



HAL
open science

Evidence of an additional center of apple domestication in Iran, with contributions from the Caucasian crab apple *Malus orientalis* Uglitzk. to the cultivated apple gene pool. Running head: Apple domestication in the Caucasus and Iran

Hamid Bina, Hamed Yousefzadeh, Anthony Venon, Carine Remoué, Agnès Rousselet, Matthieu Falque, Shadab Faramarzi, Chen Xilong, Jarkyn Samanchina, David Gill, et al.

► To cite this version:

Hamid Bina, Hamed Yousefzadeh, Anthony Venon, Carine Remoué, Agnès Rousselet, et al.. Evidence of an additional center of apple domestication in Iran, with contributions from the Caucasian crab apple *Malus orientalis* Uglitzk. to the cultivated apple gene pool. Running head: Apple domestication in the Caucasus and Iran. *Molecular Ecology*, inPress, 31 (21), pp.5581-5601. 10.1111/mec.16667. hal-03759285

HAL Id: hal-03759285

<https://hal.science/hal-03759285v1>

Submitted on 24 Aug 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.

1 Title: Evidence of an additional center of apple domestication in Iran, with contributions from the
2 Caucasian crab apple *Malus orientalis* Uglitzk. to the cultivated apple gene pool.

3
4 **Running head:** Apple domestication in the Caucasus and Iran

5
6 Bina Hamid¹, Yousefzadeh Hamed², Venon Anthony³, Remoué Carine³, Rousselet Agnès³, Falque
7 Matthieu³, Famarzi Shadab⁴, Xilong Chen³, Jarkyn Samanchina⁵, David Gill⁵, Akylai Kabaeva⁵,
8 Giraud Tatiana⁶, Hossainpour Batool⁷, Abdollahi Hamid⁸, Gabrielyan Ivan⁹, Nersesyan Anush¹⁰,
9 Cornille A.³

10
11 ¹ Department of Forestry, Tarbiat Modares University, Noor, Iran

12 ² Department of Environmental science, Biodiversity Branch, Natural resources faculty, Tarbiat
13 Modares University, Noor, Iran

14 ³ Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190 Gif-sur-
15 Yvette, France

16 ⁴ Department of Plant Production and Genetics, Faculty of Agriculture, Razi University,
17 Kermanshah, Iran

18 ⁵ Fauna & Flora International, Cambridge, UK

19 ⁶ Ecologie Systematique Evolution, Universite Paris-Saclay, CNRS, AgroParisTech, 91190 Gif-
20 sur-Yvette, France

21 ⁷ Iranian Research Organization for Science and Technology (IROST), Institute of Agriculture,
22 Department of Agriculture, Tehran, Iran

23 ⁸ Temperate Fruits Research Center, Horticultural Sciences Research Institute, Agricultural
24 Research, Education and Extension Organization (AREEO), Karaj, Iran

25 ⁹ Department of Palaeobotany, A. Takhtajyan Institute of Botany, Armenian National Academy
26 of Sciences, Acharyan Str.1, 0040 Yerevan, Armenia

27 ¹⁰ Department of Conservation of Genetic Resources of Armenian Flora, A. Takhtajyan Institute
28 of Botany, Armenian National Academy of Sciences, Acharyan Str.1, 0040 Yerevan, Armenia

29
30
31 Corresponding authors: amandine.cornille@cnrs.fr, h.yousefzadeh@modares.ir

32 **Abstract**

33 Divergence processes in crop-wild fruit tree complexes in pivotal regions for plant domestication
34 such as the Caucasus and Iran remain little studied. We investigated anthropogenic and natural
35 divergence processes in apples in these regions using 26 microsatellite markers amplified in 550
36 wild and cultivated samples. We found two genetically distinct cultivated populations in Iran that
37 are differentiated from *Malus domestica*, the standard cultivated apple worldwide. Coalescent-
38 based inferences showed that these two cultivated populations originated from specific
39 domestication events of *Malus orientalis* in Iran. We found evidence of substantial wild-crop and
40 crop-crop gene flow in the Caucasus and Iran, as has been described in apple in Europe. In addition,
41 we identified seven genetically differentiated populations of wild apple (*M. orientalis*), not
42 introgressed by the cultivated apple. Niche modeling combined with genetic diversity estimates
43 indicated that these wild populations likely resulted from range changes during past glaciations.
44 This study identifies Iran as a key region in the domestication of apple and *M. orientalis* as an
45 additional contributor to the cultivated apple gene pool. Domestication of the apple tree therefore
46 involved multiple origins of domestication in different geographic locations and substantial crop-
47 wild hybridization, as found in other fruit trees. This study also highlights the impact of climate
48 change on the natural divergence of a wild fruit tree and provides a starting point for apple
49 conservation and breeding programs in the Caucasus and Iran.

50

51 Key words: apple, Caucasus, Iran, crop-wild gene flow, domestication, fruit tree, climate,
52 introgression.

53

54

55 **Introduction**

56 Crop-wild complexes, *i.e.*, groups of crops and their associated wild species that exchange genes
57 (Anderson, 2005), provide good models for understanding how anthropogenic and natural factors
58 shape population divergence in the presence of gene flow. Indeed, crops are the result of a recent
59 anthropogenic divergence process, *i.e.*, domestication, which began around 10,000 years ago, and
60 which has often been followed by subsequent crop-wild gene flow (Besnard, Terral, & Cornille,
61 2018; Brandenburg et al., 2017; Chen et al., 2019; Cornille, Giraud, Smulders, Roldán-Ruiz, &
62 Gladieux, 2014; Cornille et al., 2012; Diez et al., 2015; Flowers et al., 2019; Gaut, Díez, & Morrell,
63 2015). Conversely, wild species allow the study of natural divergence over a longer timescale.
64 Indeed, wild species have often undergone shifts in their distribution following past climate
65 changes associated with glacial periods, and range contraction has often led to population
66 differentiation and divergence (Excoffier, Foll, & Petit, 2009; Hewitt, 1990; Hewitt, 1996; Jezkova,
67 Olah-Hemmings, & Riddle, 2011; Petit, Bialozyt, Garnier-Géré, & Hampe, 2004; Schmitt, 2007).
68 Understanding the evolutionary processes shaping the natural and anthropogenic divergence of
69 crop-wild complexes is not just an academic exercise: it will also help assess the future status of
70 wild resources. Because of the socio-economic importance of crop plants, protecting the wild
71 relatives of crops, beyond the need for preserving biodiversity (Bacles & Jump, 2011), will allow
72 us to manage the genetic resources for future breeding strategies in the face of global changes (*e.g.*,
73 climate change, emerging diseases) (Bailey-Serres, Parker, Ainsworth, Oldroyd, & Schroeder,
74 2019; Castañeda-Álvarez et al., 2016; Zhang, Mittal, Leamy, Barazani, & Song, 2017).

75 Fruit trees present several historical and biological features that make them fascinating
76 models for investigating anthropogenic and natural divergence with gene flow. Several fruit tree
77 crops are found across the world and sometimes occur in sympatry with their wild relatives
78 (Besnard, Terral, & Cornille, 2018; Cornille et al., 2019; Cornille, Giraud, Smulders, Roldán-Ruiz,
79 & Gladieux, 2014; Delplancke et al., 2011; Liu et al., 2019; Miller & Gross, 2011). Fruit trees are
80 also characterized by high levels of gene flow during divergence, which is to be expected
81 considering the typical life history traits of trees (Cornille et al., 2013; Cornille, Gladieux, &
82 Giraud, 2013; Oddou-Muratorio & Klein, 2008; Petit & Hampe, 2006). Population genetics studies
83 of natural divergence processes associated with the last glacial maximum in Europe, North America
84 and Asia in wind-dispersed trees (*e.g.*, *Abies*, *Pinus*, *Fraxinus*, *Quercus*, *Betula* (Lascoux, Palmé,
85 Cheddadi, & Latta, 2004; Petit et al., 2004)) and animal-dispersed trees (Cornille et al., 2013;

86 George et al., 2015) showed that there were high levels of gene flow between populations and that
87 trees had high dispersal capabilities. These studies also located single (Bai & Spitkovsky, 2010;
88 Tian, Li, Ji, Zhang, & Luo, 2009; Zeng et al., 2011) or multiple (Qiu, Wang, Liu, Shen, & Tang,
89 2011; Tian et al., 2009) glacial refugia where most temperate tree species persisted during the last
90 glacial maximum, and from which populations recolonized higher or lower latitudes during the
91 Holocene post-glacial expansion (Giesecke, Brewer, Finsinger, Leydet, & Bradshaw, 2017).
92 Population genetics and genomics studies also revealed the prominent role of gene flow during the
93 anthropogenic divergence of fruit trees. Domestication of several emblematic fruit tree crops such
94 as grape and apple occurred with substantial crop-crop and wild-crop gene flow and without a
95 bottleneck (Arroyo-García et al., 2006; Cornille et al., 2012; Decroocq et al., 2016; Diez et al.,
96 2015; Duan et al., 2017; Liu et al., 2019; Meyer, Duval, & Jensen, 2012; Myles et al., 2011). These
97 studies thus revealed that the domestication of fruit trees displays different patterns from that of
98 annuals which can be explained by the long lifespan, long juvenile phase and self-incompatibility
99 system of trees (Fuller, 2018; Gaut et al., 2015). However, studies of natural and anthropogenic
100 divergence processes in crop-wild fruit tree complexes are still scarce in the geographic regions
101 that were pivotal in the divergence history of these complexes.

102 The Caucasus ecoregion harbors a remarkable concentration of both economically
103 important plants and their wild relatives, in particular wheat, rye, barley and also fruit trees
104 including walnut, apricot and apple (Asanidze, Akhalkatsi, Henk, Richards, & Volk, 2014;
105 Gabrielian & Zohary, 2004; Vavilov, 1926, 1992; Yousefzadeh, Hosseinzadeh Colagar, Tabari,
106 Sattarian, & Assadi, 2012). This region covers Georgia, Armenia, Azerbaijan, the North Caucasian
107 part of the Russian Federation, the northeastern part of Turkey and the Hyrcanian Mixed Forests
108 region in northwestern Iran (Nakhutsrishvili, Zazanashvili, Batsatsashvili, & Montalvo, 2015;
109 Zazanashvili et al., 2020). Two Pleistocene refugia for temperate plants are recognized in this
110 region (Bina, Yousefzadeh, Ali, & Esmailpour, 2016; Yousefzadeh et al., 2012): the Colchis
111 refugium in the catchment basin of the Black Sea (in the Western Caucasus) and the Hyrcanian
112 refugium at the southern edge of the Caspian Sea. Glacial refugia are known to harbor higher levels
113 of species and genetic diversity (Hewitt, 2004), and this is the case for the Colchis and Hyrcanian
114 refugia. Furthermore, it has been suggested that Iran, with its close proximity to Central Asia - the
115 center of origin of emblematic fruit trees - and its historic position on the Silk Trade Routes, is a
116 possible secondary center of domestication of apple, grape and apricot (Decroocq et al., 2016;

117 Liang et al., 2019; Liu et al., 2019). However, inferences of the natural and anthropogenic
118 divergence history of wild-crop fruit tree complexes in the Caucasus have been limited by the small
119 number of samples (Decroocq et al., 2016; Liu et al., 2019) and/or genetic markers investigated so
120 far (Amirchakhmaghi et al., 2018; Asanidze et al., 2014; Cornille et al., 2013; Gharghani et al.,
121 2010; Myles et al., 2011; Volk & Cornille, 2019; Vouillamoz et al., 2006).

122 The history of domestication of the apple tree in the Caucasus and Iran remains to be
123 resolved. Some authors suggest that the local apple cultivars from several regions of the Caucasus
124 originated from *M. orientalis* (Forsline, Aldwinckle, Dickson, Luby, & Hokanson, 2003;
125 Langenfeld, 1991; Schmitt, 2007) and not from the Central Asian wild apple, *M. sieversii*, the
126 progenitor of *M. domestica* (Cornille et al., 2019; Cornille, Giraud, Smulders, Roldán-Ruiz, &
127 Gladieux, 2014; Duan et al., 2017, 2017; Harris, Robinson, & Juniper, 2002; Migicovsky et al.,
128 2021), or may even be *M. orientalis* grown in gardens and orchards. The Caucasian crab apple,
129 *Malus orientalis* Uglitzk., is an endemic wild apple species in the Caucasus. More specifically, it
130 is found in the Western Caucasus (*i.e.*, the southern part of Russia, northern Anatolia and
131 northwestern Turkey), Georgia, Armenia, the mountainous belt in northern Iran, the Hyrcanian
132 Forests (Büttner, 2001; Rechinger, 1964) and the Zagros Forests in eastern and central Iran
133 (Browicz, 1969; Rechinger, 1964) (Figure 1). This species displays high phenotypic diversity
134 across its distribution range where it occurs as scattered individuals in natural forests or at high
135 altitude in rocky mountains (Fischer & Schmidt, 1938; Rechinger, 1964). The Caucasian crab apple
136 has a high resistance to pests and diseases (Büttner, 2001) and its fruit, of high quality, are variable
137 in size (2-4 cm) and color (green to yellowish) (Cornille et al., 2014). *Malus orientalis* has been
138 present and cultivated in Iran (formerly Persia) long before *M. domestica*, the standard cultivated
139 apple worldwide, was introduced (Abdollahi, 2021; Curtis et al., 2005; Fallahi, Colt, Fallahi, &
140 Chun, 2002; Fallahi, Fallahi, & Mahdavi, 2020; Gharghani et al., 2010; Gharghani et al., 2009;
141 Spengler, 2019). Translations of cuneiform scripts from Persepolis and other historical documents
142 that are related to the Sassanid and Parthian dynasties revealed that edible apples were grown in
143 Iran in the form of dwarf apples three centuries before Christ (BC) (Abdollahi, 2021; Curtis et al.,
144 2005; Fallahi et al., 2002). Cultivation of *M. orientalis* in Persian gardens became more widespread
145 after the invasion and subjugation of large portions of the Persian Empire (334 BC) by Alexander
146 the Great (Bowe, 2004). Dwarf apples became popular in gardens and paradise gardens (derived
147 from the Persian term ‘*pardis*’) around 60 BC in Iran (Fallahi et al., 2002). Apple orchards were

148 cultivated more widely in Iran during the Safavid period (1501-1736), and apple cultivation has
149 continued in Iran until now (Fallahi et al., 2002, 2020). Nowadays, the fruit of several local
150 cultivars that are unique to Iran are harvested in the Caucasus for stewing and processed as juice
151 and other beverages (cider, wine), jelly, syrup, jam and vinegar (Amirchakhmaghi et al., 2018;
152 Büttner, 2001). *Malus domestica* is also currently grown in various regions of the Caucasus, but to
153 a much lesser extent than the local cultivars (Forsline et al., 2003; Gharghani et al., 2010;
154 Gharghani et al., 2009; Langenfeld, 1991; Schmitt, 2007). Therefore, there may have been
155 additional events of domestication of the apple tree in the Caucasus, from *M. orientalis*, or *M.*
156 *domestica*, or apple domestication in the Caucasus may be in its early stage, with the cultivation of
157 *M. orientalis* individuals in orchards. *Malus orientalis* may also be at the origin of wild-to-crop
158 introgressions in the same way as *M. sylvestris*, the European crab apple, contributed to the *M.*
159 *domestica* gene pool (Cornille et al., 2012; Gharghani et al., 2009); hybrid cultivated trees would
160 then occur in gardens and orchards in Iran. The origin of the cultivated apple tree in the Caucasus
161 and Iran therefore remains unresolved despite the large consumption of the apple fruit in these
162 regions.

