioned into nine

621 populations, namely, four wild and five cultivated apple populations.

622

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

625 **Figures**

626 **Figure 1. Genetic structure, variation, differentiation, and relationships of wild and**
627 **cultivated European and Chinese apples.** a. Population structure inferred with ADMIXTURE
628 at $K = 5$ ($N = 115$, 36,200 unlinked synonymous SNPs); each individual is represented by a
629 vertical bar partitioned into K segments that represent the proportions of ancestry of its genome
630 in K clusters. Colors represent the inferred ancestry from K ancestral genetic clusters. b.
631 Neighbor net tree depicting the relationships among the wild and cultivated apple populations
632 identified with ADMIXTURE for $K = 5$. Colors correspond to the genetic groups inferred at $K =$
633 5, and admixed samples (*i.e.*, individuals with membership coefficient < 0.55 to any given gene
634 pool) are in grey. c. Principal component analysis (PCA) representing the genetic variation
635 among wild and cultivated apples, with the total variance explained by the first two components
636 (29.72% and 16.11%, respectively). d. Pairwise genetic differentiation (F_{ST}) among populations
637 identified with ADMIXTURE at $K = 5$ (excluding wild individuals with membership coefficient
638 < 0.55 to a given gene pool).

639

640 **Figure 2. Gene flow inferred with TreeMix ($N = 114$, 36,200 SNPs; Chinese ornamentals**
641 **were removed because of low number of individuals, $N = 4$) among the wild and cultivated**
642 **apple populations detected with ADMIXTURE at $K = 5$.** The North American wild apples
643 were used as the outgroup. The best number of migration events was six (Figure S10). Arrows of
644 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

655 **References**

- 656 Aguirre-Liguori, J. A., Gaut, B. S., Jaramillo-Correa, J. P., Tenaillon, M. I., Montes-Hernández,
657 S., García-Oliva, F., ... Eguiarte, L. E. (2019). Divergence with gene flow is driven by
658 local adaptation to temperature and soil phosphorus concentration in teosinte subspecies
659 (*Zea mays parviglumis* and *Zea mays mexicana*). *Molecular Ecology*, 28(11), 2814–2830.
660 doi: 10.1111/mec.15098
- 661 Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in
662 unrelated individuals. *Genome Research*, 19(9), 1655–1664. doi: 10.1101/gr.094052.109
- 663 Allaby, R. G., Fuller, D. Q., & Brown, T. A. (2008). The genetic expectations of a protracted
664 model for the origins of domesticated crops. *PNAS*, 105(37), 13982–13986. doi:
665 10.1073/pnas.0803780105
- 666 Aravanopoulos, F. A., Ganopoulos, I., & Tsaftaris, A. (2015). Chapter Four—Population and
667 Conservation Genomics in Forest and Fruit Trees. In C. Plomion & A.-F. Adam-Blondon
668 (Eds.), *Advances in Botanical Research* (pp. 125–155). Academic Press. doi:
669 10.1016/bs.abr.2015.04.001
- 670 Besnard, G., Terral, J.-F., & Cornille, A. (2018). On the origins and domestication of the olive: A
671 review and perspectives. *Annals of Botany*, 121(3), 385–403. doi: 10.1093/aob/mcx145
- 672 Bina, H., Yousefzadeh, H., Venon, A., Remoué, C., Rousselet, A., Falque, M., ... Cornille, A.
673 (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

