bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Title: Evidence of an additional center of apple domestication in Iran, with contributions from the Caucasian crab apple *Malus orientalis*

Running head: Apple domestication in the Caucasus

Bina Hamid[1], Yousefzadeh Hamed[2], Venon Anthony[3], Remoué Carine[3], Rousselet Agnès[3], Falque Matthieu[3], Shadab Faramarzi[4], Xilong Chen[3], Jarkyn Samanchina[5], David Gill[5], Akylai Kabaeva[5], Giraud Tatiana[6], Hossainpour Batool[7], Abdollahi Hamid[8], Gabrielyan Ivan[9], Nersesyan Anush[10], Cornille Amandine[3]

[1] Department of Forestry, Tarbiat Modares University, Noor, Iran

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

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

[4] Department of Plant Production and Genetics**,** Faculty of Agriculture, Razi University, Kermanshah, Iran

[5] Fauna & Flora International, Cambridge, UK

[6] Ecologie Systematique Evolution, Universite Paris-Saclay, CNRS, AgroParisTech, Orsay, France

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

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

[9] Department of Conservation of Genetic Resources of Armenian Flora, A. Takhtajyan Institute of Botany, Armenian National Academy of Sciences, Acharyan Str.1, 0040 Yerevan, Armenia

[10] Department of Palaeobotany, A. Takhtajyan Institute of Botany, Armenian National Academy of Sciences, Acharyan Str.1, 0040 Yerevan, Armenia

Corresponding authors: amandine.cornille@universite-paris-saclay.fr, hamed.yousefzadeh@modares.ir

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Abstract

Anthropogenic and natural divergence processes in crop-wild fruit tree complexes are less studied than in annual crops, especially in the Caucasus, a pivotal region for plant domestication. We investigated anthropogenic and natural divergence processes in apples in the Caucasus using 26 microsatellite markers amplified in 550 wild and cultivated samples. We found two distinct cultivated populations in Iran, that were both genetically differentiated from *Malus domestica*, the standard cultivated apple worldwide. Coalescent-based inferences using approximate Bayesian computation showed that these two cultivated populations originated from specific domestication events in Iran. One of the Iranian clusters encompassed both cultivated and forest trees, suggesting that either farmers use of local wild apple for cultivation or that some forest trees represent feral cultivars. We found evidence of substantial wild-to-crop, crop-crop and crop-to-wild gene flow in the Caucasus, as has been previously described in apple in Europe and in fruit trees in general. In the Caucasus, we identified seven genetically differentiated populations of wild apple (*Malus orientalis*). Niche modeling combined with genetic diversity estimates indicated that these populations likely resulted from range changes during the last glaciation. This study identifies Iran as a key region in the evolution and domestication of apple and further demonstrates the role of wild-to-crop gene flow during fruit tree domestication. Our results support the view that domestication of fruit trees was likely a geographically diffuse and protracted process, involving multiple, geographically disparate, origins of domestication. We also highlight the impact of climate change on the natural divergence of a wild fruit tree and provides a base for apple conservation and breeding programs in the Caucasus.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

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

Fruit trees present several historical and biological features that make them fascinating models for investigating anthropogenic and natural divergence with gene flow. Several fruit tree crops are spread across the world and are sometimes found in sympatry with their wild relatives (Besnard et al., 2018; Cornille et al., 2019, 2014; Liu et al., 2019). Fruit trees are also characterized by high levels of gene flow during divergence, which is to be expected considering the typical life history traits of trees (Cornille et al., 2013; Cornille, Gladieux, & Giraud, 2013a; Oddou-Muratorio & Klein, 2008; Petit & Hampe, 2006). Population genetics studies of natural divergence processes associated with the last glacial maximum in Europe, North America and Asia in wind-dispersed trees (*e.g., Abies*, *Pinus*, *Fraxinus*, *Quercus*, *Betula* (Lascoux, Palmé, Cheddadi, & Latta, 2004; Petit et al., 2004)) and animal-dispersed trees (Cornille et al., 2013a) showed that there were high levels of gene flow between populations and that trees had high

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

The Caucasus ecoregion harbors a remarkable concentration of both economically important plants and their wild relatives, in particular wheat, rye, barley and also fruit trees including walnut, apricot and apple (Asanidze, Akhalkatsi, Henk, Richards, & Volk, 2014a; Gabrielian & Zohary, 2004; Vavilov, 1926, 1992; Yousefzadeh, Hosseinzadeh Colagar, Tabari, Sattarian, & Assadi, 2012). This region covers Georgia, Armenia, Azerbaijan, the North Caucasian part of the Russian Federation, the northeastern part of Turkey and the Hyrcanian Mixed Forests region in northwestern Iran (Nakhutsrishvili, Zazanashvili, Batsatsashvili, & Montalvo, 2015; Zazanashvili et al., 2020). Two Pleistocene refugia for temperate plants are recognized in this region (Bina, Yousefzadeh, Ali, & Esmailpour, 2016; Yousefzadeh et al., 2012): the Colchis refugium in the catchment basin of the Black Sea (in the Western Caucasus), and the Hyrcanian refugium at the southern edge of the Caspian Sea. Glacial refugia are known to harbor higher levels of species and genetic diversity (Hewitt, 2004), and this is the case for the Colchis and Hyrcanian refugia. The geography of the Caucasus, with two parallel mountain chains separated by contrasted climatic zones, makes this region a good model for investigating the natural divergence processes associated with the last glacial maximum. Furthermore, it has

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

been suggested that Iran, with its close proximity to Central Asia - the center of origin of emblematic fruit trees - and its historic position on the Silk Trade Routes, is a possible secondary center of domestication of apple, grape and apricot (Decroocq et al., 2016; Liang et al., 2019; Liu et al., 2019). However, inferences of the natural and anthropogenic divergence history of wild-crop fruit tree complexes in the Caucasus have been limited by the small number of samples (Decroocq et al., 2016; Liu et al., 2019) and/or genetic markers investigated so far (Amirchakhmaghi et al., 2018; Asanidze, Akhalkatsi, Henk, Richards, & Volk, 2014b; Cornille et al., 2013a; Gharghani et al., 2010; Myles et al., 2011; Volk & Cornille, 2019; Vouillamoz et al., 2006).

The Caucasian crab apple, *Malus orientalis* Uglitzk., is an endemic wild apple species occurring in the Caucasus. More specifically, it is found in the Western Caucasus (*i.e.*, the southern part of Russia, northern Anatolia and northwestern Turkey), Georgia, Armenia, the mountainous belt in northern Iran, the Hyrcanian Forests (Büttner, 2001; Rechinger, 1964) and the Zagros forests in eastern and central Iran (Browicz, 1969; Rechinger, 1964) (Figure 1). This species displays high phenotypic diversity across its distribution range where it occurs as scattered individuals in natural forests or at high altitude in rocky mountains (Fischer & Schmidt, 1938; Rechinger, 1964). It has a high resistance to pests and diseases (Büttner, 2001) and its fruit, of high quality, are variable in size (2-4 cm) and color (green to yellowish) (Cornille et al., 2014). Fruit of *M. orientalis* are harvested across the Caucasus for stewing and processed as juice and other beverages (cider, wine), jelly, syrup, jam, and vinegar (Amirchakhmaghi et al., 2018; Büttner, 2001). This has led some authors to suggest that some local apple cultivars from several regions of the Caucasus originated from *M. orientalis* (Forsline, Aldwinckle, Dickson, Luby, & Hokanson, 2003; Langenfeld, 1991; Schmitt, 2007). Other studies also suggested that *M. orientalis* has been present and cultivated in Iran long before the *M. domestica* trees were introduced (Gharghani et al., 2009; Gharghani et al., 2010; Spengler, 2019). *Malus domestica*, the standard cultivated apple, is also currently grown in various regions of the Caucasus (Forsline et al., 2003; Gharghani et al., 2010; Gharghani et al., 2009; Langenfeld, 1991; Schmitt, 2007). So far, the relationships between *M. orientalis*, the local Caucasian cultivars, *M. domestica* and its Central Asian progenitor, *Malus sieversii*, are still unknown. This raises the following questions: i) Are cultivated apples in the Caucasus derived from the same domestication event as *M. domestica*? ii) If not, has *M. orientalis* contributed to the local cultivated Caucasian apple