163 Reciprocally, gene flow from cultivated trees may have led to the presence of hybrid trees
164 in the wild, or even to feral trees. In France, the occurrence of orchards surrounding wild apple
165 populations have directly impacted the levels of crop-to-wild introgression of the European wild
166 apple populations (Cornille et al., 2015). However, the extent of crop-wild gene flow in apples in
167 the Caucasus and Iran has just begun to be investigated. A study suggested that *M. orientalis* only
168 made a minor contribution to Mediterranean *M. domestica* cultivars (Cornille et al., 2012), but
169 lacked in-depth investigation. A population genetics study revealed low levels of crop-to-wild gene
170 flow from *M. domestica* to *M. orientalis* in natural forests of Armenia, Turkey and Russia (Cornille
171 et al., 2013). Population genetic diversity and structure analyses of *M. orientalis* populations from
172 the Western Caucasus and South Caucasus identified three differentiated populations: one in
173 Turkey, one in Armenia and one in Russia (Cornille et al., 2013). On a smaller geographic scale,
174 an east-west genetic subdivision was found across the Hyrcanian Forests in Iran, with five main
175 populations showing admixture (Amirchakhmaghi et al., 2018). However, we still lack a
176 comprehensive view (beyond Armenia) of the genetic diversity and structure of *M. orientalis* to
177 understand its natural divergence history and its contribution to local cultivars.

178 Here, we therefore investigate anthropogenic and natural divergence processes in apples
179 from the Caucasus and the extent of gene flow during divergence. A total of 550 apple trees,
180 comprising local cultivated and wild apple trees from the Caucasus, *M. domestica* apple cultivars
181 and *M. sieversii*, were genotyped using 26 microsatellite markers. In addition, the Siberian wild
182 apple *Malus baccata* was used as an outgroup. From the analysis of this comprehensive genetic
183 dataset, combined with ecological niche modeling approaches, we addressed the following
184 questions: 1) What is the population genetic structure among Caucasian wild and cultivated apples,
185 *M. domestica* and *M. sieversii*, and what are their genetic relationships and levels of genetic
186 diversity? 2) What is the extent of crop-crop, crop-wild and wild-wild gene flow in apple trees in
187 the Caucasus? Are there feral or hybrid trees in the wild? Do humans cultivate wild or hybrid trees?
188 3) Is *M. orientalis* an additional contributor to the domestication history of cultivated apple? 4) Did
189 *M. orientalis* experience past range contraction and expansion associated with the last glacial
190 maximum?

191

192 **Materials and methods**

193 **Sampling, DNA extraction and microsatellite genotyping**

194 Microsatellite genotyping data for *M. orientalis*, *M. sieversii* from Kazakhstan, *M. domestica* and
195 *M. baccata* were from previously published studies, including: 207 *M. orientalis* individuals from
196 Turkey, Armenia and Russia (23 sites, Tables S1 and S2, (Cornille et al., 2013; Cornille et al.,
197 2012)), four apple cultivars from Armenia, 40 “pure” European cultivated *M. domestica* individuals
198 (*i.e.*, not introgressed by *M. sylvestris*) (Tables S1 and S2, (Cornille et al., 2013, 2012)), 20 *M.*
199 *sieversii* individuals from Kazakhstan (Cornille et al., 2013) and 22 *M. baccata* individuals from
200 Russia (Cornille et al., 2012). We collected 257 new samples in 2017 in Iran and in 2018 in
201 Kyrgyzstan (*M. sieversii*) for this study: 167 *M. orientalis* individuals from the Hyrcanian Forests
202 and the Zagros region in Iran (Table S1), 48 local Iranian apple cultivars from the *Seed and Plant*
203 *Improvement Institute* (Karaj, Iran) (Table S1) and 42 *M. sieversii* individuals from Kyrgyzstan.
204 Note that for 18 of the 48 local Iranian samples, fruit size was measured (Table S1). Collections
205 meet the requirements of the recently enacted Nagoya protocol on access to genetic resources and
206 the fair and equitable sharing of benefits. Thus, a total of 550 individuals were analyzed,
207 comprising 374 *M. orientalis*, 48 Iranian and four Armenian apple cultivars, 40 European apple

208 cultivars belonging to *M. domestica*, 62 *M. sieversii* (from Kyrgyzstan and Kazakhstan) and 22 *M.*
209 *baccata* individuals (details are provided in Table S1).

210 DNA from the new samples ($N = 257$) was extracted from dried leaves with the
211 NucleoSpin plant II DNA extraction kit (Macherey & Nagel, Düren, Germany®) following the
212 manufacturer's instructions. Multiplex microsatellite PCR amplifications were performed with a
213 multiplex PCR kit (Qiagen Inc.®) for 26 microsatellite markers as previously described (Cornille
214 et al., 2012; Patocchi, Frei, Frey, & Kellerhals, 2009). Note that on each DNA plate, we included
215 three controls, *i.e.*, one sample of *M. orientalis*, one of *M. sieversii* and one of *M. domestica* for
216 which data were already available (Cornille et al., 2013). Genotypes of the controls were compared
217 with the 2013 dataset. We retained only multilocus genotypes for which $< 20\%$ of the data was
218 missing. The suitability of these markers for population genetics analyses has been demonstrated
219 in previous studies (Cornille et al., 2013; Cornille, Gladieux, & Giraud, 2013; Cornille et al., 2012).

220

221 **Bayesian inferences of population structure, genetic differentiation among wild and** 222 **cultivated apples, and genetic diversity estimates**

223 We investigated the population structure of wild and cultivated apples with the individual-based
224 Bayesian clustering methods implemented in STRUCTURE 2.3.3 (Pritchard, Stephens, &
225 Donnelly, 2000). STRUCTURE uses Markov chain Monte Carlo (MCMC) simulations to infer the
226 proportion of ancestry of genotypes from K distinct clusters. The underlying algorithms attempt to
227 minimize deviations from Hardy–Weinberg within clusters and linkage disequilibrium among loci.
228 We ran STRUCTURE from $K = 1$ to $K = 15$. Based on 10 repeated runs of MCMC sampling from
229 500,000 iterations after a burn-in of 50,000 steps, we determined the amount of additional
230 information explained by increasing K using the ΔK statistic (Evanno, Regnaut, & Goudet, 2005)
231 as implemented in the online post-processing software Structure Harvester (Earl, 2012). However,
232 ΔK provides statistical support for the strongest but not the finest population structure
233 (Puechmaille, 2016). Natural populations can display hierarchical genetic structure, with fine-scale
234 population structure. It has been shown that interpreting K values below the one corresponding to
235 the finest substructure can lead to misinterpretations, especially with uneven sampling among
236 populations (Puechmaille, 2016). Indeed, simulations have shown that STRUCTURE is not
237 reliable for identifying groups that are not at Hardy-Weinberg equilibrium, *i.e.*, when the K is too
238 low (Cunningham et al., 2020; Kalinowski, 2011; Puechmaille, 2016). For instance, when the

239 number of clusters is suboptimal, STRUCTURE does not always group the genetically closest
240 clusters into the same cluster. The problem seems to be caused when the program is forced to
241 cluster individuals into an inappropriately small number of clusters (Cullingham et al., 2020;
242 Kalinowski, 2011; Puechmaille, 2016). A more powerful method for identifying the best K value
243 is to look at the bar plots and choose the highest K value for which no clusters is only represented
244 by admixed individuals, indicating that we have reached the highest K value for which new genuine
245 clusters could be delimited (Chen et al., 2022; Cornille et al., 2013, 2012; Cornille et al., 2015;
246 Fontaine, Gladieux, Hood, & Giraud, 2013; Vercken et al., 2010). The K value we therefore
247 considered corresponded to the finest one, which can be higher than the K value of the strongest
248 population structure level identified by ΔK .

249 We ran STRUCTURE for the whole dataset ($N = 550$) to investigate the population
250 genetic structure among the Caucasian crab apple *M. orientalis*, the Caucasian and Iranian
251 cultivated apples, *M. sieversii*, *M. domestica* and *M. baccata*. We further explored the genetic
252 variation and differentiation among the genetic groups detected with STRUCTURE using three
253 different methods. First, we ran a principal component analysis (PCA) for all individuals with the
254 *dudi.pca* function from the *adeigenet* R package (Jombart & Ahmed, 2011). For the PCA,
255 individuals that were assigned to a given cluster with a membership coefficient ≥ 0.85 were color-
256 coded according to the respective color of each cluster, and admixed individuals (*i.e.*, individuals
257 with a membership coefficient to any given cluster < 0.85) were colored in gray. We chose this
258 threshold based on the distribution of the maximum membership coefficients inferred with
259 STRUCTURE (see results). Second, we generated a neighbor-net with Splitstree v4 (Huson, 1998;
260 Huson & Scornavacca, 2012), using the PCA color code. Third, we explored the relationships
261 among populations identified with STRUCTURE (*i.e.*, clusters of individuals with a membership
262 coefficient ≥ 0.85 to a given cluster) with a neighbor joining (NJ) tree (Huson, 1998; Huson &
263 Scornavacca, 2012). The NJ tree and the neighbor-net tree were built using Nei's standard genetic
264 distance (Nei, 1987) computed among individuals or populations with the Populations software
265 v1.2.31 (<https://bioinformatics.org/populations/>).

266 We computed descriptive population genetic estimates for each population (*i.e.*, each
267 cluster inferred with STRUCTURE, excluding admixed individuals with a membership coefficient
268 < 0.85). We calculated allelic richness (A_R) and private allelic richness (A_P) with ADZE (Szpiech,
269 Jakobsson, & Rosenberg, 2008) using standardized sample sizes of $N_{ADZE} = 7$ (one individual x

270 two chromosomes), corresponding to the minimal number of observations across populations.
271 Heterozygosity (expected and observed), Weir and Cockerham F -statistics and deviations from
272 Hardy–Weinberg equilibrium were calculated with Genepop v4.2 (Raymond & Rousset, 1995;
273 Rousset, 2008).

274
275 **Identification of crop-wild hybrids in the Caucasus and historical wild-wild gene flow in the**
276 **Caucasian crab apple, *M. orientalis***

277 To assess the extent of crop-wild gene flow in the Caucasus, we removed *M. sieversii* and *M.*
278 *baccata* from the dataset (resulting in a dataset with $N = 466$, Table S2) and ran STRUCTURE
279 with the same parameters as above. We defined hybrids resulting from crop-to-wild introgression
280 as *M. orientalis* trees assigned to the *M. domestica* or the Iranian or Armenian cultivated gene pools
281 with a membership coefficient > 0.10 . We defined hybrids resulting from wild-to-crop
282 introgression as cultivars assigned to any of the wild gene pools with a membership coefficient $>$
283 0.10 . We chose this threshold based on the distribution of the maximum membership coefficients
284 inferred with STRUCTURE (see results).

285 After removing crop-wild hybrids, we estimated the extent of historical wild-wild gene
286 flow in *M. orientalis* in the Caucasus using two methods. First, we tested whether there was a
287 significant isolation-by-distance (IBD) pattern. We computed the correlation between $F_{ST}/(1-F_{ST})$
288 and the natural algorithm of geographic distance between geographical sites with SPAGeDI 1.5
289 (Hardy & Vekemans, 2002). Second, for each population, we computed the Nason’s kinship
290 coefficient F_{ij} between pairs of individuals i and j (Loiselle, Sork, Nason, & Graham, 1995) with
291 SPAGeDI 1.5 (Hardy & Vekemans, 2002), and regressed F_{ij} against the natural logarithm of
292 geographic distance, $\ln(d_{ij})$, to obtain the regression slope b . We permuted the spatial position of
293 individuals 9,999 times to test whether there was a significant spatial genetic structure between
294 sites. We then calculated the Sp statistic, defined as $Sp = -bLd/(1-F_N)$, where F_N is the mean F_{ij}
295 between neighboring individuals (Vekemans & Hardy, 2004), and $-bLd$ is the regression slope of
296 F_{ij} against $\ln(d_{ij})$. A low Sp implies low spatial population structure, which suggests high historical
297 gene flow and/or high effective population size.

298 We investigated the directionality of crop-wild, wild-wild and crop-crop gene flow (*i.e.*,
299 immigration rate m_{ji} , the proportion of immigrants in population i that arrive from population j)
300 over the past few generations with a Bayesian assignment method implemented in BAYESASS

301 v3.0 (Wilson & Rannala, 2003). As recommended by Wilson and Rannala (Wilson & Rannala,
302 2003), we run 3×10^6 iterations, sampled every 2,000 iterations with a discarded burn-in of 10^6
303 iterations, delta values were adjusted to 0.12 to ensure that chain swapping occurred in
304 approximately 50% of the total iterations. We repeated five times the analysis with different
305 random number seeds to evaluate consistency of results.