- 714 basis for conservation and breeding programs in apples. *Evolutionary Applications*, 8(4),
715 373–384. doi: 10.1111/eva.12250
- 716 Cornille, A., Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The
717 domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57–65. doi:
718 10.1016/j.tig.2013.10.002
- 719 Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., ... Roldán-Ruiz, I.
720 (2019). A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science*,
721 24(8), 770–782. doi: <https://doi.org/10.1016/j.tplants.2019.05.007>
- 722 Csilléry, K., François, O., & Blum, M. G. B. (2012). abc: An R package for approximate
723 Bayesian computation (ABC): *R package: abc. Methods in Ecology and Evolution*, 3(3),
724 475–479. doi: 10.1111/j.2041-210X.2011.00179.x
- 725 Daccord, N., Celton, J.-M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., ... Bucher, E.
726 (2017). High-quality *de novo* assembly of the apple genome and methylome dynamics of
727 early fruit development. *Nature Genetics*, 49(7), 1099–1106. doi: 10.1038/ng.3886
- 728 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000
729 Genomes Project Analysis Group. (2011). The variant call format and VCFtools.
730 *Bioinformatics*, 27(15), 2156–2158. doi: 10.1093/bioinformatics/btr330
- 731 Delplancke, M., Alvarez, N., Espíndola, A., Joly, H., Benoit, L., Brouck, E., & Arrigo, N. (2011).
732 Gene flow among wild and domesticated almond species: Insights from chloroplast and
733 nuclear markers. *Evolutionary Applications*, 5(4), 317-329.

- 734 Dray, S., & Dufour, A.-B. (2007). The ade4 Package: Implementing the Duality Diagram for
735 Ecologists. *Journal of Statistical Software*, 022(i04). Retrieved from
736 <https://ideas.repec.org/a/jss/jstsof/v022i04.html>
- 737 Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., ... Chen, X. (2017). Genome re-sequencing
738 reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature*
739 *Communications*, 8(1), 249. doi: 10.1038/s41467-017-00336-7
- 740 Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for Ancient Admixture
741 between Closely Related Populations. *Molecular Biology and Evolution*, 28(8), 2239–
742 2252. doi: 10.1093/molbev/msr048
- 743 Eaton, D. A. R., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq
744 datasets. *Bioinformatics*, 36(8), 2592–2594. doi: 10.1093/bioinformatics/btz966
- 745 Estoup, A. A., Raynal, L., Verdu, P., & Marin, J.-M. (2018). Model choice using Approximate
746 Bayesian Computation and Random Forests: Analyses based on model grouping to make
747 inferences about the genetic history of Pygmy human populations. *Journal de La Societe*
748 *Française de Statistique*, 159(3), 167–190.
- 749 Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust
750 Demographic Inference from Genomic and SNP Data. *PLOS Genetics*, 9(10), e1003905.
751 doi: 10.1371/journal.pgen.1003905

- 752 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to
753 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
754 *Resources*, 10(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x
- 755 Feurtey, A., Cornille, A., Shykoff, J. A., Snirc, A., & Giraud, T. (2017). Crop to wild gene flow
756 and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation
757 strategy of the wild apple in Europe. *Evolutionary Applications*, 10(2), 180–188. PMC
758 (PMC5253423). doi: 10.1111/eva.12441
- 759 Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., ...
760 Purugganan, M. D. (2019). Cross-species hybridization and the origin of North African
761 date palms. *Proceedings of the National Academy of Sciences*, 116(5), 1651–1658. doi:
762 10.1073/pnas.1817453116
- 763 Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population
764 structure. *Molecular Ecology Resources*, 17(1), 27–32. doi: 10.1111/1755-0998.12509
- 765 Gao, Y., Liu, F., Wang, K., Wang, D., Gong, X., Liu, L., ... Volk, G. M. (2015). Genetic diversity
766 of *Malus* cultivars and wild relatives in the Chinese National Repository of Apple
767 Germplasm Resources. *Tree Genetics & Genomes*, 11(5), 106. doi: 10.1007/s11295-015-
768 0913-7
- 769 Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of
770 Annual and Perennial Domestication. *Trends in Genetics*, 31(12), 709–719. doi:
771 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->
774 [8137.2009.02884.x](https://doi.org/10.1111/j.1469-8137.2009.02884.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](https://doi.org/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](https://doi.org/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](https://doi.org/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](https://doi.org/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