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

germplasm through wild-to-crop introgression, in the same way that *M. sylvestris*, the European crab apple, contributed to the *M. domestica* gene pool (Cornille et al., 2012; Ali Gharghani et al., 2009)?; iii) Are wild *M. orientalis* and/or crop-wild hybrid trees cultivated in gardens and orchards in Iran? Conversely, are there feral and/or hybrid trees in the wild? The extent of crop-wild gene flow in apples in the Caucasus has just begun to be investigated. One study suggested that *M. orientalis* only made a minor contribution to Mediterranean *M. domestica* cultivars (Cornille et al., 2012), but lacked in-depth investigation. A population genetics study revealed low levels of crop-to-wild gene flow from *M. domestica* to *M. orientalis* in natural forests of Armenia, Turkey and Russia (Cornille et al., 2013a). Population genetic diversity and structure analyses of *M. orientalis* populations from the Western and South Caucasus identified three differentiated populations: one in Turkey, one in Armenia and one in Russia (Cornille et al., 2013a). On a smaller geographic scale, an east-west genetic subdivision was found across the Hyrcanian Forests in Iran, with five main populations showing admixture (Amirchakhmaghi et al., 2018). However, we still lack a comprehensive view (beyond Armenia) of the genetic diversity and structure of *M. orientalis* to understand its natural divergence history. In addition, studying local cultivars from the Caucasus will shed light on the relationships between local wild apple populations, the local cultivated apple and the standard cultivated apple *M. domestica,* as well as the extent of crop-wild gene flow in apple in this region.

Here, we investigate anthropogenic and natural divergence processes in apples from the Caucasus and the extent of gene flow during divergence. A total of 550 apple trees, comprising local cultivated and wild apples from the Caucasus, *M. domestica* apple cultivars and *M. sieversii*, were sampled and genotyped using 26 microsatellite markers. In addition, the Siberian wild apple *Malus baccata* was used as an outgroup in certain analyses. From the analysis of this comprehensive genetic dataset, combined with ecological niche modeling approaches, we addressed the following questions: 1) What is the population genetic structure among Caucasian wild and cultivated apples, *M. domestica* and *M. sieversii*, and what are their genetic relationships, levels of genetic diversity and demographic history? 2) Is there any crop-crop or crop-wild gene flow in the Caucasus? Are there feral or crop-wild hybrid trees in the wild or, conversely, do humans cultivate wild or hybrid trees? 3) Did *M. orientalis* experience past range contraction and expansion associated with the last glacial maximum?

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and methods

Sampling, DNA extraction and microsatellite genotyping

Microsatellite genotyping data for *M. orientalis* (Cornille et al., 2013a, 2012), *M. sieversii* from Kazakhstan (Cornille et al., 2013a), *M. domestica* (Cornille et al., 2013a) and *M. baccata* (Cornille et al., 2012) were from previously published studies; we had previously available data for 207 *M. orientalis* individuals from Turkey, Armenia and Russia (23 sites, Tables S1 and S2), four apple cultivars from Armenia, 40 “pure” European cultivated *M. domestica* individuals (*i.e.*, not -introgressed by *M. sylvestris*) (Tables S1 and S2), 20 *M. sieversii* individuals from Kazakhstan (Cornille et al., 2012) and 22 *M. baccata* individuals from Russia (Cornille et al., 2012). We collected new samples in 2017 in Iran and in 2018 in Kyrgyzstan (*M. sieversii*). A total of 257 samples were collected for this study: 167 *M. orientalis* individuals from the Hyrcanian Forests and the Zagros region in Iran (Table S1), 48 local Iranian apple cultivars from the *Seed and Plant Improvement Institute* (Karaj, Iran) (Table S1) and 42 *M. sieversii* individuals from Kyrgyzstan. Note that for 18 of the 48 local Iranian samples, we measured fruit size (Table S1). Collections meet the requirements of the recently enacted Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits. Thus, a total of 550 individuals were analyzed, comprising 374 wild *M. orientalis*, 48 Iranian and four Armenian apple cultivars, 40 European apple cultivars belonging to *M. domestica*, 62 *M. sieversii* (from Kyrgyzstan and Kazakhstan) and 22 *M. baccata* individuals (details are provided in Table S1).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Bayesian inferences of population structure and genetic differentiation among wild and cultivated apples

We investigated the population structure of wild and cultivated apples with the individual-based Bayesian clustering methods implemented in STRUCTURE 2.3.3 (Pritchard, Stephens, & Donnelly, 2000). STRUCTURE uses Markov chain Monte Carlo (MCMC) simulations to infer the proportion of ancestry of genotypes from K distinct clusters. The underlying algorithms attempt to minimize deviations from Hardy–Weinberg within clusters and linkage disequilibrium among loci. We ran STRUCTURE from $K = 1$ to $K = 15$. Based on 10 repeated runs of MCMC sampling from 500,000 iterations after a burn-in of 50,000 steps, we determined the amount of additional information explained by increasing K using the ΔK statistic (Evanno, Regnaut, & Goudet, 2005) as implemented in the online post-processing software Structure Harvester (Earl, 2012). However, the K value identified with the ΔK statistic often does not correspond to the most biologically relevant number of clusters, as it only corresponds to the strongest population subdivision level and not the finest subdivision level (Puechmaille, 2016); we therefore visually checked the bar plots and chose the K value for which all clusters had well-assigned individuals (indicating the existence of a genuine population subdivision), while no further well-delimited and biogeographically relevant clusters could be identified for higher K values (indicating we reached the highest K value for which new genuine clusters could be delimited).

We ran STRUCTURE for the whole dataset ($N = 550$) to investigate the population genetic structure among the Caucasian wild apple *M. orientalis*, the Caucasian cultivated apples, *M. sieversii*, *M. domestica* and *M. baccata*. We further explored the genetic variation and differentiation among the genetic groups detected with STRUCTURE using three different methods. First, we ran a principal component analysis (PCA) for all individuals with the *dudi.pca* function from the *adegenet* R package (Jombart & Ahmed, 2011). For the PCA, individuals that were assigned to a given cluster with a membership coefficient ≥ 0.85 were colored according to the respective color of each cluster, and admixed individuals (*i.e.*, individuals with a membership coefficient to any given cluster < 0.85) were colored in gray. We chose this threshold based on the distribution of the maximum membership coefficients inferred with STRUCTURE (see results). The angle used to draw the PCA was 380 degrees for a better visualization of the data. Second, we generated a neighbor-net tree with Splitstree v4 (Huson, 1998; Huson & Scornavacca, 2012), using the PCA color code. Third, we explored the relationships among

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

populations identified with STRUCTURE (*i.e.*, clusters of individuals with a membership coefficient ≥ 0.85 to a given cluster) with a neighbor joining (NJ) tree (Huson, 1998; Huson & Scornavacca, 2012). The NJ tree and the neighbor-net tree were built using Nei's standard genetic distance (Nei, 1987) computed among individuals or populations with the Populations software v1.2.31 (https://bioinformatics.org/populations/).

Genetic diversity estimates and test for the occurrence of a bottleneck in wild and cultivated apples

We computed descriptive population genetic estimates for each population (*i.e.*, each cluster inferred with STRUCTURE, excluding admixed individuals with a membership coefficient < 0.85). We calculated allelic richness (A_R) and private allelic richness (A_P) with ADZE (Szpiech, Jakobsson, & Rosenberg, 2008) using standardized sample sizes of N_{ADZE} = 7 (one individual x two chromosomes), corresponding to the minimal number of observations across populations. Heterozygosity (expected and observed), Weir and Cockerham *F*-statistics and deviations from Hardy–Weinberg equilibrium were calculated with Genepop v4.2 (Raymond & Rousset, 1995; Rousset, 2008).

Approximate Bayesian computation to reconstruct the domestication history of the Iranian cultivated apple

We used approximate Bayesian computation to test whether cultivated apples in the Caucasus were derived from the same domestication event as *M. domestica*. We used the newly developed ABC method based on a machine learning tool named "random forest" (ABC-RF) to perform model selection and parameter estimates (Estoup et al., 2018; Pudlo et al., 2016; Raynal et al., 2019). In brief, this method creates a "forest" of bootstrapped decision trees that ranks scenarios based on the summary statistics of the datasets. Some simulations are not used to build the trees and can thus be used to cross-validate the analysis by computing a "prior error rate". This approach allows the comparison of complex demographic models (Pudlo et al., 2016) by comparing groups of scenarios with a specific type of evolutionary event with other groups with different types of evolutionary events instead of considering all scenarios separately (Estoup et al., 2018).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Using the ABC-RF framework, we compared different scenarios of domestication of Iranian cultivars, *i.e.*, with an origin from i) *M. domestica*, ii) *M. sieversii* from Kazakhstan, iii) another, unknown ancestor. Populations were defined as the clusters detected with STRUCTURE, removing putative hybrid individuals (*i.e.*, individuals with a membership coefficient < 0.85 to any given cluster). We removed putative hybrids in order to retrace the divergence history and historical gene flow among populations; more recent admixture events being detectable directly from the STRUCTURE bar plots. We assumed bidirectional gene flow among wild and cultivated apple populations. The model parameters used were: the divergence time between X and Y populations ($T_{X\text{-}Y}$), the effective population size of population X ($N_{E\text{-}X}$) and the migration rate per generation between X and Y populations ($m_{X\text{-}Y}$). Prior values for divergence time were drawn from the log-uniform distribution bounded between the distributions used in the approximate Bayesian computations and are given in Table S3.