306

307 **Approximate Bayesian computation to reconstruct the domestication history of apple**

308 We used approximate Bayesian computation to test whether Iranian cultivated apples (see results
309 below, we removed Armenian cultivated apples as they were represented by only four samples)
310 were derived from the same domestication event as *M. domestica*. We used the ABC method based
311 on a machine learning tool named “random forest” (ABC-RF) to perform model selection and
312 parameter estimates (Estoup et al., 2018; Pudlo et al., 2016; Raynal et al., 2019). In brief, this
313 method creates a “forest” of bootstrapped decision trees that ranks scenarios based on the summary
314 statistics of the datasets. Some simulations are not used to build the trees and can thus be used to
315 cross-validate the analysis by computing a “prior error rate”. This approach allows the comparison
316 of complex demographic models (Pudlo et al., 2016) by comparing groups of scenarios with a
317 specific type of evolutionary event with other groups with different types of evolutionary events
318 instead of considering all scenarios separately (Estoup et al., 2018).

319 Using the ABC-RF framework, we compared different scenarios of domestication of
320 Iranian cultivars, *i.e.*, with an origin from *M. domestica*, *M. sieversii* from Kazakhstan or *M.*
321 *orientalis*. Populations were defined as the clusters detected with STRUCTURE with the whole
322 dataset ($N = 550$), removing putative hybrid individuals (*i.e.*, individuals with a membership
323 coefficient < 0.85 to any given cluster), and excluding *M. baccata* and Armenian cultivars (see
324 Results part). We removed putative hybrids in order to retrace the divergence history and historical
325 gene flow among populations; more recent admixture events being detectable directly from the
326 STRUCTURE bar plots. We assumed bidirectional gene flow among wild and cultivated apple
327 populations. The model parameters used were: the divergence time between X and Y populations
328 (T_{X-Y}), the effective population size of population X (N_{E-X}) and the migration rate per generation
329 between X and Y populations (m_{X-Y}). Prior values for divergence time were drawn from the log-
330 uniform distribution bounded between the distributions used in the approximate Bayesian
331 computations and are given in Table S3.

332 For all models, microsatellite datasets were simulated for 14 out of the 26 markers that
333 had perfect repeats (Ch01h01, Ch01h10, Ch02c06, Ch02d08, Ch05f06, Ch01f02, Hi02c07,
334 Ch02c09, Ch03d07, Ch04c07, Ch02b03b, MS06g03, Ch04e03, Ch02g01 (Cornille et al., 2013;
335 Cornille et al., 2012; Patocchi, Fernández-Fernández, et al., 2009). We checked that the population
336 structure inferred with 14 microsatellite markers did not differ significantly from the inferences
337 obtained with 26 SSR markers (data not shown). We assumed a generalized stepwise model of
338 microsatellite evolution (Slatkin, 1995). Mutation rates were allowed to vary across loci, with
339 locus-specific mutation rates where μ is the mutation rate per generation, with a log-uniform prior
340 distribution for μ (10^{-4} , 10^{-3}) (Table S3).

341 We used ABCtoolbox (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010)
342 with fastsimcoal 2.5 (Excoffier & Foll, 2011) to simulate datasets, using model parameters drawn
343 from prior distributions (Table S3). We performed 8,000 simulations per scenario. For each
344 simulation, we calculated three summary statistics per population with arlsumstats v 3.5 (Excoffier
345 and Lischer 2010): H , the mean heterozygosity across loci, GW , the mean Garza-Williamson
346 statistic over populations (Garza & Williamson, 2001) and the pairwise F_{ST} between populations
347 (Weir & Cockerham, 1984).

348 We used the abcrf v.1.7.0 R statistical package (Pudlo et al., 2016) to carry out the ABC-
349 RF analysis. This analysis provides a classification vote that represents the number of times a
350 scenario is selected as the best one among n trees in the constructed random forest. For each ABC
351 step, we selected the scenario, or the group of scenarios, with the highest number of classification
352 votes as the best scenario, or best group of scenarios, among a total of 500 classification trees
353 (Breiman, 2001). We computed the posterior probabilities and prior error rates (*i.e.*, the probability
354 of choosing a wrong group of scenarios when drawing model index and parameter values from the
355 priors of the best scenario) over 10 replicate analyses (Estoup et al., 2018) for each ABC step. We
356 also checked visually that the simulated models were compatible with the observed dataset by
357 projecting the simulated and the observed datasets onto the first two linear discriminant analysis
358 (LDA) axes (Pudlo et al., 2016), and by checking that the observed dataset fell within the clouds
359 of simulated datasets. We then calculated parameter inferences using the final selected model. The
360 median and 90% confidence interval (CI, 5%-95%) are given for each model parameter estimate.
361 Note that the ABC-RF approach includes the model checking step that was performed *a posteriori*
362 in previous ABC methods.

363
364 **Spatial pattern of genetic diversity in the Caucasian crab apple**
365 We investigated spatial patterns of diversity in “pure” *M. orientalis*. To this aim, we excluded the
366 crop-to-wild hybrids detected in the second STRUCTURE analysis (*i.e.*, excluding *M. baccata* and
367 *M. sieversii*), as well as *M. domestica* and the Iranian and Armenian cultivars. Spatial patterns of
368 genetic diversity in “pure” *M. orientalis* were visualized by mapping the variation across space
369 (A_R) at 36 sites (*i.e.*, geographic locations for which at least five individuals were successfully
370 genotyped for each marker, Table S2) with the geometry-based inverse distance weighted
371 interpolation in QGIS (Quantum GIS, GRASS, SAGA GIS). We calculated allelic richness (A_R)
372 and private allelic richness (A_P) per site with ADZE (Szpiech et al., 2008) using standardized
373 sample sizes of $N_{ADZE} = 6$ (one individual x two chromosomes), corresponding to the minimal
374 number of observations across sites.

375
376 **Species distribution modeling**
377 The BIOMOD2 R package (Thuiller, Georges, Engler & Breiner, 2016) was used to project past
378 and present distributions of *M. orientalis* following the species distribution modeling methods of
379 Leroy *et al.* (2014). A set of 19 bioclimatic variables from WorldClim.org was used in addition to
380 monthly temperature and precipitation values. Climate data were obtained for past conditions from
381 the last glacial maximum and for the current period between 1960 and 1990. The climate projection
382 at the 2.5-minute spatial resolution from the CCSM4 global climate model was used
383 (<https://www.worldclim.org/data/worldclim21.html#>), as we previously showed that it was the
384 most accurate for apple trees (Cornille et al., 2013). Past and present distributions were projected
385 using three modeling algorithms: a generalized linear model (GLM), a generalized additive model
386 (GAM) and artificial neural networks (ANN).

387 The location of 339 “pure” *M. orientalis* trees (*i.e.*, individuals assigned to a wild apple
388 gene pool with a membership coefficient > 0.9 , see results from the second STRUCTURE analysis)
389 provided the longitude and latitude coordinates. Duplicate data points were removed, resulting in
390 57 presence points for *M. orientalis* (Table S4). We did not have absence data so we randomly
391 selected pseudo-absences to serve as “absence” points for the model, and weighted presence and
392 absence points equally as per Barbet-Massin *et al.* (2012). Models were calibrated using the set of
393 bioclimatic variables and model evaluation was calculated with Jaccard’s indices. Ensemble model

394 forecasting was completed by pulling the average trend of the three modeling algorithms and
395 retaining only the uncorrelated bioclimatic variables with a Pearson correlation threshold greater
396 than 0.75 (Table S5). The model was run again using only variables with high predictive power.

397

398 **Results**

399

400 **Clear population genetic structure among Caucasian wild and cultivated apples, *M.*** 401 ***domestica* and *M. sieversii***

402 The ΔK statistic indicated that the strongest level of population subdivision was at $K = 3$ (Figures
403 S1 and S2). At $K = 3$, a genetic cluster encompassed *M. orientalis* from Turkey, Russia, Armenia
404 and *M. baccata* samples, a second cluster encompassed *M. sieversii* from Central Asia (Kazakhstan
405 and Kyrgyzstan), and a third cluster comprised *M. domestica* and Iranian cultivars. However,
406 several lines of evidence indicated that $K = 3$ did not correspond to a biologically relevant
407 population structure. Indeed, *M. baccata* is known to be highly genetically differentiated from *M.*
408 *orientalis*, *M. sieversii* and *M. domestica* ((Cornille et al., 2014, 2012) and see the PCA below), yet
409 it clustered with *M. orientalis* at $K=3$. In addition, we observed multiple ΔK peaks, up to $K=12$
410 (Figure S1a) and further genetic subdivisions were revealed for $K > 3$ (Figure S2), with well-
411 delimited and biologically meaningful clusters. For example, a population of *M. sieversii* admixed
412 at $K=3$ formed a well-separated cluster without admixture at $K=12$ (the light blue cluster at $K=12$
413 in Figure S1). We therefore visually examined the bar plots (Figure S2) for K values that had ΔK
414 peaks (Figure S1), and chose from these K values the highest one for which all clusters had fully
415 assigned individuals, indicating the finest level of genetic subdivision. At $K = 12$, STRUCTURE
416 identified twelve well-delimited clusters corresponding to species and/or geographic regions
417 (Figures 1 and S2). We therefore considered these twelve clusters as the most relevant genetic
418 structure to investigate the domestication history of the apple tree.

419 Among these twelve clusters were two distinct genetic clusters of *M. sieversii*, one from
420 Kazakhstan and one from Kyrgyzstan (in two shades of light blue, respectively, Figure 1), and a
421 specific genetic cluster of *M. baccata* (orange red). We identified seven distinct genetic groups of
422 *M. orientalis*: a genetic group from the Western Caucasus (Russia, Turkey and northwestern
423 Armenia; orange), a central Armenian group (blue), a southern Armenian group (brown), four
424 genetic groups in Iran corresponding to two gene pools spread across the Zagros Forests (including

425 samples from the Lorestan province in light green and from the Kurdistan province in red), and
426 two gene pools (pink and purple) spread across the Hyrcanian Forests (Figure 1).

427 The *M. domestica* apple cultivars formed a specific genetic group (yellow) that was well
428 separated from the wild *M. orientalis* and the Iranian and Armenian cultivars. The Iranian apple
429 cultivars formed two genetic clusters: one that included only cultivars (dark green) and another
430 (purple) that included cultivated trees and wild individuals from the Hyrcanian Forests. We also
431 detected Iranian cultivated trees that were highly admixed between the Iranian cultivated dark
432 green cluster, the Iranian purple cluster, the *M. domestica* yellow cluster and with two other clusters
433 (red and orange); the latter two included several wild *M. orientalis* individuals from the Zagros
434 Forests in the Kurdistan province in Iran and the Western Caucasus, respectively (Figures 1 and
435 S2). The four Armenian cultivated apple trees fell in the blue and orange clusters, which also
436 included *M. orientalis* trees.

437

438 **Genetic variation and differentiation among wild and cultivated apples**

439 We assigned individuals with a membership coefficient > 0.85 (Figure S3) to a given cluster to the
440 corresponding 12 populations to assess the genetic variation and differentiation among wild and
441 cultivated apples.

442 The PCA showed that *M. baccata* was highly differentiated from the other genetic groups
443 (Figure S4). The PCA without *M. baccata* showed that the European (yellow) and the two Iranian
444 cultivated genetic clusters (green and purple) formed well differentiated gene pools (Figure 2a and
445 S5). *Malus sieversii* from Kazakhstan was closer to the Armenian and Iranian wild apple than *M.*
446 *sieversii* from Kyrgyzstan. The neighbor-net (Figure 2b) confirmed that the apple cultivars (*M.*
447 *domestica* and the purple and dark green Iranian cultivar clusters) were distinct from the wild
448 populations, with the exception of trees from the wild Hyrcanian purple population and *M. sieversii*
449 from Kazakhstan; trees of *M. domestica* and *M. sieversii* from Kazakhstan, and of the wild and
450 cultivated Hyrcanian purple population, were indeed intermingled in the neighbor net (Figure 2b).
451 *Malus orientalis* from the Western Caucasus (orange) and Central and Southern Armenian (brown
452 and blue) were closer to each other than to the Iranian wild apples (light green, red and pink), which
453 clustered together. Most of the NJ tree nodes were not statistically supported, except three nodes
454 that showed a sister relationship of 1) the two cultivated Iranian populations (green and purple), 2)
455 *M. sieversii* from Kyrgyzstan and the Lorestan population, and, 3) among the three Armenian

456 populations (Figure S6). The close relationship between *M. domestica* and *M. sieversii* was not
457 statistically supported but was demonstrated in previous studies (Cornille et al., 2019; Cornille,
458 Giraud, Smulders, Roldán-Ruiz, & Gladieux, 2014; Duan et al., 2017, 2017; Harris, Robinson, &
459 Juniper, 2002; Migicovsky et al., 2021).

460 The two cultivated apple populations from Iran were genetically highly differentiated
461 (mean F_{ST} = 0.24, Table S6), and had lower levels of genetic diversity and fewer private alleles than
462 *M. domestica* ($P < 0.01$, Tables 1 and S7); the purple gene pool displayed the lowest level of genetic
463 diversity and the lowest number of private alleles. The close genetic relationship between apple
464 trees sampled in the Hyrcanian Forests and cultivars from the purple gene pool (Figures 1 and 2),
465 together with the lower levels of genetic diversity in these populations (Table 1), suggest that the
466 trees sampled in the Hyrcanian Forests assigned to the purple genetic group may be feral.
467 Alternatively, the cultivated trees from the purple gene pool may represent the first step of apple
468 domestication in Iran, *i.e.*, wild trees cultivated by humans.

469 STRUCTURE analysis, PCA and neighbor net thus revealed insights into the history of
470 apples in the Caucasus. First, cultivated apples in Iran (dark green and purple clusters) may have
471 resulted from additional domestication events, distinct from that of *M. domestica*, as their
472 respective clusters are not the most closely related in the PCA and the neighbor net. These
473 additional domestication events could have involved *M. orientalis* as a progenitor in Iran;
474 alternatively, cultivated apples in Iran could be derived from *M. sieversii* or *M. domestica* with
475 subsequent local gene flow from *M. orientalis*. Second, the full membership of a substantial
476 number of cultivated trees to genetic clusters of Iranian and Armenian *M. orientalis* suggests that
477 wild trees are grown in orchards for consumption without any strong domestication process and/or
478 that feral individuals occur (*e.g.*, the purple genetic cluster may represent a cultivated group that is
479 also found in the wild as feral or an early domestication step). Third, the high level of admixture in
480 several Iranian cultivated apple trees with wild *M. orientalis* gene pools indicates substantial wild-
481 crop gene flow. Fourth, the spatial population structure of *M. orientalis* in the Caucasus may result
482 from past range contraction and expansion associated with the last glacial maximum. We tested
483 these hypotheses below. First, we estimated the extent of crop-wild and wild-wild gene flow in
484 apple trees in the Caucasus (*i.e.*, excluding *M. baccata* and *M. sieversii*). Second, we inferred the
485 anthropogenic divergence history of the Iranian cultivated apple using a coalescent-based inference
486 method for assessing the probability of different domestication scenarios. Third, we investigated

487 both spatial genetic distribution of genetic diversity and modeled the past and present climatic
488 niches of *M. orientalis* to test whether this species underwent past contraction and expansion in the
489 Caucasus during past glaciations.