- 791 into diversification of a fruit tree crop. *Nature Communications*, 6(1), 8824. doi:
792 10.1038/ncomms9824
- 793 Hu, C. G., Honda, C., Kita, M., Zhang, Z., Tsuda, T., & Moriguchi, T. (2002). A simple protocol
794 for RNA isolation from fruit trees containing high levels of polysaccharides and
795 polyphenol compounds. *Plant Molecular Biology Reporter*, 20(1), 69–69. doi:
796 10.1007/BF02801935
- 797 Huang, J., Zhang, C., Zhao, X., Fei, Z., Wan, K., Zhang, Z., ... Li, X. (2016). The Jujube
798 Genome Provides Insights into Genome Evolution and the Domestication of
799 Sweetness/Acidity Taste in Fruit Trees. *PLOS Genetics*, 12(12), e1006433. doi:
800 10.1371/journal.pgen.1006433
- 801 Huson, D. H., & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary
802 Studies. *Molecular Biology and Evolution*, 23(2), 254–267. doi: 10.1093/molbev/msj030
- 803 Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the
804 main genetic clusters within species: Simulations and implications for human population
805 structure. *Heredity*, 106(4), 625–632.
- 806 Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low
807 memory requirements. *Nature Methods*, 12(4), 357–360. doi: 10.1038/nmeth.3317
- 808 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak:
809 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.

- 865 Peace, C. P., Bianco, L., Troggio, M., van de Weg, E., Howard, N. P., Cornille, A., ...
866 Vanderzande, S. (2019). Apple whole genome sequences: Recent advances and new
867 prospects. *Horticulture Research*, 6(1), 59. doi: 10.1038/s41438-019-0141-7
- 868 Picard toolkit. (2019). In *Broad Institute, GitHub repository*. Broad Institute. Retrieved from
869 <http://broadinstitute.github.io/picard/>
- 870 Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from
871 Genome-Wide Allele Frequency Data. *PLOS Genetics*, 8(11), e1002967. doi:
872 10.1371/journal.pgen.1002967
- 873 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using
874 Multilocus Genotype Data. *Genetics*, 155(2), 945–959. doi: 10.1093/genetics/155.2.945
- 875 Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable
876 ABC model choice via random forests. *Bioinformatics*, 32(6), 859–866. doi:
877 10.1093/bioinformatics/btv684
- 878 Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population
879 structure when sampling is uneven: Subsampling and new estimators alleviate the
880 problem. *Molecular Ecology Resources*, 16(3), 608–627. doi: 10.1111/1755-0998.12512
- 881 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., ... Sham, P.
882 C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based
883 Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559–575.

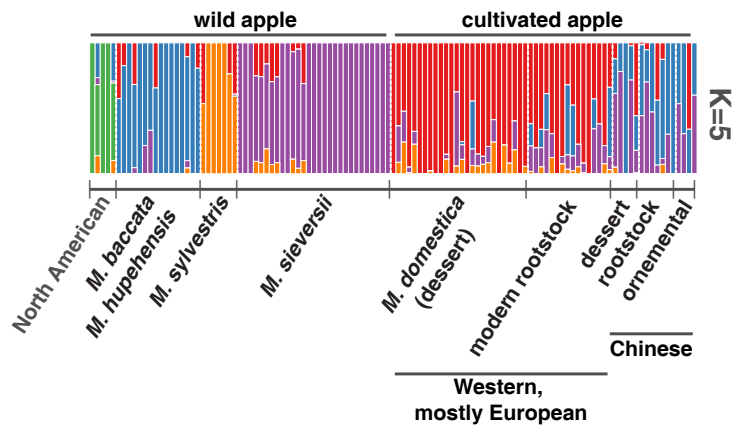
- 884 Purugganan, M. D. (2019). Evolutionary Insights into the Nature of Plant Domestication.
885 *Current Biology*, 29(14), R705–R714. doi: 10.1016/j.cub.2019.05.053
- 886 Robinson, J. P., Harris, S. A., & Juniper, B. E. (2001). Taxonomy of the genus *Malus* Mill.
887 (Rosaceae) with emphasis on the 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