For all models, microsatellite datasets were simulated for 14 out of the 26 markers that had perfect repeats (Ch01h01, Ch01h10, Ch02c06, Ch02d08, Ch05f06, Ch01f02, Hi02c07, Ch02c09, Ch03d07, Ch04c07, Ch02b03b, MS06g03, Ch04e03, Ch02g01, (Cornille et al., 2013; Cornille et al., 2012; Patocchi, Fernàndez-Fernàndez, et al., 2009). We checked that the population structure inferred with 14 microsatellite markers did not differ significantly from the inferences obtained with 26 SSR markers (data not shown). We assumed a generalized stepwise model of microsatellite evolution. Mutation rates were allowed to vary across loci, with locus-specific mutation rates drawn from a gamma distribution (α, α/μ) where μ is the mutation rate per generation and α is a shape parameter. We assumed a log-uniform prior distribution for μ (1e-5, 1e-3) and a uniform distribution for α (1.30) (Table S3).

We used ABCtoolbox (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010) with fastsimcoal 2.5 (Excoffier & Foll, 2011) to simulate datasets, using model parameters drawn from prior distributions (Table S3). We performed 7,500 simulations per scenario. For each simulation, we calculated three summary statistics per population with arlsumstats v 3.5 (Excoffier and Lischer 2010): H, the mean heterozygosity across loci, *sd(GW)*, the standard deviation of the mean Garza-Williamson statistic over populations (Garza and Williamson, 2001) and the pairwise F_{ST} between populations (Weir and Cockerham, 1984).

We used the abcrf v.1.7.0 R statistical package (Pudlo et al., 2016) to carry out the ABC-RF analysis. This analysis provides a classification vote that represents the number of times

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

a scenario is selected as the best one among *n* trees in the constructed random forest. For each ABC step, we selected the scenario, or the group of scenarios, with the highest number of classification votes as the best scenario, or best group of scenarios, among a total of 500 classification trees (Breiman, 2001). We computed the posterior probabilities and prior error rates (*i.e.*, the probability of choosing a wrong group of scenarios when drawing model index and parameter values from the priors of the best scenario) over 10 replicate analyses (Estoup et al., 2018) for each ABC step. We also checked visually that the simulated models were compatible with the observed dataset by projecting the simulated and the observed datasets onto the two first linear discriminant analysis (LDA) axes (Pudlo et al., 2016), and by checking that the observed dataset fell within the clouds of simulated datasets. We then calculated parameter inferences using the final selected model. Note that the ABC-RF approach includes the model checking step that was performed *a posteriori* in previous ABC methods.

Identification of crop-wild hybrids and historical gene flow in the wild apple *M. orientalis* in the Caucasus

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

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

Spatial pattern of genetic diversity in the Caucasian crab apple

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

Species distribution modeling

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

The location of 339 "pure" *M. orientalis* trees (*i.e.*, individuals assigned to a wild apple gene pool with a membership coefficient > 0.9, see results from the second STRUCTURE analysis) provided the longitude and latitude coordinates. Duplicate data points were removed, resulting in 57 presence points for *M. orientalis* (Table S4). We did not have absence data so we randomly selected pseudo-absences to serve as "absence" points for the model, and weighted presence and absence points equally as per Barbet-Massin *et al.* (2012). Models were calibrated using the set of bioclimatic variables and model evaluation was calculated with Jaccard's indices. Ensemble model forecasting was completed by pulling the average trend of the three modeling algorithms and retaining only the uncorrelated bioclimatic variables with a Pearson correlation threshold greater than 0.75 (Table S5). The model was run again using only variables with high predictive power.

Results

Clear genetic structure and variation in cultivated and wild apples

The ΔK statistic indicated that the strongest level of population subdivision was at $K = 3$ (Figure S1 a, b). However, further genetic subdivisions were observed for $K > 3$, with well delimited and biologically meaningful clusters. We therefore visually examined the bar plots and chose the K value at which all clusters had fully assigned individuals, indicating the finest level of genetic subdivision. At $K = 12$, STRUCTURE identified twelve well-delimited clusters (Figures 1 a, b and S2) corresponding to species and/or geographic regions (Figure 1). We therefore considered these twelve clusters as the most relevant genetic structure.

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

The *M. domestica* apple cultivars formed a specific genetic group (yellow) that was well separated from the wild *M. orientalis* and the Iranian and Armenian cultivars. The Iranian apple cultivars formed two gene pools: one that included only cultivars (dark green), and another

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(purple) that included cultivated trees and wild *M. orientalis* individuals from the Hyrcanian Forests. We also detected Iranian cultivated trees that were highly admixed between the Iranian cultivated dark green cluster, the Iranian purple cluster, the *M. domestica* yellow cluster and with two other clusters (red and orange), the latter two included several wild *M. orientalis* individuals from the Zagros Forests in the Kurdestan province in Iran and the Western Caucasus, respectively (Figures 1 and S2). The four Armenian cultivated apple fell within the blue and orange clusters, which also included wild *M. orientalis* trees.

We assigned individuals with a membership coefficient > 0.85 to a given cluster to the corresponding population (Figure S3) to assess the genetic variation among wild and cultivated apples. The PCA (Figure 1c) showed that *M. baccata* was highly differentiated from the other genetic groups (Figure 1c, upper left). The European (yellow) and the two Iranian cultivated genetic clusters (green and purple) formed well differentiated gene pools (Figure 1c, lower right). *Malus orientalis* from the Western Caucasus (orange) and Central and Southern Armenia (brown and blue) were closer to each other than to the Iranian wild apples (light green, red and pink), which clustered together. *Malus sieversii* from Kazakhstan was closer to the Armenian and Iranian wild apple than *M. sieversii* from Kyrgyzstan.

STRUCTURE analysis and PCA thus revealed three main aspects in the history of apples in the Caucasus. First, cultivated apples in Iran (dark green) may have resulted from a domestication event that was independent from the domestication of *M. domestica*, as their respective clusters are not closely related in the PCA. The close genetic relationships between the wild and cultivated Iranian clusters indicates that the cultivated apples in Iran could have been domesticated from the local wild apple, *M. orientalis*, although they could also be derived from *M. sieversii* or *M. domestica* with subsequent local gene flow in Iran. Second, the full membership of a substantial number of cultivated trees to genetic clusters of wild *M. orientalis* suggests that wild trees are grown in orchards for consumption without any strong domestication process and/or feral individuals occur (*e.g.*, the purple genetic cluster may represent a cultivated group that is also found in the wild as feral). In addition, the high level of admixture in several Iranian cultivated apple trees with wild *M. orientalis* gene pools indicate substantial wild-crop gene flow. Third, the spatial population structure of *M. orientalis* in the Caucasus may result from past range contraction and expansion associated with the last glacial maximum. We tested these hypotheses as described in the methods. First, we investigated the genetic diversity, genetic differentiation and evolutionary history of cultivated and wild apple populations using a statistical framework for assessing the probability of different domestication scenarios. Second, we estimated the extent of crop-wild and wild-wild gene flow in apples in the Caucasus. Third, we investigated past contraction and expansion of wild apple populations in the Caucasus using both genetic diversity estimates and ecological niche modeling.

Genetic diversity and differentiation among wild and cultivated apples

The neighbor-net tree (Figure 2a) and the NJ tree (Figure 2b) confirmed that the apple cultivars (*M. domestica* and the purple and dark green Iranian cultivar clusters) were distinct from the wild populations (with the exception of the wild Hyrcanian purple group, see below). Note that we excluded admixed individuals (*i.e.*, individuals assigned with a membership coefficient < 0.85 to a given cluster) from the NJ analysis to better assess the genetic relationships among pure cultivated and wild populations.