490

491 **Substantial crop-wild, crop-crop and wild-wild gene flow in apples in the Caucasus**

492 The second STRUCTURE analysis, focused on the Caucasus, revealed the same genetic clustering
493 for wild apples and *M. domestica* at $K = 9$ (Figures S1 and S7) as in the previous analysis ($K = 12$)
494 (Figures 1 and 2). At $K = 9$, we detected 150 apple genotypes that could be considered hybrids (*i.e.*,
495 individuals assigned to a gene pool with a membership coefficient < 0.9 , this cut-off being chosen
496 on the basis of the distribution of the cumulated membership coefficients for each individual at K
497 $= 9$, Figure S8); these 150 hybrids represented 32% of the total dataset (Table 2). The Iranian
498 cultivars had the highest proportion of hybrids (67%), mostly admixed with the wild and cultivated
499 gene pools from Iran, but also with the *M. domestica* gene pool. Hybrids of the wild Armenian
500 apple were mostly an admixture of the wild Armenian gene pools (*i.e.*, Western, Central and
501 Southern), suggesting local gene flow between crop and wild populations.

502 We removed the 150 hybrids and all apple cultivars (Tables 2 and S2) and focused on
503 gene flow within the “pure” Caucasian crab apple *M. orientalis*. We detected a significant but weak
504 isolation by distance pattern across the Caucasus ($P < 0.001$, $R^2 = 0.07$, Figure S9), suggesting a
505 high level of gene flow among the sampled geographic sites. We estimated Sp values for
506 populations with at least five sampling sites and 20 individuals, *i.e.*, the Hyrcanian (pink) and the
507 Central Armenian (blue) *M. orientalis* populations. Sp values were low but significant (Sp
508 $_{\text{Hyrcanian_pink}} = 0.0076$, $Sp_{\text{Central_blue}} = 0.0027$, $P < 0.001$) suggesting a high level of historical gene
509 flow within populations. However, the Sp value was higher for the Iranian population than for the
510 Armenian population, suggesting a lower level of historical gene flow within the Hyrcanian (pink)
511 population than the Central Armenian (blue) population. Contemporary estimates of gene flow
512 further confirmed the occurrence of wild-to-wild, crop-wild and crop-crop gene flow in the
513 Caucasus (Table S8 and Figure S10). Our results therefore are in agreement with the observations
514 obtained above for the full dataset, suggesting substantial crop-crop, crop-wild and wild-wild gene
515 flow in apples in Iran and the Caucasus.

516

517 **Additional domestication events from *M. orientalis* in Iran inferred with ABC-RF**

518 We defined the populations used in the ABC framework from the clusters detected with
519 STRUCTURE at $K=12$ for 550 wild and cultivated apple accessions, excluding admixed
520 individuals (*i.e.*, with a membership coefficient < 0.85 to any given cluster, Figure S3, as recent
521 gene flow can easily be seen from visual inspection of the bar plots). We excluded *M. baccata* as
522 it was the most genetically differentiated species (Figure S4 and Table S6), and the Armenian
523 cultivars as they were represented by only four individuals. We included two *M. sieversii*
524 populations (from Kazakhstan and Kyrgyzstan), the cultivated (green and purple) populations from
525 Iran, the standard *M. domestica* population and the seven populations of *M. orientalis* (four
526 populations from Iran and three populations from Armenia, Table 1). We also simulated an
527 ancestral population. In total, we therefore simulated thirteen populations (Figures S11 and S12).

528 To keep the number of scenarios tractable, we fixed the divergence histories of the wild
529 apple populations as previously known and inferred from the trees (Figures 2b and S6), and
530 assumed bidirectional gene flow among populations which is congruent with the high levels of
531 crop-wild and crop-crop gene flow shown above. We built eight groups of scenarios to test whether
532 the two Iranian cultivated apple populations diverged from the same wild population or whether
533 they diverged from different populations (Figures S11 and S12). In order to assess the origin of the
534 wild purple population, these eight groups of scenarios were simulated under two alternative
535 hypotheses (Figures S11 and S12, respectively): the wild purple population was either assumed to
536 have diverged from the ancestral population (Figure S11), or the Iranian wild and crop purple
537 populations were assumed to be sister groups (Figure S12). A total of 46 scenarios was therefore
538 simulated, representing 16 groups of hypotheses (Figures S11 and S12).

539 For each step of the ABC-RF approach, the projection of the reference table and the
540 observed datasets onto the two LDA axes that explained most of the variance of the summary
541 statistics showed that the observed data fell within the distribution of the simulated summary
542 statistics (Figure S13), forming distinct clouds for each scenario or groups of scenarios. Visual
543 inspection of the LDA plots indicated that we had enough power to discriminate and select
544 scenarios; results were subsequently validated by the ABC-RF inferences presented below.

545 We used a five-step nested ABC-RF approach. In the first step, we tested the divergence
546 history of the purple wild and cultivated populations. All ten replicate ABC-RF analyses supported
547 a sister relationship between the cultivated and wild purple populations (groups of scenarios 9 to
548 16, Figure S11, an average of 282 votes out of the 500 RF-trees; posterior probabilities = 65%,

549 prior error rate = 0.98%, Table S9). In a second step, we compared the remaining groups (groups
550 9 to 16) to reconstruct the history of the Iranian cultivated apple populations. All ten replicate ABC-
551 RF analyses supported that the two Iranian cultivated apple populations (green and purple)
552 diverged from the same wild apple population (groups 9 to 16, an average of 296 votes out of the
553 500 RF-trees; posterior probabilities = 82%, prior error rate = 1.1%, Table S10). In the third step,
554 we compared the remaining groups (groups 9 to 12) to test whether these Iranian cultivated apple
555 populations diverged from *M. sieversii*, *M. domestica* or *M. orientalis*. All ten replicate ABC-RF
556 analyses supported that the two Iranian cultivated apple populations originated from *M. orientalis*
557 (groups 11 and 12, an average of 290 votes out of the 500 RF-trees; posterior probabilities = 79%,
558 prior error rate = 0.2%, Table S11). In the fourth step, we compared the remaining groups (groups
559 11 and 12) to test whether the two Iranian cultivated apple populations were derived from *M.*
560 *orientalis* from Iran or from Armenia. All ten replicate ABC-RF analyses supported that the two
561 Iranian cultivated apple populations originated from *M. orientalis* from Iran (group 11, an average
562 of 316 votes out of the 500 RF-trees; posterior probabilities = 92%, prior error rate = 0.10%, Table
563 S12). For step five, we compared the last three remaining individual scenarios to select the final
564 model. All ten replicate ABC-RF analyses supported that the two Iranian cultivated apple
565 populations originated from the Iranian red *M. orientalis* population (sc45, an average of 316 votes
566 out of the 500 RF-trees; posterior probabilities = 90%, prior error rate = 0.11%, Table S13). This
567 nested ABC approach avoids comparing too many complex models with too many parameters, and
568 is more powerful than testing all scenarios individually to disentangle the main evolutionary events
569 characterizing demography and divergence (Estoup et al., 2018).

570 ABC-RF inferences therefore provided support for additional apple domestication events
571 in Iran, from the red *M. orientalis* population (Figures 3 and S12). The time of domestication was
572 inferred to be *ca.* 826 years ago (confidence interval (CI): 130-5,030) for the green Iranian crop
573 population and *ca.* 803 ya (CI: 160-3,969) for the purple crop population, respectively; by contrast
574 *M. domestica* was inferred to have diverged from *M. sieversii* *ca.* 3,898 ya (CI: 270-7,670) (Table
575 S14, Figure 3). The divergence time between the wild and the crop purple populations was
576 estimated to be 418 ya (CI: 100-2,500). Note that the confidence intervals were relatively high for
577 some parameters (*e.g.*, migration rates) as illustrated by the high normalized mean absolute error
578 values (NMAE). High posterior probabilities and low *prior* error rates for model choice

579 nevertheless indicate high support for the final model, even if some parameters such as migration
580 rates cannot be precisely estimated.

581
582 **Range expansion and contraction of the wild apple *M. orientalis* in the Caucasus associated**
583 **with the past glaciations**

584 We investigated the spatial variation of genetic diversity and used ecological niche modeling to
585 test the existence of past contraction and expansion of the range of the wild apple in the Caucasus.
586 After removing the 150 crop-wild hybrids identified from the second STRUCTURE analysis, we
587 found a significant positive correlation between longitude and allelic richness (Figure S14, average
588 adjusted $R^2 = 0.66$, $P < 0.0001$) and a significant negative correlation between latitude and allelic
589 richness (Figure S14, average adjusted $R^2 = -0.43$, $P < 0.001$). We also found that the Western
590 (orange) population had the highest level of allelic richness (Tables 1 and S15 and Figure 4). The
591 level of allelic richness was also significantly lower in the wild apple populations from the Zagros
592 Forests, *i.e.*, Lorestan (light green) and Kurdistan (red), than in the other wild populations (Tables
593 1 and S7). The Western Caucasus may therefore have been a glacial refugium in the past. In
594 addition to high levels of genetic diversity in the west, across northeastern Turkey and the Lesser
595 Caucasus mountains in Armenia, we observed local hotspots of genetic diversity in the Hyrcanian
596 Forests and the High Caucasus mountains (Figure 4), suggesting that these mountainous regions
597 may have been potential glacial refugia.

598 Ecological niche modeling further indicated past contraction and expansion of the *M.*
599 *orientalis* range. Model performance as assessed with AUC and TSS was high (Table S15),
600 indicating that the ANN, GLM and GAM algorithms fitted the data well (Allouche, Tsoar, &
601 Kadmon, 2006; Fieldings & Bell, 1997; Monserud & Leemans, 1992). The following six
602 bioclimatic variables were found to have high predictive power: mean diurnal range temperature
603 (bio2), temperature seasonality (bio4), minimum temperature of the wettest quarter (bio8),
604 minimum temperature of the driest quarter (bio9), annual precipitation (bio12) and precipitation of
605 the coldest quarter (bio19). These bioclimatic variables were used to calibrate the models to predict
606 the past and present distribution of *M. orientalis*. The MIROC model (Figure 4) predicted that the
607 areas suitable for *M. orientalis* during the LGM contracted to the western Lesser Caucasus and
608 northeastern Turkey along the Black Sea and into the Colchis region, and also in the eastern part
609 of the Hyrcanian Forests, near Azerbaijan, in agreement with the genetic data (Figure 4). The

610 climatic model therefore suggested that populations of the Caucasian wild apple *M. orientalis* may
611 have been maintained in at least two glacial refugia.

612

613 **Discussion**

614 This study provides insights into the natural and anthropogenic divergence history of apples in
615 hotspots of crop diversity, the Caucasus and Iran. Our results reveal that the evolution of the
616 domesticated apple involved an additional wild species, *M. orientalis*. We identified two distinct
617 cultivated gene pools in Iran that were well differentiated from the standard *M. domestica* apple
618 cultivars and were derived from the same wild Iranian *M. orientalis* population. In addition, several
619 cultivated apple genotypes from the Caucasus were found to belong to the Caucasian *M. orientalis*
620 gene pool, suggesting that local farmers use the Caucasian crabapple for cultivation, which may
621 represent the early stages of domestication. *Malus orientalis* has also contributed to the Caucasian
622 cultivated apple germplasm through wild-to-crop introgression; a similar process has been
623 previously described in apples in Europe (Cornille et al., 2014, 2012). Conversely, crop-to-wild
624 gene flow, which has been reported in Europe (Cornille et al., 2015), was also detected in the
625 Caucasus. We also showed that « pure » *M. orientalis* in this region displays a clear spatial genetic
626 structure with at least seven populations. The combination of niche modeling and population
627 genetics approaches suggested that these populations resulted from range contraction and
628 expansion associated with the past glaciations. Iran is therefore an additional center of apple
629 domestication, where distinct cultivated gene pools evolved from their progenitor *M. orientalis*.
630 *Malus orientalis* has also contributed to cultivated gene pools in the Caucasus and Iran through
631 wild-to-crop gene flow. Our results thereby reinforce the view that apple tree domestication has
632 involved multiple wild apple species from multiple geographic regions, a pattern that has also been
633 found in apricot and pear trees (Groppi, Liu, Cornille, Decroocq, & Decroocq, 2021; Liu et al.,
634 2019; Wu et al., 2018). We also found evidence of substantial hybridization between domestic and
635 wild forms, and this has also been described in other fruit trees (Groppi et al., 2021; Liu et al.,
636 2019; Wu et al., 2018). In addition, this study highlights the impact of climate change on the natural
637 divergence of a wild fruit tree and provides a starting point for apple conservation and breeding
638 programs in the Caucasus and Iran.