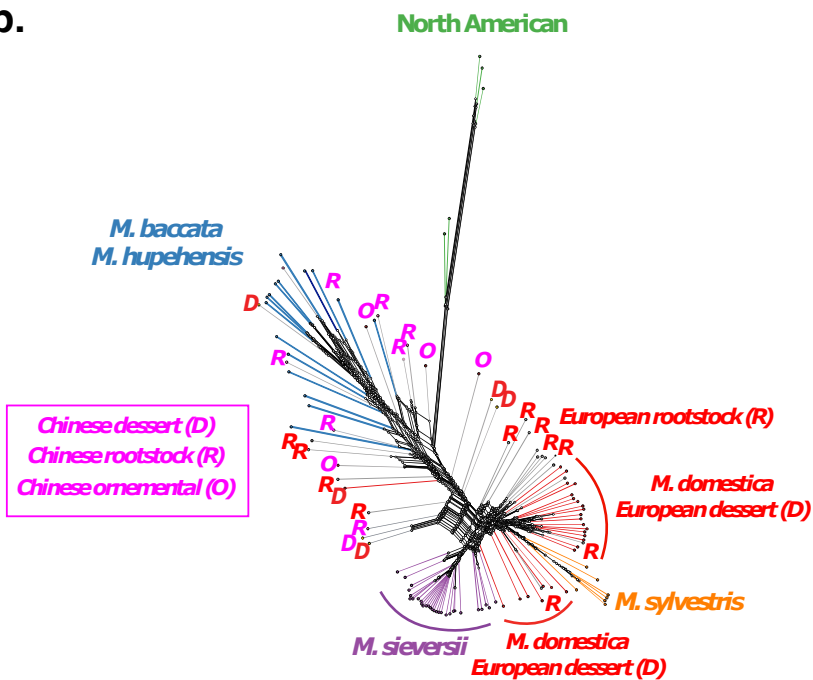
- 922 Volk, G. M., Cornille, A., Durel, C.-E., & Gutierrez, B. (2021). Botany, Taxonomy, and Origins
923 of the Apple. In S. S. Korban (Ed.), *The Apple Genome* (pp. 19–32). Cham: Springer
924 International Publishing. doi: 10.1007/978-3-030-74682-7_2
- 925 Wakeley, J., & Hey, J. (1997). Estimating Ancestral 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

- 941 Yang, H., Duan, K., & Zhang, W. (2008). Biology and physiology of *Malus hupehensis* for the
942 apogamic plant resource. *Acta Horticulturae*, 769, 441–447.
- 943 Yunong, L. (1999). An Investigation and Studies on the Origin and Evolution of *Malus*
944 *domestica* Borkh. In the World. *Acta Horticulturae Sinica*, 4.
- 945 Zhang, F., & Batley, J. (2020). Exploring the application of wild species for crop improvement in
946 a changing climate. *Biotic Interactions. AGRI 2019*, 56, 218–222. doi:
947 10.1016/j.pbi.2019.12.013
- 948 Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B.-H. (2017). Back into the wild—
949 Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary*
950 *Applications*, 10(1), 5–24. doi: <https://doi.org/10.1111/eva.12434>
- 951 Zhang, Q., Zhang, H., Sun, L., Fan, G., Ye, M., Jiang, L., ... Cheng, T. (2018). The genetic
952 architecture of floral traits in the woody plant *Prunus mume*. *Nature Communications*,
953 9(1), 1702. doi: 10.1038/s41467-018-04093-z
- 954 Zhi-Qin, Z. (1999). The apple genetic resources in China: The wild species and their
955 distributions, informative characteristics and utilisation. *Genetic Resources and Crop*
956 *Evolution*, 46(6), 599–609. doi: 10.1023/A:1008747709534
- 957 Zohary, D., & Spiegel-Roy, P. (1975). Beginnings of fruit growing in the Old World. *Science*,
958 319–327.
- 959

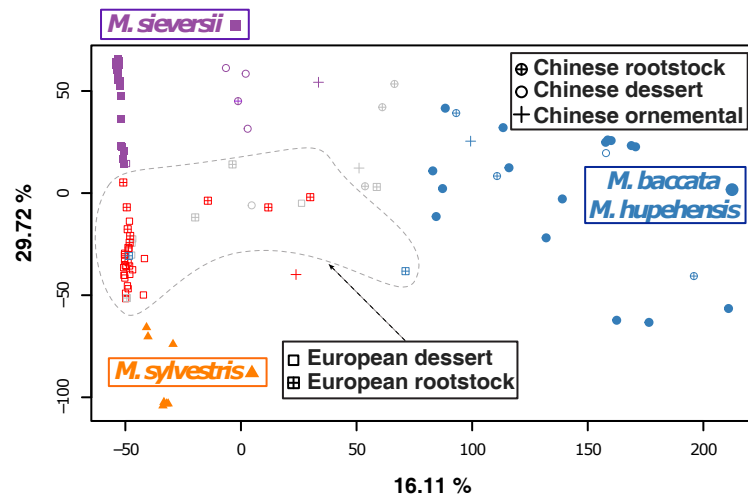
a.