Individuals of *M. domestica* and *M. sieversii* from Kazakhstan were intermingled (Figure 2a) and the genetic clusters of these species were sister groups (Figure 2b). Neither of the two distinct Iranian cultivated gene pools (green and purple) was not sister groups with *M. domestica*, supporting the view that specific domestication events have occurred in Iran. The two cultivated apple populations from Iran were genetically highly differentiated (Table S6), were sister groups (Figure 2) and had lower levels of genetic diversity and fewer private alleles than *M. domestica* ($P < 0.01$, Tables 1 and S7), with the purple gene pool displaying the lowest level of genetic diversity and the least number of private alleles. The close relationships between trees sampled in the Hyrcanian Forests and cultivars from the purple gene pool (Figures 1 and 2), and the lower levels of genetic diversity in both of these populations suggest that the trees sampled in the Hyrcanian Forests assigned to the purple genetic group may be feral. Alternatively, it may be that the cultivated trees from the purple gene pool represent the first step of apple domestication in Iran, *i.e.*, wild trees cultivated by humans.

Malus orientalis and *M. sieversii* did not form a monophyletic group (Figure 2b). Wild *M. orientalis* populations from the Western Caucasus, (orange) and from central (blue) and southern (brown) Armenia grouped together, the latter two being sister groups. The *M. orientalis* population from the Lorestan province (Iran; light green) was intermingled with the *M. sieversii* population from Kyrgyzstan (cyan), (Figure 2a); when considered as separate populations *(i.e.*, excluding admixed individuals Figure 2b), the two populations formed sister groups (Figure 2b). Although the wild Hyrcanian (pink and purple) populations were intermingled (with the exception of a few wild purple individuals clustering with the cultivated purple population) in the neighbor-net tree, the NJ tree indicated that the wild purple population was closer to the cultivated purple population. Some *M. sieversii* trees from Kazakhstan were intermingled with *M. sieversii* from Kyrgyzstan and clustered with the wild Iranian populations. *Malus sieversii* from Kazakhstan formed a distinct group, placed as a sister group to *M. domestica* in both neighbor-net and NJ trees (Figure 2). The level of allelic richness was significantly lower in the wild apple populations from the Zagros Forest (*i.e.*, Lorestan (light green) and Kurdestan (red) than in the other wild populations (Tables 1 and S7).

Additional domestication events in Iran inferred with ABC-RF

We defined the populations used in the ABC framework from the clusters detected with STRUCTURE at *K*=12 for 550 wild and cultivated apple accessions, excluding admixed individuals (*i.e.*, with a membership coefficient < 0.85 to any given cluster, Figure S2, as recent gene flow can easily be seen from visual inspection of the barplots). We excluded the Armenian cultivars as they were represented by only four individuals; we also excluded *M. baccata* and all wild Iranian and Armenian populations to avoid intractably complex scenarios with too many populations. We therefore included only *M. sieversii* from Kazakhstan, the cultivated (green) and wild + cultivated (purple) populations from Iran.

We used a two-step nested ABC-RF approach. In the first step we inferred i) the sequence of divergence of the purple wild and cultivated populations (ABC step 1, Figure 3), including scenarios assuming a sister relationship between the cultivated and wild populations, which would support that either the wild trees from the purple group are feral or that the cultivated trees from the purple group represent an early domestication step of a wild genetic cluster. In the next step, we inferred ii) the progenitor of the Iranian cultivated apple, *i.e.*, either *M. sieversii*, *M. domestica* or an unknown ancestral population (ABC step 2, Figure 3). To build the scenarios, we assumed that the purple Iranian cultivated population was younger than the green Iranian cultivated population, as the former population had a lower number of private alleles and lower genetic diversity (Table 1). For all scenarios, we assumed bidirectional gene flow among

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

populations. This nested ABC approach avoids comparing overly complex models with too many populations and parameters, and is more powerful than testing all scenarios individually to disentangle the main evolutionary events characterizing demography and divergence (Estoup et al., 2018).

For each step of the ABC-RF approach, the projection of the reference table and the observed datasets onto the two LDA axes that explained most of the variance of the summary statistics showed that the observed data fell within the distribution of the simulated summary statistics (Figure S4), forming distinct clouds for each scenario or groups of scenarios. Visual inspection of the LDA plots indicated this method had a high power to discriminate and select scenarios; results were subsequently validated by the ABC-RF inferences presented below.

For all ten replicates, the step 1 of the ABC-RF method supported the sister group status of the cultivated and wild purple populations, indicating that either this wild population is feral or that this cultivated group represents an early domestication step without strong genetic differentiation (scenarios sc1-after, sc2-after, sc3-after, sc4-after, average of 298 votes out of the 500 RF-trees; posterior probabilities = 58%, prior error rate = 38.31%, Figure 3, Table S8). Using the group of scenarios selected in step 1 (Figure 3) as a backbone, all ten replicates of the second step of the ABC-RF analyses provided support for the domestication of the purple and the green Iranian cultivated apples from an unknown ancestral population (scenario sc1-feral, an average of 308 votes out of the 500 RF-trees; posterior probabilities = 64.8%, prior error rate = 8.5 %, Figure 3, Table S9). ABC-RF inferences provided support for the independent domestication of the two cultivated Iranian populations: *i.e.*, the dark green and purple cultivated apple populations diverged independently from an unknown ancestral population *c.* 2,190 years ago (ya) and 2,890 ya, respectively, while *M. domestica* diverged from *M. sieversii c*. 3,550 ya (Table S10).

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

The second STRUCTURE analysis, focusing on the Caucasus, revealed the same genetic clustering for wild apples and *M. domestica* at $K = 9$ (Figure S5) as in the previous analysis ($K = 12$) (Figures 1 and 2). At $K = 9$, 150 apple genotypes could be considered hybrids (*i.e.*, individuals assigned to a gene pool with a membership coefficient < 0.9, this cut-off being chosen on the basis of the distribution of the cumulated membership coefficients for each individual at $K = 9$, Figure S6); these 150 hybrids represented 32% of the total dataset (Table 2).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The Iranian cultivars had the highest proportion of hybrids (67%), mostly admixed with the wild and cultivated gene pools from Iran, but also with the *M. domestica* gene pool. Hybrids of the wild Armenian apple were mostly an admixture of the wild Armenian gene pools (*i.e.*, Western, Central and Southern), suggesting local gene flow between crop and wild populations.

We removed the 150 hybrids and all apple cultivars (Tables 2 and S2) and focused on the extent of gene flow in the "pure" Caucasian wild apple *M. orientalis*. We detected a significant but weak isolation by distance pattern across the Caucasus ($P < 0.001$, *R-squared* = 0.07, Figure S7), suggesting a high level of gene flow among the sampled geographic sites. We estimated *Sp* values for populations with at least five sampling sites and 20 individuals, *i.e.*, the Hyrcanian (pink) and the Central Armenian (blue) wild apple populations. *Sp* values were low but significant ($Sp_{\text{Hyrcanian\_pink}} = 0.0076$, $Sp_{\text{Central\_blue}} == 0.0027$, $P < 0.001$) suggesting a high level of historical gene flow within populations. However, the *Sp* value was higher for the Iranian population than for the Armenian population suggesting a lower level of historical gene flow within the Hyrcanian (pink) population than the Central Armenian (blue) population. Our results therefore suggest substantial crop-crop, crop-wild and wild-wild gene flow in apples in Iran and the Caucasus.