639

640 **Additional events of apple domestication from *M. orientalis* in Iran**

641 The occurrence of cultivated populations in Iran which are genetically differentiated from the
642 cultivated apple *M. domestica* and are derived from *M. orientalis*, suggests that Iran is an additional
643 center of apple domestication. The existence of two Iranian cultivated groups (green and purple),
644 genetically closely related, suggests that they diverged from the same wild population, perhaps
645 corresponding to two successive domestication steps, or represent separate domestication events
646 from the same progenitor. We tested these hypotheses and inferred more precisely the Iranian
647 cultivated apple domestication history using coalescent-based methods combined with
648 approximate Bayesian computation. This confirmed that the two Iranian cultivated populations
649 diverged from a *M. orientalis* population in Iran, most likely from the Kurdistan province (red
650 population). Thus, despite the spread of the cultivated apple *M. domestica* along the Silk Trade
651 Routes that crossed Iran and the South Caucasus to reach Turkey (Canepa, 2010; Spengler, 2019),
652 specific domestication events in this region have resulted in local cultivated apple gene pools. It is
653 still not clear, however, whether these two cultivated apple populations represent *de novo*
654 domestication events. Distinct genetic ancestries may not reflect independent *de novo*
655 domestication, but may instead represent a single domestication event with multiple origins (Choi
656 et al., 2017; Gros-Balthazard & Flowers, 2021). The occurrence of independent domestication
657 events in many crop species is a source of ongoing debate (Besnard et al., 2018; Choi et al., 2017;
658 Gros-Balthazard & Flowers, 2021). In apricot and pear trees, there is evidence of independent
659 domestication events in Europe and Asia, with distinct cultivated gene pools of different ancestries
660 which display different genomic regions showing footprints of selection (Groppi et al., 2021; Liu
661 et al., 2019; Wu et al., 2018). Here, our data shows that there are distinct cultivated apple
662 populations, with different geographic origins and which are derived from distinct ancestral gene
663 pools. Microsatellite markers however do not allow to investigate history of selection in
664 populations and have lower power than nucleotide sequences to infer demographic history as
665 mutation models for microsatellite markers are less accurate than for nucleotide sequences (Putman
666 & Carbone, 2014). The low bootstrap support for some of the nodes of the NJ tree and conflicting
667 results with the ABC inferences may be due to the limited power of microsatellite markers to
668 resolve relationships among genetically close populations that exchanged gene flow. The resolution
669 of NJ tree nodes can indeed depends on the number of loci (Koskinen, Hirvonen, Landry, &
670 Primmer, 2004; Takezaki & Nei, 2008), and when admixture among populations occurred (Cavalli-
671 Sforza L. Luca, 1998), as seen in the apple tree. While inferences with ABC coalescent simulate

672 admixture and gene flow among populations, NJ tree does not. ABC inferences are thus more
673 powerful to infer the history of apple domestication, as actually shown by high posterior
674 probabilities and low prior error rates. Sequencing genomes of Iranian and European varieties will
675 likely give a more accurate view of the domestication history of apple in the Caucasus and Iran.
676 First, it would help to get more genetic markers and resolve the relationships among apple
677 populations (X. Chen et al., 2022). Second, the existence of distinct genomic regions under
678 selection in the European and Iranian cultivated gene pools, for instance related to apple taste or
679 size, would be a hallmark of multiple *de novo* domestication events, as shown in apricot and pear
680 trees (Groppi et al., 2021; Wu et al., 2018). In addition, a broader view of the domestication history
681 of the apple tree will require additional sampling of *M. sylvestris*, which is another contributor to
682 the *M. domestica* gene pool (Cornille et al., 2014, 2012), as well as of local varieties across Eurasia,
683 which are missing in this study.

684
685 **Wild-to-crop gene flow from *M. orientalis* to local cultivars and cultivation of the local wild**
686 **apple in the Caucasus**

687 The Caucasian crab apple has considerably contributed to the Caucasian apple germplasm through
688 wild-to-crop introgression. We found evidence of substantial wild-crop and crop-crop gene flow
689 in the Caucasus. Indeed, we found that 41.6% of Iranian cultivars were introgressed by local wild
690 apple gene pools or were an admixture of two cultivated gene pools. This extensive wild-to-crop
691 and crop-crop gene flow is strikingly similar to the pattern documented in apples in Europe. *Malus*
692 *sylvestris* has been shown to be a significant contributor to the *M. domestica* gene pool through
693 recurrent and recent hybridization and introgression events ever since the cultivated apple was
694 introduced in Europe by the Greeks around 1,500 years ago (Cornille et al., 2012). Conversely,
695 substantial crop-to-wild gene flow has been reported from *M. domestica* to *M. sylvestris* (Cornille
696 et al., 2015). Similarly, we found many crop-to-wild hybrids (*M. orientalis* introgressed with *M.*
697 *domestica*) in the forests of Armenia and Iran. Extensive gene flow has been found during the
698 domestication of other fruit trees (Arroyo-García et al., 2006; Cornille et al., 2012; Decroocq et al.,
699 2016; Diez et al., 2015; Duan et al., 2017; Liu et al., 2019; Meyer, Duval, et al., 2012; Myles et al.,
700 2011). The evolutionary consequences of crop-to-wild gene flow remains unclear in fruit trees
701 (Feurtey, Cornille, Shykoff, Snirc, & Giraud, 2017); the extent to which crop-to-wild gene flow
702 may threaten the Caucasian crab apple remains to be tested.

703 Our findings also suggest that farmers in the Caucasus and Iran grow *M. orientalis* rather
704 than *M. domestica*. Indeed, the four Armenian cultivars share their gene pools with the Western
705 (orange) and Central (blue) Caucasian crabapple populations, and some Iranian cultivars are fully
706 assigned to wild populations (red and purple) in Iran. Coalescent-based methods suggest that the
707 wild purple population found in the forest is feral or results from a recent domestication event. It
708 is difficult to discriminate between these two hypotheses using only genotypes, as both predict a
709 sister relationship between the wild and cultivated purple populations. Individual isolated trees
710 belonging to the purple population are found in natural mountainous areas close to fruit trees
711 grafted on *Crataegus*, although they are not themselves cultivated (personal observation. H.
712 Yousefzadeh). In addition, the wild purple population found in the forest had a lower level of
713 genetic diversity than the other wild Iranian populations. Altogether, this may suggest that this
714 population has recently escaped from cultivation. Like domestication, feralization can be seen as a
715 process that is accompanied by admixture and introgression, and can be accompanied by a range
716 of genetic, phenotypic and demographic changes (Mabry, Rowan, Pires, & Decker, 2021). Feral
717 populations have also been documented in other fruit trees, including olive (Besnard et al., 2018),
718 almond (Balaguer-Romano et al., 2021) and apricot (Spengler, Chang, & Tourtellotte, 2013).
719 However, the trees found in the Hyrcanian Forests could also be the result of a first domestication
720 step, as wild apples are widely used for cultivation in Iran (personal communication H.
721 Yousefzadeh and from our results). Additional phenotypic data to identify shifts in phenotypic
722 traits associated with domestication or feralization are required to further disentangle these
723 hypotheses. The cultivation of wild genotypes (here, from the red or purple populations) would not
724 be surprising as *M. orientalis* can grow in mountainous areas, is highly resistant to pests, diseases
725 and drought (Amirchakhmaghi et al., 2018; Büttner, 2001; Höfer et al., 2013; Volk et al., 2008),
726 and has high-quality fruits that have several features that are intermediate between those of *M.*
727 *sylvestris* and *M. sieversii* (Cornille et al., 2014). The use of local wild apples has also been
728 documented in Europe for specific purposes at different times in history (Tardío, Arnal, & Lázaro,
729 2020).

730
731 **The natural divergence history of the Caucasian wild apple was shaped by the past glaciations**
732 The climatic variations since the last glacial maximum, along with the landscape features of the
733 Caucasus, have likely shaped the population structure and diversity of the Caucasian wild apple.

734 We identified seven populations of *M. orientalis* in the Caucasus and Iran: one highly genetically
735 differentiated population in the Western Caucasus (Turkey, Russia and northwestern Armenia),
736 two in Armenia (a southern and a Central population) and four in Iran, including two in the Zagros
737 Forest (one in the Kurdistan province and one in the Lorestan province) and two in the Hyrcanian
738 Forests bordering the southern Caspian Sea. These wild apple populations likely arose from
739 isolation in several refugia during the last glacial maximum. This hypothesis is supported by the
740 observation of a large hotspot of genetic diversity located in the Western Caucasus, and several
741 local hotspots of genetic diversity in Armenia and the Hyrcanian Forests (Zazanashvili et al., 2020).
742 Ecological niche modeling further supported the existence of strong contractions in the range of
743 *M. orientalis* in the Western Caucasus bordering the Black Sea (including the Colchis region), as
744 well as in the Lesser Caucasus and in some parts of the Hyrcanian Forests. Additional samples
745 from the Western Caucasus are required to confirm this hypothesis. These glacial refugia have been
746 described in relation to other species (Parvizi et al., 2019). Indeed, two refugia are recognized in
747 the Caucasus (Tarkhnishvili et al. 2012; Yousefzadeh et al. 2012; Bina et al. 2016; Aradhya et al.
748 2017): a major forest refugium between the western Lesser Caucasus and northeastern Turkey
749 (including the Colchis region in the catchment basin of the Black Sea) and the Hyrcanian refugium
750 at the southern edge of the Caucasus. Further sampling of *M. orientalis* in the far Western and
751 Eastern Caucasus and genotyping with the same microsatellite markers are needed to uncover the
752 role of these two refugia in the divergence history of *M. orientalis*.

753 We also found that the natural divergence history of the Caucasian wild apple involved
754 gene flow across the Caucasus. The weak but significant isolation-by-distance pattern further
755 supports the existence of substantial gene flow among wild apple populations in the Caucasus.
756 Widespread gene flow during divergence associated with the last glacial maximum has been
757 documented for another wild apple relative *M. sylvestris* (Cornille et al., 2013). Calculation of the
758 *Sp* parameter within the largest populations revealed high levels of historical gene flow within
759 populations. *Sp* can also be used to compare the dispersal capacities of *M. orientalis* with that of
760 other plants (Cornille et al., 2013; Cornille, Gladieux, & Giraud, 2013; Vekemans & Hardy, 2004).
761 The Caucasian wild apple showed dispersal capacities that were similar to previous estimates in
762 other wild apple species and lower than that of wind-dispersed trees. Wild apples can disperse over
763 kilometers (Cornille et al., 2015; Feurtey, Cornille, Shykoff, Snirc, & Giraud, 2017). The spatial
764 population structure was somewhat stronger in Iran than in Armenia, suggesting lower levels of

765 gene flow in the Hyrcanian population. In addition to having a stronger genetic structure, the
766 Iranian populations had lower genetic diversity than the Armenian populations, especially the
767 Zagros and Kurdistan populations. In Iran, traditional animal husbandry is a widespread practice
768 (Soofi et al., 2018). Such intensive farming environments may lead to forest fragmentation and
769 may impact wild apple populations, which form low density populations. The future of Iranian wild
770 apple populations, especially in the south where genetic diversity is low, may depend on our ability
771 to protect them through conservation programs. However, *M. orientalis* in the Caucasian ecoregion
772 is highly diverse and further investigations and additional sampling are necessary, as well as a
773 better assessment of its species status and genetic relationship with *M. sieversii*. Here, the
774 microsatellite markers yielded a neighbor-net with intermingled *M. sieversii* and *M. orientalis* from
775 Iran, but the microsatellite markers may lack power to discriminate closely related genetic clusters,
776 as discussed above. A better understanding of the functional genetic diversity and of the ecological
777 relationships of wild apples in their ecosystem are needed for developing and implementing
778 effective conservation genetic strategies in this region (Teixeira & Huber, 2021), and is crucial to
779 better understand crop domestication. This question needs to be resolved urgently, as *M. sieversii*
780 is endangered across its distribution (Omasheva et al., 2017; Zhang, Li, & Li, 2018).

781

782 **Conclusion**

783 We identified Iran as a key center in the evolution and domestication of apple, and *M. orientalis* as
784 a contributor to the evolutionary history of cultivated apples in Iran and in the Caucasus. We also
785 provided insights into the processes underlying the natural divergence of this emblematic wild
786 species and identified several populations that could be the target of conservation programs. Our
787 study revealed the role of gene flow and human practices in natural and anthropogenic divergence
788 processes of an emblematic fruit tree in the Caucasus and Iran. Our results are consistent with those
789 reported for other woody perennials, including apricot (Groppi et al., 2021; Liu et al., 2019), olive
790 (Besnard et al., 2018; Diez et al., 2015), pear (Volk & Cornille, 2019; Wu et al., 2018), and date
791 palm (Flowers et al., 2019). This study also supports the view that the domestication of fruit trees
792 has often been a geographically diffuse and protracted process, involving multiple, geographically
793 disparate origins of domestication, as also reported for example in apricots and pears (Groppi et
794 al., 2021; Wu et al., 2018).

795

796 **Acknowledgements**

797 We thank the Franco-Iranian Campus France program « Gundhishapur » 2016-2018, the Institut
798 Diversité Écologie et Évolution du Vivant (IDEEV), ATIP-Avenir for funding, and the ANR JCJC
799 “PLEASURE”. We thank Bolotbek Tagaev (Sustainable Livelihoods Coordinator of FFI-
800 Kyrgyzstan) for sampling and prospection, Fauna & Flora International and more specifically the
801 Global Trees Campaign (GTC) Program. We also thank Adrien Falce, Olivier Langella and Benoit
802 Johannet for help and support on the INRAE-Génétique Quantitative et Evolution- Le Moulon lab
803 cluster and the genotyping platform GENTYANE INRA UMR 1095. We thank the INRAE
804 MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing help and support, in
805 particular Véronique Martin, Eric Montaubon and Valentin Loux. We also thanks Céline Bellard
806 for her advices for ecological niche modeling analyses.

807

808 **Data Availability**

809 SSR data are available on the ZENODO repository 10.5281/zenodo.6981530. Scripts are available
810 https://forgemia.inra.fr/amandine.cornille/caucasian_apple_domestication_ssr

811

812 **Author Contributions**

813 AC, HY conceived and designed the experiments; AC, HY obtained funding; HB, HY, SF, HBa,
814 IG, AN, AC, JS, DG, AK sampled the material; AV, CR, AR, MF performed the molecular work;
815 AC, HB analyzed the data; AC, HB: wrote the original draft and preparation of the figures; AC,
816 HB, HY, TG, XC, IG, AN and all co-authors: gave critical inputs in final draft and revisions.