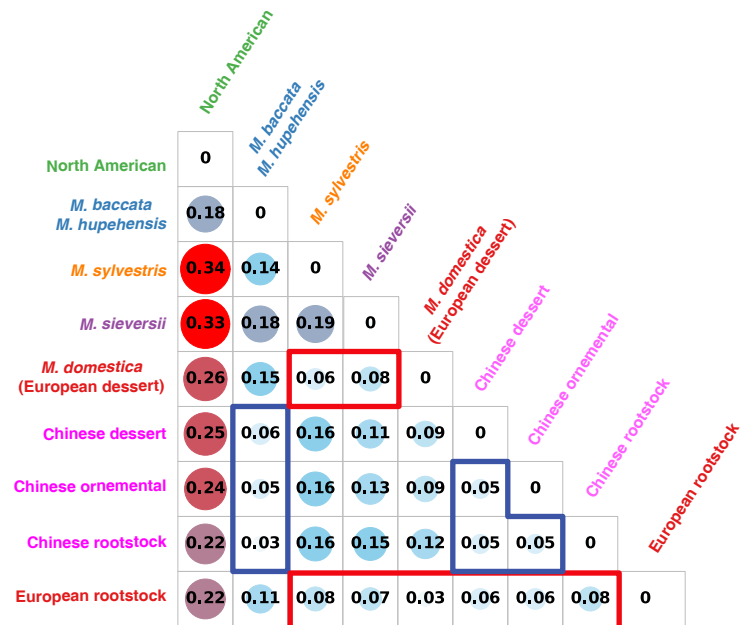
b.

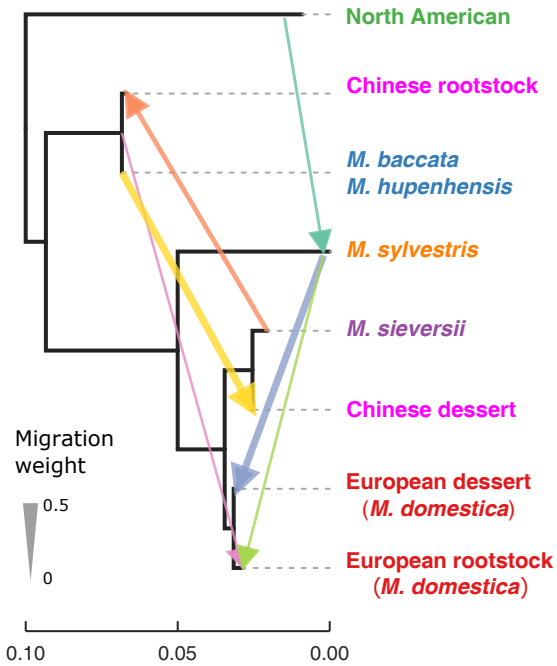


c.



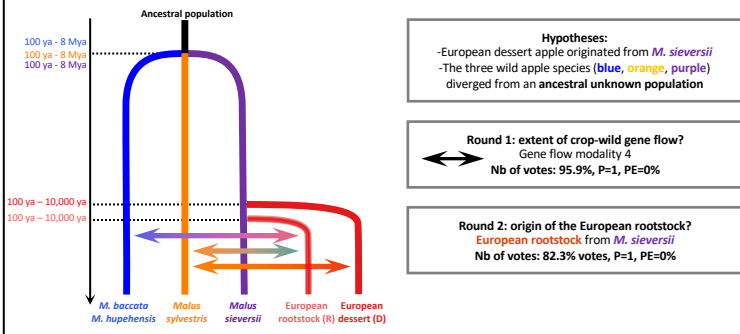
d.





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