Range expansion and contraction of the wild apple *M. orientalis* in the Caucasus associated with the last glacial maximum

We investigated the spatial variation of genetic diversity and used ecological niche modeling to test the existence of past range contraction and expansion for the wild apple in the Caucasus. After removing the 150 crop-wild hybrids identified from the second STRUCTURE analyses, we found a significant positive correlation between longitude and allelic richness (Figure S8, average adjusted *R-squared* = 0.66, $P < 0.0001$) and a significant negative correlation between latitude and allelic richness (Figure S8, average adjusted $R^2 = -0.43$, $P < 0.001$). We also found that the western (orange) population had the highest level of allelic richness (Tables 1 and S7 and Figure 4). The western Caucasus may therefore have been a glacial refugium in the past. In addition to high levels of genetic diversity in the west, across northeastern Turkey and the Lesser Caucasus mountains in Armenia, we observed local hotspots of genetic diversity in the Hyrcanian Forests and the High Caucasus mountains (Figure 4) suggesting that these mountainous regions may have been potential glacial refugia.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ecological niche modeling further indicated past contraction and expansion of the *M. orientalis* range. Model performance as assessed with AUC and TSS was high (Table S11), indicating that the ANN, GLM and GAM algorithms fitted the data well (Allouche, Tsoar, & Kadmon, 2006; Fieldings & Bell, 1997; Monserud & Leemans, 1992). The following six bioclimatic variables were found to have high predictive power: mean diurnal range temperature (bio2), temperature seasonality (bio4), minimum temperature of the wettest quarter (bio8), minimum temperature of the driest quarter (bio9), annual precipitation (bio12) and precipitation of the coldest quarter (bio19). These bioclimatic variables were used to calibrate the models to predict the past and present distribution of *M. orientalis*. The MIROC model (Figure 4) predicted that the areas suitable for *M. orientalis* during the LGM contracted to the western Lesser Caucasus and northeastern Turkey along the Black Sea and into the Colchis region, and also in the eastern part of the Hyrcanian Forests, near Azerbaijan, in agreement with the genetic data (Figure 4). The climatic model therefore suggested that populations of the Caucasian wild apple *M. orientalis* may have been maintained in at least two glacial refugia.

Discussion

Our study provides insights into the natural and anthropogenic divergence history of apples in a hotspot of crop diversity, the Caucasus. First, we identified two distinct cultivated gene pools in Iran that were well differentiated from the standard *M. domestica* apple cultivars and were not derived from *M. sieversii* or *M. domestica*. Our results thus suggest that Iran is an additional center of apple domestication. Second, we showed that the evolution of the domesticated apple involved an additional wild species: *M. orientalis*. Indeed, cultivated apple from the Caucasus belonged to the Caucasian *M. orientalis* gene pool, suggesting that local farmers use the Caucasian crabapple for cultivation. Furthermore, *M. orientalis* has contributed to the Caucasian cultivated apple germplasm through wild-to-crop introgression; a similar process has been previously described in apples in Europe (Cornille et al., 2014, 2012). Reciprocally, we detected crop-to-wild gene flow in the Caucasus, as it has been reported in Europe (Cornille et al., 2015). Third, « pure » *M. orientalis* in this region showed a clear spatial genetic structure with at least seven populations spread across the Caucasus. The combination of niche modeling and population genetics approaches suggested that these populations resulted from range contraction and expansion associated with the last glaciation. Thus, this study revealed the existence of an additional center of apple domestication, which seems to have followed processes that are strikingly similar to those observed in Europe, *i.e.*, substantial wild-to-crop and crop-crop gene flow. Several origins of domestication with substantial hybridization between domestic and wild forms have also been described in other fruit trees (Groppi, Liu, Cornille, Decroocq, & Decroocq, 2021; Liu et al., 2019; Wu et al., 2018).

An additional center of apple domestication in Iran, with two specific cultivated gene pools that originated independently

The occurrence of two specific cultivated populations in Iran, which are genetically differentiated from the cultivated apple *M. domestica* and are not derived from *M. sieversii* or *M. domestica*, suggests that Iran is an additional center of apple domestication. Domesticated populations are expected to be nested within their source population because they recently diverged from a subset of individuals within the source population (Matsuoka et al., 2002); however, the green Iranian cultivated population was not nested within any wild population (by contrast, *M. domestica* that was nested within *M. sieversii*) suggesting that, despite our considerable sampling effort, we have so far been unable to identify the wild population that is most closely related to this cultivated (green) population. The purple Iranian cultivated population was nested within the wild purple population, but the status of this wild purple population (either feral or wild trees that are cultivated) remains unclear as we lack phenotypic data (see below). However, the monophyly of the two Iranian cultivated groups (green and purple) suggests that they diverged from the same wild population, perhaps corresponding to two successive domestication steps, or represent independent domestication events from the same progenitor. We tested these hypotheses and inferred more precisely the Iranian cultivated apple domestication history using coalescent-based methods combined with approximate Bayesian computation. We confirmed that the two Iranian cultivated populations did not diverge from *M. sieversii* or *M. domestica*, but independently from the same ancestral, unidentified population. Thus, despite the spread of the cultivated apple *M. domestica* along the Silk Trade Routes that crossed Iran and the South Caucasus to reach Turkey (Canepa, 2010; Spengler, 2019), specific domestication events in this region led to local cultivated apple gene pools.

The occurrence of independent domestication events in many crop species is a source of ongoing debate (Besnard et al., 2018; Choi et al., 2017; Gros-Balthazard & Flowers, 2021). In

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

apricot and pear, there is evidence of independent domestication events in Europe and Asia (Groppi et al., 2021; Liu et al., 2019; Wu et al., 2018). Although we could not identify the source population of the Iranian cultivated apple populations, our demographic inferences provide strong support for domestication from an ancestral gene pool other than *M. sieversii* from Kazakhstan. Distinct genetic ancestries may not reflect independent *de novo* domestication, but may instead represent a single domestication event with multiple origins (Choi et al., 2017; Gros-Balthazard & Flowers, 2021). We cannot ascertain that there were independent domestication events as we could not identify the source population. However, our data shows that the cultivated apple had multiple origins and was derived from distinct ancestral gene pools, irrespective of whether the ancestral populations were independently domesticated. Studies of the regions under selection during domestication in the Iranian and European cultivated apple genomes will likely provide insights into whether multiple *de novo* domestication events occurred. Independent selection regimes in each of the cultivated genetic groups would be a hallmark of multiple domestications.

The genetic relationships between the two wild contributors and the status of the wild Hyrcanian purple population are still unclear

The relationship between *M. sieversii* and *M. orientalis* is still unclear. We found that *M. sieversii* from Kazakhstan was nested within the *M. domestica* gene pool, while *M. sieversii* from Kyrgyzstan was closer to *M. orientalis* from the Zagros and Hyrcanian regions. *Malus orientalis* thus seems to be polyphyletic and intermingled with *M. sieversii*. Deciphering the relationship between *M. sieversii* and *M. orientalis*, and even the species status of *M. orientalis,* is not only a taxonomic exercise but is needed to better understand apple domestication. This question needs to be resolved urgently, *M. sieversii* being endangered across its distribution (Omasheva et al., 2017; H.-X. Zhang, Li, & Li, 2018), as confirmed here by the lower level of genetic diversity of the *M. sieversii* population in Kyrgyzstan. Likewise, the status of the wild purple population in the Hyrcanian Forests is still unclear. Individual trees are isolated in mountainous areas, but found close to grafted fruit trees on *Craetagus*, although they are not cultivated (personal observation. H. Yousefzadeh). Like domestication, feralization can be seen as a process that is accompanied by admixture and introgression, and can be accompanied by a range of genetic, phenotypic and demographic changes (Mabry, Rowan, Pires, & Decker, 2021). The purple wild population had a lower level of genetic diversity than the other wild Iranian population, which

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

may suggest that this population has recently escaped from cultivation. Feral populations have also been found in other fruit trees including olive (Besnard et al., 2018), almond (Balaguer-Romano et al., 2021) and apricot (Robert N. Spengler, Chang, & Tourtellotte, 2013). However, the purple population found in the Hyrcanian Forests could also be the result of a first domestication step, as wild apples are widely used for cultivation in Iran (personal communication H. Yousefzadeh). Additional phenotyping of cultivated and wild Iranian populations, and possibly historical records, are now required to better understand whether those trees represent an ongoing domestication event or form a feral population.

Malus orientalis is an additional contributor to the cultivated apple gene pool

It seems that local farmers in the Caucasus grow the local wild *M. orientalis* rather than *M. domestica*. Indeed, the four Armenian cultivars shared their gene pools with the Western (orange) and Central (blue) Caucasian wild apple populations, and Iranian cultivars were highly admixed and even in some cases fully assigned to wild populations (red and purple) in Iran. This is not surprising as this wild species can grow in mountainous areas, is highly resistant to pests, diseases and drought (Amirchakhmaghi et al., 2018; Büttner, 2001; Höfer et al., 2013; Volk et al., 2008) and has high-quality fruits that have several features that are intermediate between those of *M. sylvestris* and *M. sieversii* (Cornille et al., 2014). The use of the local wild apples has also been documented in Europe for specific purposes at different times in history (Tardío, Arnal, & Lázaro, 2020).