817

818 **References**

- 819 Abdollahi, H. (2021). An illustrated review on manifestation of pome fruit germplasm in the
820 historic miniatures of ancient Persia. *Genetic Resources and Crop Evolution*, 68(7),
821 2775–2791. doi: 10.1007/s10722-021-01244-y
- 822 Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution
823 models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*,
824 43(6), 1223–1232.
- 825 Amirchakhmaghi, N., Yousefzadeh, H., Hosseinpour, B., Espahbodi, K., Aldaghi, M., &
826 Cornille, A. (2018). First insight into genetic diversity and population structure of the
827 Caucasian wild apple (*Malus orientalis* Uglitzk.) in the Hyrcanian forest (Iran) and its
828 resistance to apple scab and powdery mildew. *Genetic Resources and Crop Evolution*,
829 65(4), 1255–1268. doi: 10.1007/s10722-018-0611-z
- 830 Anderson, E., . (2005). *Plants, man, and life: Illustrated*. Mineola, N.Y.: Dover Publications. /z-
831 wcorg/.
- 832 Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M. A., Arnold, C., ...
833 Martínez-Zapater, J. M. (2006). Multiple origins of cultivated grapevine (*Vitis vinifera* L.
834 ssp. *Sativa*) based on chloroplast DNA polymorphisms. *Molecular Ecology*, 15(12),
835 3707–3714. doi: 10.1111/j.1365-294X.2006.03049.x
- 836 Asanidze, Z., Akhalkatsi, M., Henk, A. D., Richards, C. M., & Volk, G. M. (2014). Genetic
837 relationships between wild progenitor pear (*Pyrus* L.) species and local cultivars native to
838 Georgia, South Caucasus. *Flora: Morphology, Distribution, Functional Ecology of*
839 *Plants*, 209(9), 504–512. doi: 10.1016/j.flora.2014.06.013
- 840 Bacles, C. F. E., & Jump, A. S. (2011). Taking a tree’s perspective on forest fragmentation
841 genetics. *Trends in Plant Science*, 16(1), 13–18. doi: 10.1016/j.tplants.2010.10.002

842 Bai, X. N., & Spitkovsky, A. (2010). Uncertainties of modeling gamma-ray pulsar light curves
843 using vacuum dipole magnetic field. *Astrophysical Journal*, *715*(2), 1270–1281. doi:
844 10.1088/0004-637X/715/2/1270

845 Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D., & Schroeder, J. I. (2019).
846 Genetic strategies for improving crop yields. *Nature*, *575*(7781), 109–118. doi:
847 10.1038/s41586-019-1679-0

848 Balaguer-Romano, R., Barea-Marquez, A., Ocaña-Calahorra, F. J., Gomez, J. M., Schupp, E. W.,
849 Zhang, J., & de Casas, R. R. (2021). The potential role of synzoochory in the
850 naturalization of almond tree. *Basic and Applied Ecology*, *50*, 97–106. doi:
851 10.1016/j.baae.2020.11.004

852 Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absence for
853 species distribution models: How, where and how many? *Methods in Ecology and*
854 *Evolution*, *in press*(3), 327–338. doi: <https://doi.org/10.1111/j.2041-210X.2011.00172.x>

855 Besnard, G., Terral, J. F., & Cornille, A. (2018). On the origins and domestication of the olive: A
856 review and perspectives. *Annals of Botany*, *121*(3), 385–403. doi: 10.1093/aob/mcx145

857 Bina, H., Yousefzadeh, H., Ali, S. S., & Esmailpour, M. (2016). Phylogenetic relationships,
858 molecular taxonomy, biogeography of *Betula*, with emphasis on phylogenetic position of
859 Iranian populations. *Tree Genetics and Genomes*, *12*(5). doi: 10.1007/s11295-016-1037-4

860 Bowe, Patrick. (2004). *Gardens of the Roman world*. Los Angeles: J. Paul Getty Museum. /z-
861 wcorg/.

862 Brandenburg, J. T., Mary-Huard, T., Rigaiil, G., Hearne, S. J., Corti, H., Joets, J., ... Tenailon,
863 M. I. (2017). Independent introductions and admixtures have contributed to adaptation of
864 European maize and its American counterparts. *PLoS Genetics*, *13*(3), 1–30. doi:
865 10.1371/journal.pgen.1006666

866 Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi:
867 10.1023/A:1010933404324

868 Browicz, K. (1969). Amygdalus. *Flora Iranica*, 66, 166–168.

869 Büttner, R. (2001). Malus. *Hanelt P, Institute of Plant Genetics and Crop Plant Research (Eds)*
870 *Mansfelds Encyclopedia of Agricultural and Horticultural Crops*, 471–482.

871 Canepa, M. (2010). Distant displays of power: Understanding Cross-Cultural Interaction Among
872 the Elites of Rome, Sasanian Iran, and Sui-Tang China. *Ars Orientalis*, 38, 121–154.
873 JSTOR. Retrieved from JSTOR.

874 Castañeda-Álvarez, N. P., Khoury, C. K., Achicanoy, H. A., Bernau, V., Dempewolf, H.,
875 Eastwood, R. J., ... Toll, J. (2016). Global conservation priorities for crop wild relatives.
876 *Nature Plants*, 2(4), 16022. doi: 10.1038/nplants.2016.22

877 Cavalli-Sforza L. Luca. (1998). The Chinese Human Genome Diversity Project. *Proceedings of*
878 *the National Academy of Sciences*, 95(20), 11501–11503. doi: 10.1073/pnas.95.20.11501

879 Chen, J., Li, L., Milesi, P., Jansson, G., Berlin, M., Karlsson, B., ... Lascoux, M. (2019).
880 Genomic data provide new insights on the demographic history and the extent of recent
881 material transfers in Norway spruce. *Evolutionary Applications*, 12(8), 1539–1551. doi:
882 10.1111/eva.12801

883 Chen, X., Cornille, A., An, H., Xing, L., Ma, J., Zhao, C., ... Dong, Z. (2022). The East Asian
884 wild apples, *Malus baccata* (L.) Borkh and *Malus hupehensis* (Pamp.) Rehder., are
885 additional contributors to the genomes of cultivated European and Chinese varieties.
886 *Molecular Ecology*, 0, 1–15. doi: 10.1111/mec.16485

887 Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing (邢禹依), Y.-I., Wing, R. A., & Purugganan, M. D.
888 (2017). The Rice Paradox: Multiple Origins but Single Domestication in Asian Rice.
889 *Molecular Biology and Evolution*, 34(4), 969–979. doi: 10.1093/molbev/msx049

890 Cornille, A., Antolín, F., Garcia, E., Vernesi, C., Fietta, A., Brinkkemper, O., ... Roldán-Ruiz, I.
891 (2019). A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science*,
892 24(8), 770–782. doi: <https://doi.org/10.1016/j.tplants.2019.05.007>

893 Cornille, A., Giraud, T., Bellard, C., Tellier, A., Le Cam, B., Smulders, M. J. M., ... Gladieux, P.
894 (2013). Post-glacial recolonization history of the European crabapple (*Malus sylvestris*
895 Mill.), a wild contributor to the domesticated apple. *Molecular Ecology*, 22(8), 2249–
896 2263.

897 Cornille, A., Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Le Cam, B., ...
898 Giraud, T. (2012). New insight into the history of domesticated apple: Secondary
899 contribution of the European wild apple to the genome of cultivated varieties. *PLoS*
900 *Genet*, 8(5), e1002703.

901 Cornille, Amandine, Feurtey, A., Gélín, U., Ropars, J., Misvanderbrugge, K., Gladieux, P., &
902 Giraud, T. (2015). Anthropogenic and natural drivers of gene flow in a temperate wild
903 fruit tree: A basis for conservation and breeding programs in apples. *Evolutionary*
904 *Applications*, 8(4), 373–384. doi: 10.1111/eva.12250

905 Cornille, Amandine, Giraud, T., Smulders, M. J. M., Roldán-Ruiz, I., & Gladieux, P. (2014). The
906 domestication and evolutionary ecology of apples. *Trends in Genetics*, 30(2), 57–65. doi:
907 10.1016/j.tig.2013.10.002

908 Cornille, Amandine, Gladieux, P., & Giraud, T. (2013). Crop-to-wild gene flow and spatial
909 genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary*
910 *Applications*, 6(5), 737–748. doi: 10.1111/eva.12059

911 Cornille, Amandine, Gladieux, P., Smulders, M. J. M., Roldán-Ruiz, I., Laurens, F., Le Cam, B.,
912 ... Giraud, T. (2012). New insight into the history of domesticated apple: Secondary
913 contribution of the European wild apple to the genome of cultivated varieties. *PLoS*
914 *Genetics*, 8(5). doi: 10.1371/journal.pgen.1002703

915 Cullingham, C. I., Miller, J. M., Peery, R. M., Dupuis, J. R., Malenfant, R. M., Gorrell, J. C., &
916 Janes, J. K. (2020). Confidently identifying the correct K value using the ΔK method:
917 When does $K = 2$? *Molecular Ecology*, 29(5), 862–869. doi: 10.1111/mec.15374

918 Curtis, J., Tallis, N., Andre-Salvini, B., André-Salvini, B., Museum, B., & Armbruster, B. (2005).
919 *Forgotten Empire: The World of Ancient Persia*. University of California Press. Retrieved
920 from <https://books.google.fr/books?id=kJnaKu9DdNEC>

921 Decroocq, S., Cornille, A., Tricon, D., Babayeva, S., Chague, A., Eyquard, J. P., ... Decroocq, V.
922 (2016). New insights into the history of domesticated and wild apricots and its
923 contribution to Plum pox virus resistance. *Molecular Ecology*, 25(19), 4712–4729. doi:
924 10.1111/mec.13772

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

928 Diez, C. M., Trujillo, I., Martinez-Urdiroz, N., Barranco, D., Rallo, L., Marfil, P., & Gaut, B. S.
929 (2015). Olive domestication and diversification in the Mediterranean Basin. *New*
930 *Phytologist*, 206(1), 436–447. doi: 10.1111/nph.13181

931 Duan, N., Bai, Y., Sun, H., Wang, N., Ma, Y., Li, M., ... Chen, X. (2017). Genome re-
932 sequencing reveals the history of apple and supports a two-stage model for fruit
933 enlargement. *Nature Communications*, 8(1). doi: 10.1038/s41467-017-00336-7

934 Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing
935 STRUCTURE output and implementing the Evanno method. *Conservation Genetics*
936 *Resources*, 4(2), 359–361.

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

941 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals
942 using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611–
943 2620.

944 Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic
945 diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, 27(9), 1332–
946 1334. doi: 10.1093/bioinformatics/btr124

947 Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. *Annual*
948 *Review of Ecology, Evolution, and Systematics*, 40(1), 481–501. doi:
949 10.1146/annurev.ecolsys.39.110707.173414

950 Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to
951 perform population genetics analyses under Linux and Windows. *Molecular Ecology*
952 *Resources*, 10(3), 564–567. doi: 10.1111/j.1755-0998.2010.02847.x

953 Fallahi, E., Colt, W. M., Fallahi, B., & Chun, I.-J. (2002). The Importance of Apple Rootstocks
954 on Tree Growth, Yield, Fruit Quality, Leaf Nutrition, and Photosynthesis with an

955 Emphasis on 'Fuji'. *HortTechnology Horttech*, 12(1), 38–44. doi:
956 10.21273/HORTTECH.12.1.38

957 Fallahi, E., Fallahi, P., & Mahdavi, S. (2020). Ancient Urban Gardens of Persia: Concept,
958 History, and Influence on Other World Gardens. *HortTechnology Hortte*, 30(1), 6–12.
959 doi: 10.21273/HORTTECH04415-19

960 Feurtey, A., Cornille, A., Shykoff, J. A., Snirc, A., & Giraud, T. (2017a). Crop-to-wild gene flow
961 and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation
962 strategy of the wild apple in Europe. *Evolutionary Applications*, 10(2), 180–188. PMC
963 (PMC5253423). doi: 10.1111/eva.12441

964 Feurtey, A., Cornille, A., Shykoff, J. A., Snirc, A., & Giraud, T. (2017b). Crop-to-wild gene flow
965 and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation
966 strategy of the wild apple in Europe. *Evolutionary Applications*, 10(2), 180–188. PMC
967 (PMC5253423). doi: 10/f9wm3v

968 Fieldings, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors
969 in conservation presence/absence models. *Environmental Conservation*, 24, 38–49.

970 Fischer, A., & Schmidt, M. (1938). Wilde Kern-und Steinobstarten, ihre Heimat und ihre
971 Bedeutung für die Entstehung der Kultursorten und die Züchtung. *Der Züchter*, 10(6),
972 157–167.

973 Flowers, J. M., Hazzouri, K. M., Gros-Balthazard, M., Mo, Z., Koutroumpa, K., Perrakis, A., ...
974 Purugganan, M. D. (2019). Cross-species hybridization and the origin of North African
975 date palms. *Proceedings of the National Academy of Sciences of the United States of*
976 *America*, 116(5), 1651–1658. doi: 10.1073/pnas.1817453116

977 Fontaine, M. C., Gladieux, P., Hood, M. E., & Giraud, T. (2013). History of the invasion of the
978 anther smut pathogen on *Silene latifolia* in North America. *New Phytologist*, *198*(3), 946–
979 956. doi: 10.1111/nph.12177

980 Forsline, P. L., Aldwinckle, H. S., Dickson, E. E., Luby, J. J., & Hokanson, S. C. (2003). *Of Wild*
981 *Apples of Central Asia* (Vol. 29).

982 Fuller, D. Q. (2018). Long and attenuated: Comparative trends in the domestication of tree fruits.
983 *Vegetation History and Archaeobotany*, *27*(1), 165–176. doi: 10.1007/s00334-017-0659-2

984 Gabrielian, E. T., & Zohary, D. (2004). Wild relatives of food crops native to Armenia and
985 Nakhichevan. *Flora Mediterranea*, *14*, 5–80.

986 Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data
987 from microsatellite loci. *Molecular Ecology*, *10*(2), 305–318. doi: 10.1046/j.1365-
988 294X.2001.01190.x

989 Gaut, B. S., Díez, C. M., & Morrell, P. L. (2015). Genomics and the Contrasting Dynamics of
990 Annual and Perennial Domestication. *Trends in Genetics*, *31*(12), 709–719. doi:
991 10.1016/j.tig.2015.10.002

992 George, J.-P., Konrad, H., Collin, E., Thevenet, J., Ballian, D., Idzajt, M., ... Geburek, T.
993 (2015). High molecular diversity in the true service tree (*Sorbus domestica*) despite
994 rareness: Data from Europe with special reference to the Austrian occurrence. *Annals of*
995 *Botany*, *115*(7), 1105–1115. doi: 10.1093/aob/mcv047

996 Gharghani, A., Zamani, Z., Talaie, A., Fattahi, R., Hajnajari, H., Oraguzie, N. C., ... Gardiner, S.
997 E. (2010). The Role of Iran (Persia) in Apple (*Malus × domestica* Borkh.) Domestication,
998 Evolution and Migration via the Silk Trade Route. *Acta Horticulturae*, (859), 229–236.