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

The natural divergence history of the Caucasian wild apple was shaped by the last glaciation

The climatic variations since the last glacial maximum, along with the landscape features of the Caucasus, have likely shaped the population structure and diversity of the Caucasian wild apple. We identified seven populations of *M. orientalis* in the Caucasus and Iran: one highly genetically differentiated population in the Western Caucasus (Turkey, Russia and northwestern Armenia), two in Armenia (a southern and a Central population) and four in Iran, including two in the Zagros Forest (one in the Kurdestan province and one in the Lorestan province) and two in the Hyrcanian Forests bordering the southern Caspian Sea. These wild apple populations likely arose from isolation in several refugia during the last glacial maximum. This hypothesis is supported by the observation of a large hotspot of genetic diversity located in Western Caucasus, and several local hotpots of genetic diversity in Armenia and the Hyrcanian Forests (Zazanashvili et al., 2020). Ecological niche modeling further supported the existence of strong contractions in the range of *M. orientalis* in the Western Caucasus bordering the Black Sea (including the Colchis region), as well as in the Lesser Caucasus and in some parts of the Hyrcanian Forests. Additional samples from the Western Caucasus are required to confirm this hypothesis. These glacial refugia have been described in relation to other species (Parvizi et al., 2019). Indeed, two refugia are recognized in the Caucasus (Tarkhnishvili *et al.* 2012; Yousefzadeh *et al.* 2012; Bina *et al.* 2016; Aradhya *et al.* 2017): a major forest refugium between the western Lesser Caucasus and northeastern Turkey (including the Colchis region in the catchment basin of the Black Sea) and the Hyrcanian refugium at the southern edge of the Caucasus. Further sampling of *M. orientalis* in the far Western and Eastern Caucasus and genotyping with the same microsatellite markers is needed to uncover the role of these two refugia for *M. orientalis*.

We also found that the natural divergence history of the Caucasian wild apple involved gene flow across the Caucasus. The weak but significant isolation-by-distance pattern further supported the existence of substantial gene flow among wild apple populations in the Caucasus. Widespread gene flow during divergence associated with the last glacial maximum has been documented for another wild apple relative *M. sylvestris* (Cornille et al., 2013). Calculation of the *Sp* parameter within the largest populations revealed high levels of historical gene flow within populations. *Sp* can also be used to compare the dispersal capacities of *M. orientalis* with that of other plants (Cornille et al., 2013; Cornille et al., 2013a; Vekemans & Hardy, 2004b). The Caucasian wild apple showed dispersal capacities that were similar to previous estimates in other wild apple species and lower than that of wind-dispersed trees. Wild apples can thus spread over kilometers (Cornille et al., 2015; Feurtey, Cornille, Shykoff, Snirc, & Giraud, 2017). The spatial population structure was somewhat stronger in Iran than in Armenia suggesting lower levels of gene flow in the Hyrcanian population. In addition to having a stronger genetic structure, the Iranian populations had lower genetic diversity then the Armenian populations, especially the Zagros and Kurdestan populations. In Iran, traditional animal husbandry is a widespread practice (Soofi et al., 2018). Such intensive farming environments may lead to forest fragmentation and may impact wild apple populations, which form low density populations. The future of Iranian wild apple populations, especially in the south where genetic diversity is low, will depend on our ability to protect them through sustainable conservation programs.

Conclusion

We identified Iran as a key center in the evolution and domestication of apple, and *M. orientalis* as an additional contributor to the evolutionary history of cultivated apple. We also provided insights into the processes underlying the natural divergence of this emblematic wild species and identified several populations that could be the target of conservation programs. However, *M. orientalis* in the Caucasian ecoregion is highly diverse and further investigations and additional sampling are necessary, as well as a better assessment of its species status and genetic relationship with *M. sieversii*. Indeed, a better understanding of the properties of functional genetic diversity and the ecological relationship of wild apples in their ecosystem is needed for developing and implementing effective conservation genetic strategies in this region (Teixeira & Huber, 2021). Our study revealed the role of gene flow and human practices in natural and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

anthropogenic divergence processes of an emblematic fruit tree in the Caucasus. Our results are consistent with those reported for other woody perennials, including apricots (Groppi et al., 2021; Liu et al., 2019), olive (Besnard et al., 2018; Diez et al., 2015), pears (Volk & Cornille, 2019; Wu et al., 2018), or date palms (Flowers et al., 2019). This study also supports the view that domestication of fruit trees was likely a geographically diffuse and protracted process, involving multiple, geographically disparate origins of domestication (Groppi et al., 2021; Wu et al., 2018).

Acknowledgements

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

Data Availability

SSR data are available on the DRYAD repository XXXX.

Author Contributions

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

Figures and Tables

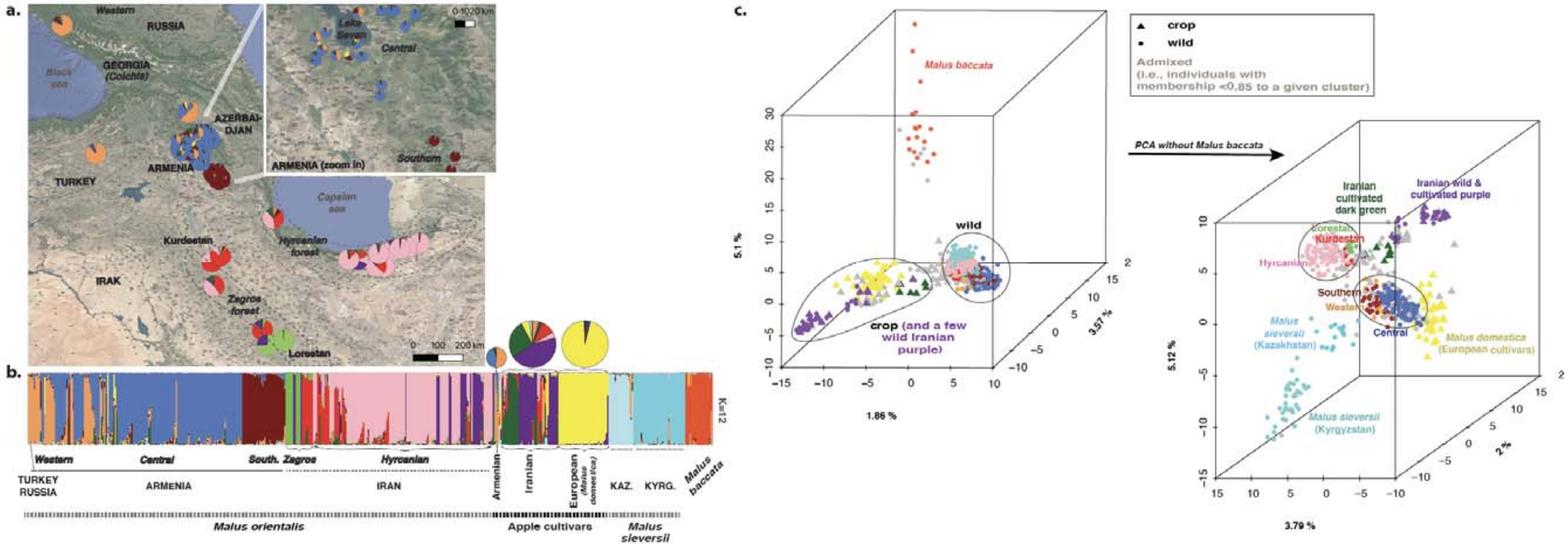


Figure 1. Population genetic structure and differentiation in cultivated and wild apples from the Caucasus, *Malus domestica*, *Malus sieversii* and *Malus baccata* based on 26 microsatellite markers. a. Spatial population genetic structure inferred with STRUCTURE at $K = 12$ ($N = 550$); the map represents membership proportions averaged over each geographic site for the Caucasian wild apple *M. orientalis* ($N = 374$, 43 sites across Turkey, Russia, Armenia and Iran). In the bottom right corner, the mean membership proportions for the apple cultivars from Armenia ($N = 4$), Iran ($N = 48$) and Europe (*M. domestica*, $N = 40$). The size of the pie charts is proportional to the number of samples per site. **b.** STRUCTURE bar plot ($N = 550$) at $K = 12$ showing 12 distinct genetic clusters. Each vertical line represents an individual. Colors represent the inferred ancestry from K ancestral genetic clusters. Sites are grouped by country for the wild apple samples (*i.e.*, Turkey, Russia, Armenia, Iran) and *M. sieversii* (*i.e.*, Kazakhstan and Kyrgyzstan), apple cultivars are grouped according to their origin: Armenia ($N = 4$), Iran ($N = 48$) and *M. domestica* ($N = 40$). Countries (Kazakhstan,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Kyrgyzstan, Armenia) and/or main regions in the Caucasus (the Western Caucasus, *i.e.*, Turkey and Russia, Zagros and Hyrcanian Forests, Central and Southern Armenia) are shown on the map. Reference samples from previously published studies of each species are: *M. orientalis* from the Western Caucasus and Central and Southern Armenia, *M. domestica* (European cultivars), *M. sieversii* from Kazakhstan (Cornille et al., 2013a) and *M. baccata* (Cornille et al., 2012) c. Principal component analysis (PCA) of 550 individuals (upper left), and after removing the outgroup *M. baccata* (lower right, N = 530), with the respective total variation explained by each component.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

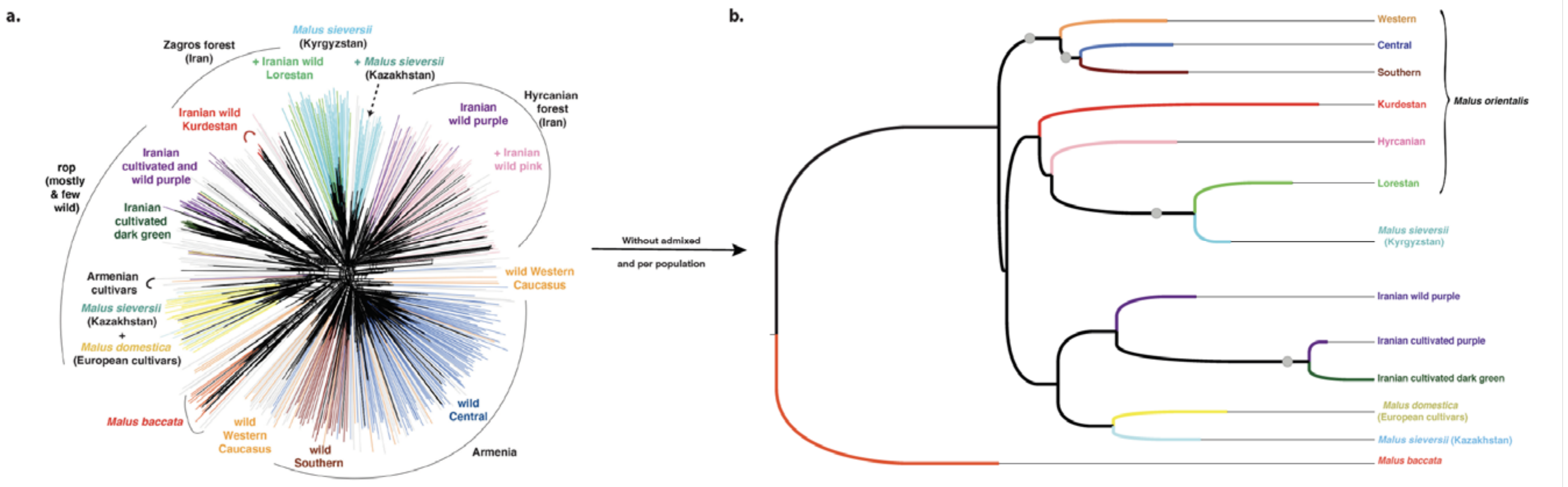
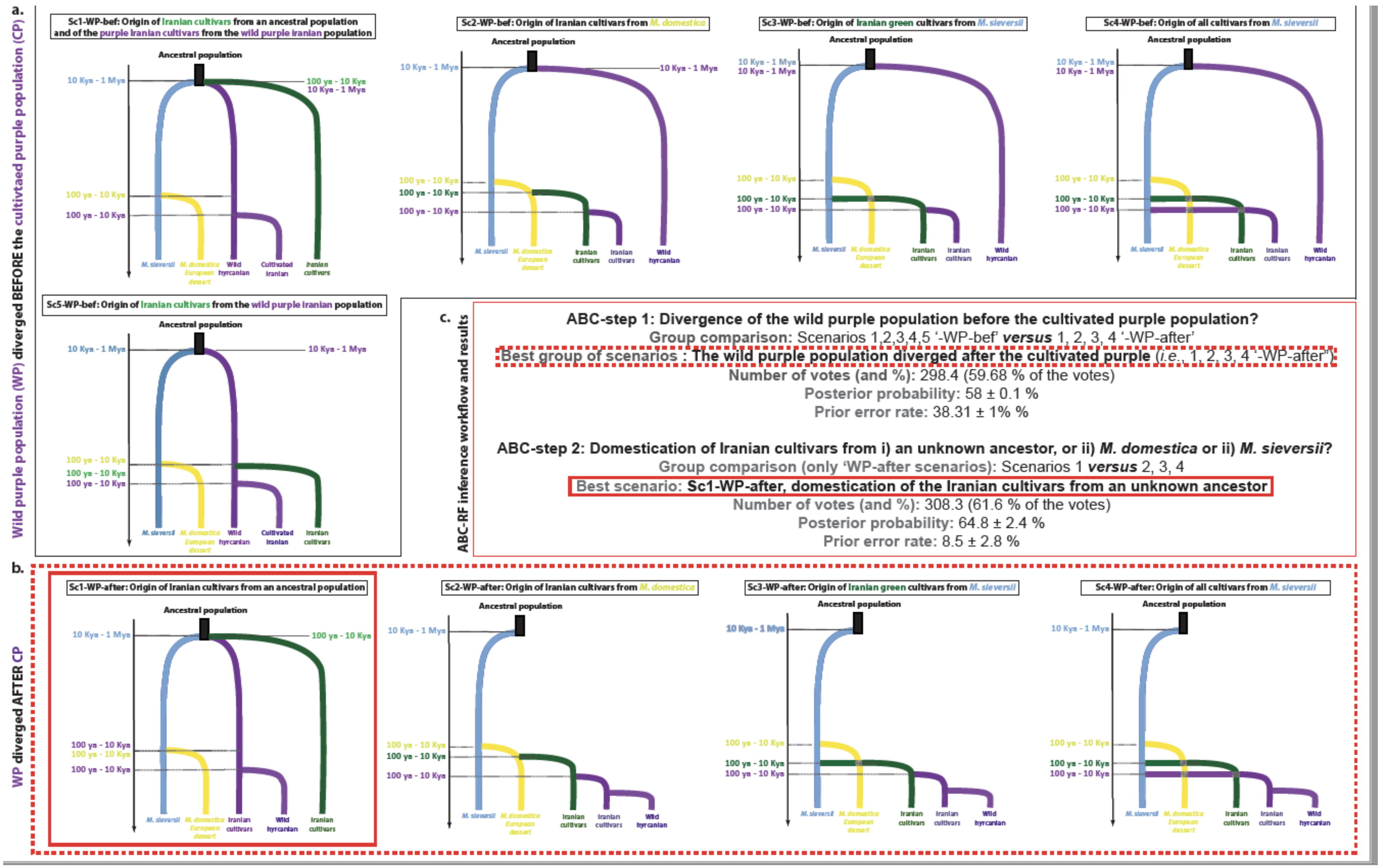


Figure 2. Genetic differentiation among cultivated and wild apples from the Caucasus, *Malus domestica*, *Malus sieversii* and *Malus baccata* based on 26 microsatellite markers. a. Neighbor-net representing the genetic relationships among wild and cultivated individuals inferred with STRUCTURE at $K = 12$. Colors correspond to the genetic groups inferred with STRUCTURE at $K = 12$ and admixed samples are in grey. b. Neighbor-joining tree representing the distance among the twelve populations detected with STRUCTURE at $K = 12$, excluding admixed individuals (*i.e.*, individuals with a membership coefficient < 0.85 to any given cluster), and rooted with *M. baccata*. Each branch is coloured according to the population inferred with STRUCTURE at $K = 12$, nodes with a grey circle represent bootstrap values > 80%.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



ABC-step 1: Divergence of the wild purple population before the cultivated purple population?
Group comparison: Scenarios 1,2,3,4,5 '-WP-bef' ***versus*** 1, 2, 3, 4 '-WP-after'
Best group of scenarios : **The wild purple population diverged after the cultivated purple** (*i.e.*, 1, 2, 3, 4 '-WP-after")
Number of votes (and %): 298.4 (59.68 % of the votes)
Posterior probability: 58 ± 0.1 %
Prior error rate: 38.31 ± 1% %

ABC-step 2: Domestication of Iranian cultivars from i) an unknown ancestor, or ii) *M. domestica* or ii) *M. sieversii*?
Group comparison (only 'WP-after scenarios): Scenarios 1 ***versus*** 2, 3, 4
Best scenario: **Sc1-WP-after, domestication of the Iranian cultivars from an unknown ancestor**
Number of votes (and %): 308.3 (61.6 % of the votes)
Posterior probability: 64.8 ± 2.4 %
Prior error rate: 8.5 ± 2.8 %