999 Gharghani, Ali, Zamani, Z., Talaie, A., Oraguzie, N. C., Fatahi, R., Hajnajari, H., ... Gardiner, S.
1000 E. (2009). Genetic identity and relationships of Iranian apple (*Malus × domestica* Borkh.)

1001 cultivars and landraces, wild *Malus* species and representative old apple cultivars based
1002 on simple sequence repeat (SSR) marker analysis. *Genetic Resources and Crop Evolution*,
1003 56(6), 829–842.

1004 Giesecke, T., Brewer, S., Finsinger, W., Leydet, M., & Bradshaw, R. H. W. (2017). Patterns and
1005 dynamics of European vegetation change over the last 15,000 years. *Journal of*
1006 *Biogeography*, 44(7), 1441–1456. doi: 10.1111/jbi.12974

1007 Groppi, A., Liu, S., Cornille, A., Decroocq, S., & Decroocq, D. (2021). Population genomics of
1008 apricots unravels domestication history and adaptive events. *Nature Communications*, 12,
1009 3956.

1010 Gros-Balthazard, M., & Flowers, J. M. (2021). A Brief History of the Origin of Domesticated
1011 Date Palms. In J. M. Al-Khayri, S. M. Jain, & D. V. Johnson (Eds.), *The Date Palm*
1012 *Genome, Vol. 1: Phylogeny, Biodiversity and Mapping* (pp. 55–74). Cham: Springer
1013 International Publishing. doi: 10.1007/978-3-030-73746-7_3

1014 Hardy, O. J., & Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyse spatial
1015 genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2(4),
1016 618–620.

1017 Harris, S. A., Robinson, J. P., & Juniper, B. E. (2002). Genetic clues to the origin of the apple.
1018 *Trends Genet*, 18(8), 426–430.

1019 Hewitt, G. M. (1990). Divergence and speciation as viewed from an insect hybrid zone.
1020 *Canadian Journal of Zoology*, 68(8), 1701–1715. doi: 10.1139/z90-251

1021 Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary.
1022 *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*,
1023 359(1442), 183–195. doi: 10.1098/rstb.2003.1388

1024 Hewitt, Godfrey M. (1996). Some genetic consequences of ice ages, and their role in divergence
1025 and speciation. *Biological Journal of the Linnean Society*, 58(3), 247–276. doi:
1026 10.1006/bijl.1996.0035

1027 Höfer, M., Flachowsky, H., Hanke, M.-V., Seměnov, V., Šlāvas, A., Bandurko, I., ... Alexanian,
1028 S. (2013). Assessment of phenotypic variation of *Malus orientalis* in the North Caucasus
1029 region. *Genetic Resources and Crop Evolution*, 60(4), 1463–1477. doi: 10.1007/s10722-
1030 012-9935-2

1031 Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics*
1032 (*Oxford, England*), 14(1), 68–73.

1033 Huson, D. H., & Scornavacca, C. (2012). Dendroscope 3: An interactive tool for rooted
1034 phylogenetic trees and networks. *Systematic Biology*, 61(6), 1061–1067.

1035 Jezkova, T., Olah-Hemmings, V., & Riddle, B. R. (2011). Niche shifting in response to warming
1036 climate after the last glacial maximum: Inference from genetic data and niche assessments
1037 in the chisel-toothed kangaroo rat (*Dipodomys microps*). *Global Change Biology*, 17(11),
1038 3486–3502. doi: 10.1111/j.1365-2486.2011.02508.x

1039 Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide
1040 SNP data. *Bioinformatics*, 27(21), 3070–3071. doi: 10.1093/bioinformatics/btr521

1041 Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the
1042 main genetic clusters within species: Simulations and implications for human population
1043 structure. *Heredity*, 106(4), 625–632.

1044 Koskinen, M. T., Hirvonen, H., Landry, P. A., & Primmer, C. R. (2004). The benefits of
1045 increasing the number of microsatellites utilized in genetic population studies: An
1046 empirical perspective. *Hereditas*, 141(1), 61–67. doi: 10.1111/j.1601-5223.2004.01804.x

1047 Langenfeld, W. T. (1991). Apple trees. *Morphological Evolution, Phylogeny, Geography and*
1048 *Systematics. Riga (Zinatne), 232.*

1049 Lascoux, M., Palmé, A. E., Cheddadi, R., & Latta, R. G. (2004). Impact of Ice Ages on the
1050 genetic structure of trees and shrubs. *Philosophical Transactions of the Royal Society B:*
1051 *Biological Sciences, 359(1442), 197–207.* doi: 10.1098/rstb.2003.1390

1052 Leroy, B., Bellard, C., Dubos, N., Colliot, A., Vasseur, M., Courtial, C., ... Ysnel, F. (2014).
1053 Forecasted climate and land use changes, and protected areas: The contrasting case of
1054 spiders. *Diversity and Distributions, 20(6), 686–697.* doi: 10.1111/ddi.12191

1055 Liang, Z., Duan, S., Sheng, J., Zhu, S., Ni, X., Shao, J., ... Dong, Y. (2019). Whole-genome
1056 resequencing of 472 *Vitis* accessions for grapevine diversity and demographic history
1057 analyses. *Nature Communications, 10(1), 1190.* doi: 10.1038/s41467-019-09135-8

1058 Liu, S., Cornille, A., Decroocq, S., Tricon, D., Chague, A., Eyquard, J., ... Decroocq, V. (2019).
1059 The complex evolutionary history of apricots: Species divergence, gene flow and multiple
1060 domestication events. In *Molecular Ecology.* doi: 10.1111/mec.15296

1061 Loiselle, B. A., Sork, V. L., Nason, J., & Graham, C. (1995). Spatial genetic structure of a
1062 tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of*
1063 *Botany, 82(11), 1420–1425.*

1064 Mabry, M. E., Rowan, T. N., Pires, J. C., & Decker, J. E. (2021). Feralization: Confronting the
1065 Complexity of Domestication and Evolution. *Trends in Genetics, 37(4), 302–305.* doi:
1066 10.1016/j.tig.2021.01.005

1067 Meyer, R. S., Duval, A. E., & Jensen, H. R. (2012). Patterns and processes in crop domestication:
1068 An historical review and quantitative analysis of 203 global food crops. *New Phytologist,*
1069 *196(1), 29–48.* doi: 10.1111/j.1469-8137.2012.04253.x

1070 Migicovsky, Z., Gardner, K. M., Richards, C., Chao, C. T., Schwaninger, H. R., Fazio, G., ...
1071 Myles, S. (2021). Genomic consequences of apple improvement. *Horticulture Research*,
1072 8(1), 1–13. doi: 10.1038/s41438-020-00441-7

1073 Miller, A. J., & Gross, B. L. (2011). From forest to field: Perennial fruit crop domestication.
1074 *American Journal of Botany*, 98(9), 1389–1414. doi: 10.3732/ajb.1000522

1075 Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa
1076 statistic. *Ecological Modelling*, 62(4), 275–293.

1077 Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., ... Buckler, E.
1078 S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the*
1079 *National Academy of Sciences of the United States of America*, 108(9), 3530–3535. doi:
1080 10.1073/pnas.1009363108

1081 Nakhutsrishvili, G., Zazanashvili, N., Batsatsashvili, K., & Montalvo, CS. (2015). Colchic and
1082 Hyrcanian forests of the Caucasus: Similarities, differences and conservation status. *Flora*
1083 *Mediterranea*, 25, 185–192.

1084 Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press. Retrieved from
1085 <https://books.google.es/books?id=UhRSsLkxDgC>

1086 Oddou-Muratorio, S., & Klein, E. K. (2008). Comparing direct vs. Indirect estimates of gene
1087 flow within a population of a scattered tree species. *Molecular Ecology*, 17(11), 2743–
1088 2754.

1089 Omasheva, M. Y., Flachowsky, H., Ryabushkina, N. A., Pozharskiy, A. S., Galiakparov, N. N., &
1090 Hanke, M.-V. (2017). To what extent do wild apples in Kazakhstan retain their genetic
1091 integrity? *Tree Genetics & Genomes*, 13(3), 52. doi: 10.1007/s11295-017-1134-z

1092 Parvizi, E., Keikhosravi, A., Naderloo, R., Solhjoui-Fard, S., Sheibak, F., & Schubart, C. D.
1093 (2019). Phylogeography of *Potamon ibericum* (Brachyura: Potamidae) identifies

1094 Quaternary glacial refugia within the Caucasus biodiversity hot spot. *Ecology and*
1095 *Evolution*, 9(8), 4749–4759. doi: 10.1002/ece3.5078

1096 Patocchi, A., Fernández-Fernández, F., Evans, K., Gobbin, D., Rezzonico, F., Boudichevskaia,
1097 A., ... van de Weg, W. (2009). Development and test of 21 multiplex PCRs composed of
1098 SSRs spanning most of the apple genome. *Tree Genetics and Genomes*, 5(1), 211–223.

1099 Patocchi, A., Frei, A., Frey, J. E., & Kellerhals, M. (2009). Towards improvement of marker
1100 assisted selection of apple scab resistant cultivars: *Venturia inaequalis* virulence surveys
1101 and standardization of molecular marker alleles associated with resistance genes.
1102 *Molecular Breeding*, 24(4), 337–347. doi: 10.1007/s11032-009-9295-6

1103 Petit, R. J., Bialozyt, R., Garnier-Géré, P., & Hampe, A. (2004). Ecology and genetics of tree
1104 invasions: From recent introductions to Quaternary migrations. *Forest Ecology and*
1105 *Management*, 197(1–3), 117–137. doi: 10.1016/j.foreco.2004.05.009

1106 Petit, R. J., & Hampe, A. (2006). Some Evolutionary Consequences of Being a Tree. *Annual*
1107 *Review of Ecology, Evolution, and Systematics*, 37(1), 187–214. doi:
1108 10.1146/annurev.ecolsys.37.091305.110215

1109 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
1110 multilocus genotype data. *Genetics*, 155(2), 945–959.

1111 Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., & Robert, C. P. (2016). Reliable
1112 ABC model choice via random forests. *Bioinformatics*, 32(6), 859–866. doi:
1113 10.1093/bioinformatics/btv684

1114 Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population
1115 structure when sampling is uneven: Subsampling and new estimators alleviate the
1116 problem. *Molecular Ecology Resources*, 16(3), 608–627. doi: 10.1111/1755-0998.12512

1117 Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite
1118 data for population genetic studies. *Ecology and Evolution*, 4(22), 4399–4428. doi:
1119 10.1002/ece3.1305

1120 Qiu, J., Wang, L., Liu, M., Shen, Q., & Tang, J. (2011). An efficient and simple protocol for a
1121 PdCl₂-ligandless and additive-free Suzuki coupling reaction of aryl bromides.
1122 *Tetrahedron Letters*, 52(48), 6489–6491. doi: 10.1016/j.tetlet.2011.09.115

1123 Raymond, M., & Rousset, F. (1995). An exact test for population differentiation. *Evolution*,
1124 49(6), 1280–1283.

1125 Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P., & Estoup, A. (2019). ABC random
1126 forests for Bayesian parameter inference. *Bioinformatics*, 35(10), 1720–1728. doi:
1127 10.1093/bioinformatics/bty867

1128 Rechinger, K. H. (1964). Flora Iranica, Akademische Druck-und Verlagsanstalt Graz. *University*
1129 *of Tehran, Iran*, 549.

1130 Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for
1131 Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106.

1132 Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends.
1133 *Frontiers in Zoology*, 4, 1–13. doi: 10.1186/1742-9994-4-11

1134 Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele
1135 frequencies. *Genetics*, 139(1), 457–462. PubMed (7705646). doi:
1136 10.1093/genetics/139.1.457

1137 Soofi, M., Ghoddousi, A., Zeppenfeld, T., Shokri, S., Soufi, M., Jafari, A., ... Waltert, M. (2018).
1138 Livestock grazing in protected areas and its effects on large mammals in the Hyrcanian
1139 forest, Iran. *Biological Conservation*, 217, 377–382. doi: 10.1016/j.biocon.2017.11.020

1140 Spengler, R.N. (2019). *Fruit from the Sands: The Silk Road Origins of the Foods We Eat*.
1141 University of California Press. Retrieved from
1142 <https://books.google.fr/books?id=fr6SDwAAQBAJ>

1143 Spengler, Robert N., Chang, C., & Tourtellotte, P. A. (2013). Agricultural production in the
1144 Central Asian mountains: Tuzusai, Kazakhstan (410–150 b.c.). *Journal of Field*
1145 *Archaeology*, 38(1), 68–85. doi: 10.1179/0093469012Z.000000000037

1146 Spengler, Robert Nicholas. (2019). Origins of the Apple: The Role of Megafaunal Mutualism in
1147 the Domestication of Malus and Rosaceous Trees. *Frontiers in Plant Science*, 10. doi:
1148 10.3389/fpls.2019.00617

1149 Szpiech, Z. A., Jakobsson, M., & Rosenberg, N. A. (2008). ADZE: a rarefaction approach for
1150 counting alleles private to combinations of populations. *Bioinformatics*, 24(21), 2498–
1151 2504.

1152 Takezaki, N., & Nei, M. (2008). Empirical Tests of the Reliability of Phylogenetic Trees
1153 Constructed With Microsatellite DNA. *Genetics*, 178(1), 385–392. doi:
1154 10.1534/genetics.107.081505

1155 Tardío, J., Arnal, A., & Lázaro, A. (2020). Ethnobotany of the crab apple tree (*Malus sylvestris*
1156 (L.) Mill., Rosaceae) in Spain. *Genetic Resources and Crop Evolution*. doi:
1157 10.1007/s10722-020-01026-y

1158 Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in
1159 conservation genetics. *Proceedings of the National Academy of Sciences*, 118(10),
1160 e2015096118. doi: 10.1073/pnas.2015096118

1161 Thuiller, W., Georges, D., Engler, R., & Breiner, F. (2016). *Biomod2: Ensemble platform for*
1162 *species distribution modeling. R package version 3.3-7*.

- 1163 Tian, F., Li, B., Ji, B., Zhang, G., & Luo, Y. (2009). Identification and structure-activity
1164 relationship of gallotannins separated from *Galla chinensis*. *LWT - Food Science and*
1165 *Technology*, 42(7), 1289–1295. doi: 10.1016/j.lwt.2009.03.004
- 1166 Vavilov, N. I. (1926). Studies on the origin of cultivated plants. *Trudy Byuro. Prikl. Bot.*, 16,
1167 139–245.
- 1168 Vavilov, N. I. (1992). *Origin and geography of cultivated plants*. Cambridge: Cambridge
1169 University Press. CABDirect.
- 1170 Vekemans, X., & Hardy, O. J. (2004a). New insights from fine-scale spatial genetic structure
1171 analyses in plant populations. *Molecular Ecology*, 13(4), 921–935.
- 1172 Vekemans, X., & Hardy, O. J. (2004b). New insights from fine-scale spatial genetic structure
1173 analyses in plant populations. *Molecular Ecology*, 13(4), 921–935. doi: 10.1046/j.1365-
1174 294X.2004.02076.x
- 1175 Vercken, E., Fontaine, M. C., Gladieux, P., Hood, M. E., Jonot, O., & Giraud, T. (2010). Glacial
1176 refugia in pathogens: European genetic structure of anther smut pathogens on *Silene*
1177 *latifolia* and *Silene dioica*. *PLoS Pathogens*, 6(12), e1001229.
- 1178 Volk, G. M., & Cornille, A. (2019). Genetic Diversity and Domestication History in *Pyrus*. In
1179 *The Pear Genome* (pp. 51–62). Springer.
- 1180 Volk, G. M., Richards, C. M., Reilley, A. A., Henk, A. D., Reeves, P. A., Forsline, P. L., &
1181 Aldwinckle, H. S. (2008). Genetic diversity and disease resistance of wild *Malus*
1182 *orientalis* from Turkey and Southern Russia. *Journal of the American Society for*
1183 *Horticultural Science*, 133(3), 383–389.
- 1184 Vouillamoz, J. F., McGovern, P. E., Ergul, A., Söylemezoğlu, G., Tevzadze, G., Meredith, C. P.,
1185 & Grando, M. S. (2006). Genetic characterization and relationships of traditional grape

1186 cultivars from Transcaucasia and Anatolia. *Plant Genetic Resources*, 4(2), 144–158. doi:
1187 10.1079/pgr2006114

1188 Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A
1189 versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, 11(1), 116.
1190 doi: 10.1186/1471-2105-11-116

1191 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population
1192 structure. *Evolution*, 38(6), 1358–1370.

1193 Wilson, G. A., & Rannala, B. (2003). Bayesian Inference of Recent Migration Rates Using
1194 Multilocus Genotypes. *Genetics*, 163(3), 1177–1191. doi: 10.1093/genetics/163.3.1177

1195 Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., ... Zhang, S. (2018). Diversification and
1196 independent domestication of Asian and European pears. *Genome Biology*, 19(1), 77. doi:
1197 10.1186/s13059-018-1452-y

1198 Yousefzadeh, H., Hosseinzadeh Colagar, A., Tabari, M., Sattarian, A., & Assadi, M. (2012).
1199 Utility of ITS region sequence and structure for molecular identification of *Tilia* species
1200 from Hyrcanian forests, Iran. *Plant Systematics and Evolution*, 298(5), 947–961. doi:
1201 10.1007/s00606-012-0604-x

1202 Zazanashvili, N., Sanadiradze, G., Garforth, M., Bitsadze, M., Manvelyan, K., Askerov, E., ...
1203 Devranoğlu Tavşel, S. (2020). *Ecoregional Conservation Plan (ECP) For The Caucasus*
1204 *2020 Edition*. WWF, KfW, Tbilisi.

1205 Zeng, G., Zhang, J., Chen, Y., Yu, Z., Yu, M., Li, H., ... Hu, C. (2011). Relative contributions of
1206 archaea and bacteria to microbial ammonia oxidation differ under different conditions
1207 during agricultural waste composting. *Bioresour Technol*, 102(19), 9026–9032. doi:
1208 10.1016/j.biortech.2011.07.076

1209 Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B. H. (2017). Back into the wild—
1210 Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary*
1211 *Applications*, 10(1), 5–24. doi: 10.1111/eva.12434

1212 Zhang, H.-X., Li, H.-Y., & Li, Y.-X. (2018). Identifying evolutionarily significant units for
1213 conservation of the endangered *Malus sieversii* using genome-wide RADseq data. *Nordic*
1214 *Journal of Botany*, 36(7), e01733. doi: 10.1111/njb.01733

1215

1216

1217 Figures and Tables

1218

1219 **Figure 1. Population genetic structure in cultivated and wild apples from the Caucasus and**
1220 **Iran, *Malus domestica*, *Malus sieversii* and *Malus baccata* based on 26 microsatellite markers.**

1221 a. Spatial population genetic structure inferred with STRUCTURE at $K = 12$ ($N = 550$); the map
1222 represents membership proportions averaged over each geographic site for the Caucasian crabapple
1223 *M. orientalis* ($N = 374$, 43 sites across Turkey, Russia, Armenia and Iran). In the bottom right
1224 corner, the mean membership proportions for the apple cultivars from Armenia ($N = 4$), Iran ($N =$
1225 48) and Europe (*M. domestica*, $N = 40$). The size of the pie charts is proportional to the number of
1226 samples per site. b. STRUCTURE bar plot ($N = 550$) at $K = 12$ showing 12 distinct genetic clusters.
1227 Each vertical line represents an individual. Colors represent the inferred ancestry from K ancestral
1228 genetic clusters. Sites are grouped by country for the wild apple samples (*i.e.*, Turkey, Russia,
1229 Armenia, Iran) and *M. sieversii* (*i.e.*, Kazakhstan and Kyrgyzstan), apple cultivars are grouped
1230 according to their origin: Armenia ($N = 4$), Iran ($N = 48$) and *M. domestica* ($N = 40$). Countries
1231 (Kazakhstan, Kyrgyzstan, Armenia) and/or main regions in the Caucasus (the Western Caucasus,
1232 *i.e.*, Turkey and Russia, Zagros and Hyrcanian Forests, Central and Southern Armenia) are shown
1233 on the map. Reference samples from previously published studies of each species are: *M. orientalis*
1234 from the Western Caucasus and Central and Southern Armenia, *M. domestica* (European cultivars),
1235 *M. sieversii* from Kazakhstan (Cornile et al., 2013) and *M. baccata* (Cornille et al., 2012).

1236

1237

1238 **Figure 2. Genetic variation and differentiation among cultivated and wild apples from the Caucasus, *Malus domestica*, *Malus***
1239 ***sieversii* and *Malus baccata* based on 26 microsatellite markers.** a. Principal component analysis after removing the outgroup *M.*
1240 *baccata* ($N = 530$), with the respective total variation explained by each of the two first components. b. Neighbor-net representing the
1241 genetic relationships among wild and cultivated individuals inferred with STRUCTURE at $K = 12$. Colors correspond to the genetic
1242 groups inferred with STRUCTURE at $K=12$, and admixed samples are in grey.

1243

1244 **Figure 3. Most likely scenario of domestication of cultivated apples in Iran using random-forest approximate Bayesian**
1245 **computation (ABC-RF) combined with coalescent-based simulations (SC45, Figure S12).** Population names correspond to the ones
1246 detected with STRUCTURE for $K = 12$, excluding admixed individuals (*i.e.*, individuals with a membership coefficient < 0.85 to any
1247 given cluster), *M. baccata* was excluded from the inferences. Bidirectional gene flow between populations was assumed. For clarity,
1248 only estimates of divergence time of the cultivated populations are provided, other parameters estimates are provided in Table S13.

1249

1250

1

2 **Figure 4. Spatial diversity and past contraction and expansion of *Malus orientalis* across the Caucasus. a.** Spatial genetic
3 diversity (allelic richness) at 36 sites ($N = 339$). **b. and c.** Ensemble forecasting of the three different algorithms (ANN, GLM and
4 GAM) predicting the current and last glacial maximum (LGM) distribution range of suitable areas for *M. orientalis*, respectively. The
5 probabilities of being a suitable habitat are given in the legend. The Colchis and Hyrcanian regions are shown on the maps.

1 **Table 1. Genetic diversity estimates for wild and cultivated apple populations detected with STRUCTURE at $K = 12$ ($N = 424$,**
2 *i.e.*, individuals with a membership coefficient < 0.85 to any given cluster were excluded from the analysis). Note that the purple cluster
3 was split between cultivated and wild samples. Thus, samples were partitioned into 13 populations, including 10 wild and three cultivated
4 apple populations.
5

Wild or cultivated	Species	Country of origin	Population	N	H_O	H_E	F_{IS}	$A_R (G=7)$	$A_P (G=7)$
wild	<i>Malus orientalis</i>	Armenia	Western (orange)	20	0.82	0.86	0.05**	5.04±0.13	0.97±0.10
			Central (blue)	109	0.78	0.79	0.01 ^{NS}	4.03±0.14	0.60±0.08
			Southern (brown)	29	0.76	0.79	0.03 ^{NS}	4.38±0.17	0.63±0.12
		Iran	Lorestan (light green)	10	0.84	0.46	-0.82***	2.03±0.12	0.23±0.06
			Kurdestan (red)	12	0.75	0.71	-0.05*	3.71±0.15	0.50±0.09
			Hyrcanian (pink)	88	0.71	0.76	0.07***	4.12±0.18	0.65±0.10
			Hyrcanian (purple)	19	0.76	0.63	-0.21***	3.24±0.13	0.21±0.03
	<i>Malus sieversii</i>	Kazakhstan	(light blue)	40	0.74	0.76	0.03*	4.15±0.16	0.99±0.13
		Kyrkyzstan	(cyan)	17	0.70	0.73	0.04*	3.47±0.14	0.30±0.06
	<i>Malus baccata</i>	Russia	(light red)	17	0.52	0.55	0.06 ^{NS}	3.23±0.19	1.08±0.14
cultivated	<i>Malus domestica</i>	Mostly Europe	European cultivars	38	0.78	0.78	-0.00 ^{NS}	4.27±0.09	0.85±0.10

	?	Iran	Iranian cultivars (purple)	18	0.77	0.61	-0.26 ***	3.08±0.13	0.10±0.02
			Iranian cultivars (dark green)	7	0.71	0.65	-0.08 *	3.39±0.13	0.52±0.09
TOTAL				424					

1

2 *N*: number of individuals assigned to a focal cluster with a membership coefficient > 0.85; *H_O* and *H_E*: observed and expected
3 heterozygosity; *F_{IS}*: inbreeding coefficient; *A_R*: mean allelic richness across loci, corrected by the rarefaction method, estimated for a
4 sample size of 7; *A_P*: number of private alleles, corrected by the rarefaction method, estimated for a sample size of 7; *: 0.05 < *P* < 0.01;
5 ***: *P* < 0.001; NS: non-significant.

6

1 **Table 2. Distribution of hybrids (i.e., individuals with a membership coefficient < 0.90 to any given genetic cluster, as inferred**
 2 **with STRUCTURE for $K = 9$) in cultivated and wild apple in the Caucasus ($N = 466$, 26 microsatellite markers).**
 3

Gene pool		Crop				Wild			
		Armenia $N_{tot}=3$	Iran $N_{tot}=48$	<i>M. domestica</i> $N_{tot}=40$		Armenia $N_{tot}=196$	Iran $N_{tot}=167$	Russia $N_{tot}=5$	Turkey $N_{tot}=6$
<i>Malus domestica</i>	mean introgression rate	0	0.13	0.66		0.04	0.02	0	0
	N (% over the total)	0	11 (23%)	2 (5%)		4 (2%)	5 (2.9%)	0	0
Wild and cultivated Hyrcanian (purple)	mean introgression rate	0	0.36	0.33		0	0.16	0	0
	N (%)	0	18 (37.5%)	2 (5%)		0	13 (7.8%)	0	0
Cult. Iran (dark_green)	mean introgression rate	0	0.17	0		0.05	0.06	0	0
	N (%)	0	14 (30%)	0		11 (5.6%)	8 (4.7%)	0	0
	Ncrop-crop hybrids (and %)	0	43 (89.5%)	4 (1%)	Ncrop-to-wild hybrids (and %)	15 (7.6%)	26 (18.1%)	0	0
Hyrcanian (pink)	mean introgression rate	0	0.05	0		0	0.41	0	0
	N (%)	0	6 (12.5%)	0		0	34 (20.3%)	0	0
Lorestan (light green)	mean introgression rate	0	0.01	0		0	0.04	0	0
	N (%)	0	2 (4%)	0		0	6 (3.6%)	0	0
Wild Kurdistan (red)	mean introgression rate	0	0.17	0		0	0.27	0	0
	N (%)	0	12 (35%)	0		0	30 (17.9%)	0	0
Western (orange)	mean introgression rate	0.65	0.08	0		0.28	0.02	0.49	0.84
	N (%)	1 (33%)	3 (6.2%)	0		33 (16.8%)	4 (2.3%)	1 (20%)	2 (33%)
Southern (brown)	mean introgression rate	0	0.02	0		0.17	0.02	0.35	0
	N (%)	0	2 (4.1%)	0		13 (6.6%)	2 (1.1%)	1 (20%)	0

Central (blue)	mean introgression rate	0.32	0.02	0		0.46	0.01	0	0.11	
	N (%)	1 (33%)	1 (2%)	0		46 (23.4%)	1 (0.5%)	0	1 (16.6%)	
	Nwild-to-crop hybrids (and %)	2 (66%)	26 (54.1%)	0	Nwild-to-wild hybrids (and %)	92 (46.9%)	77 (46.1%)	2 (40%)	3 (50%)	
	Total number of hybrids (and %)	1	32 (66%)	2 (5%)		58 (28%)	54 (32%)	1 (20%)	2 (33%)	150 (32%)

1 *N_{tot}*: total number of samples in each group (*i.e.*, cultivated or wild, from different regions); *N* (and %): number of hybrids assigned to
2 each gene pool, and the respective percentage over the total number of samples from each group; mean introgression rate: mean
3 membership coefficient to this gene pool. Note that some admixed trees were assigned to several gene pools with a membership
4 coefficient < 0.90, therefore, the total number of hybrids associated with each cluster (TOTAL) is given on the last line of the table. We
5 also showed the distribution of crop-crop, crop-to-wild, wild-to-crop and wild-to-wild hybrids.