Figure 3. Inference of independent domestication events of cultivated apples in Iran using random-forest approximate Bayesian computation (ABC-RF) combined with coalescent-based simulations. a. Five scenarios testing the domestication history of Iranian cultivars (purple and green) assuming the wild Iranian purple population diverged before the cultivated purple population

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(sc1-WP-bef, sc2-WP- bef, sc3-WP- bef, sc4-WP- bef, sc5-WP- bef). b. Four scenarios testing the domestication history of Iranian cultivars (purple and green) assuming the wild Iranian purple population diverged more recently than the cultivated purple (sc1-WP-after, sc2-WP-after, sc3-WP-after, sc4-WP-after). For a and b, scenarios assumed bidirectional gene flow between crop and wild populations. **c**. Most likely groups of scenarios (step 1) or scenario (step 2) of Iranian apple domestication inferred from ABC-RF analyses. The corresponding most likely scenarios, or groups of scenarios, are framed in red (dashed or solid lines for steps 1 and 2, respectively). Population names correspond to the ones detected with STRUCTURE for K = 12, excluding admixed individuals (*i.e.*, individuals with a membership coefficient < 0.85 to any given cluster).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

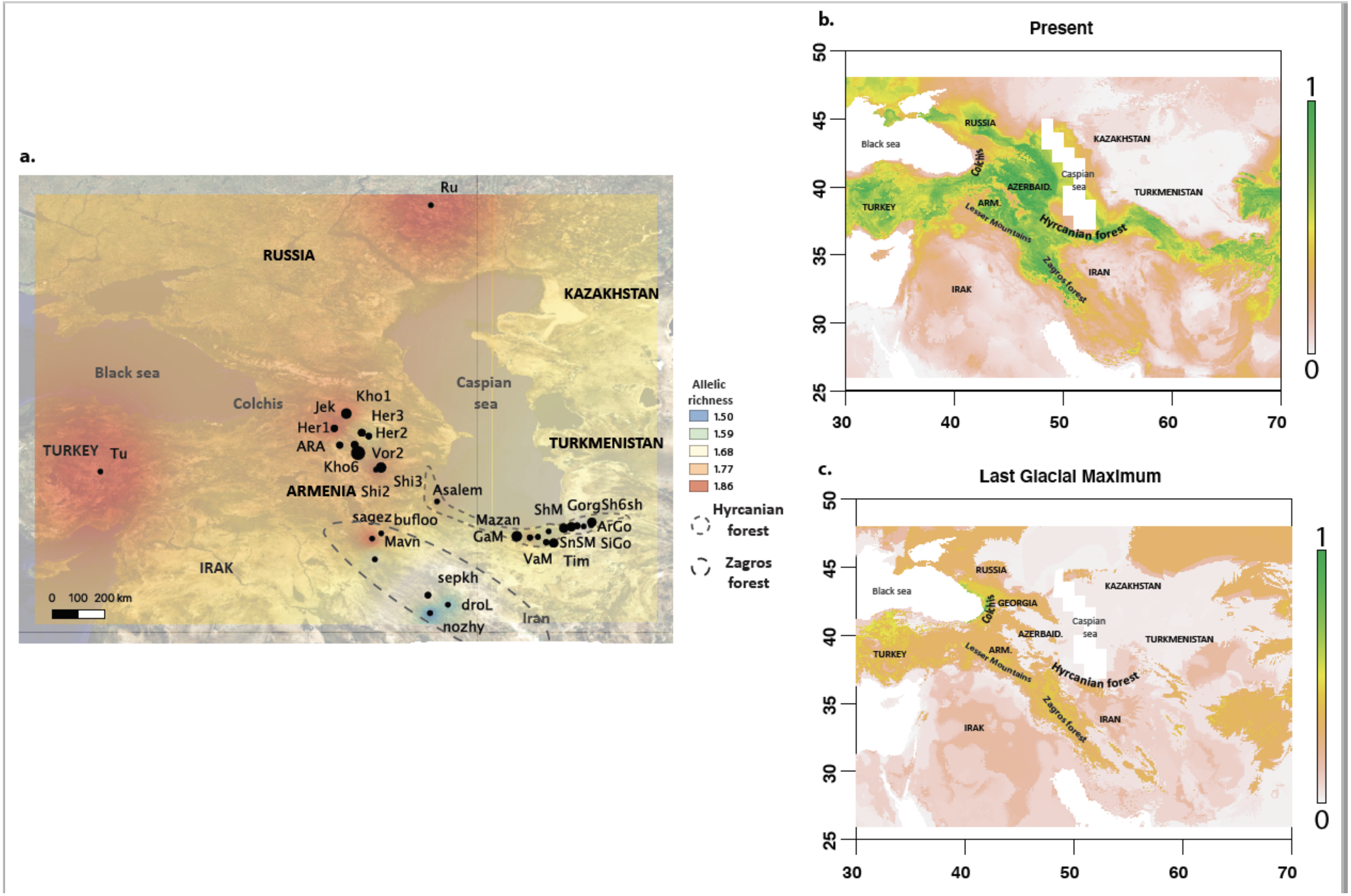


Figure 4. Spatial diversity and past contraction and expansion of *Malus orientalis* across the Caucasus. Spatial genetic diversity

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(allelic richness) at 36 sites (*N* = 339). b. Ensemble forecasting of the three different algorithms (ANN, GLM and GAM) predicting the current (a) and last glacial maximum (LGM) (b) distribution range of suitable areas for *M. orientalis*. The probabilities of being a suitable habitat are given in the legend. **The Colchis and Hyrcanian regions are shown on the maps.**

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Wild or cultivated	Species	Country of origin	Population	N	H_O	H_E	F_{IS}	A_R ($G=7$)	A_P ($G=7$)
wild	*Malus orientalis*	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
	Malus sieversii	Kazahkstan	(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
	Malus baccata	Russia	(light red)	17	0.52	0.55	0.06^{NS}	3.23±0.19	1.08±0.14
cultivated	*Malus domestica*	Mostly Europe	European cultivars	38	0.78	0.78	-0.00^{NS}	4.27±0.09	0.85±0.10

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

	?	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**					

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 2. Distribution of hybrids (*i.e.*, individuals with a membership coefficient < 0.90 to any given genetic cluster, as inferred with STRUCTURE for *K* = 9) in cultivated and wild apple in the Caucasus (*N* = 466, 26 microsatellite markers). For each group (cultivated or wild, from different regions), N_{tot} is the total number of samples in each group, *N* is the number of hybrids assigned to each gene pool and % is the respective percentage over the total number of samples from each group, the mean introgression rate is the mean membership coefficient to this gene pool. Note that some admixed trees were assigned to several gene pools with a membership coefficient < 0.90; the total number of hybrids associated with each cluster (TOTAL) is given on the last line of the table. We also showed the distribution of crop-crop, crop-to-wild, wild-to-crop and wild-to-wild hybrids.

		Crop				Wild			
Genepool		Armenia $N_{tot}=3$	Iran $N_{tot}=48$	*M. domestica* $N_{tot}=40$		Armenia $N_{tot}=196$	Iran $N_{tot}=167$	Russia $N_{tot}=5$	Turkey $N_{tot}=6$
Malus domestica	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**

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

			)		%)		)		
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%)

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

		)		%)	)	)	)		
Total number of hybrids (and %)	**1**	**32 (66%)**	**2 (5%)**		**58 (28%)**	**54 (32%)**	**1 (20%)**	**2 (33%)**	**150 (32%)**

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, *43*(6), 1223–1232.

Amirchakhmaghi, N., Yousefzadeh, H., Hosseinpour, B., Espahbodi, K., Aldaghi, M., & Cornille, A. (2018). First insight into genetic diversity and population structure of the Caucasian wild apple (*Malus orientalis* Uglitzk.) in the Hyrcanian forest (Iran) and its resistance to apple scab and powdery mildew. *Genetic Resources and Crop Evolution*, *65*(4), 1255–1268. doi: 10.1007/s10722-018-0611-z

Arroyo-García, R., Ruiz-García, L., Bolling, L., Ocete, R., López, M. A., Arnold, C., … Martinez-Zapater, J. M. (2006). Multiple origins of cultivated grapevine (Vitis vinifera L. ssp. Sativa) based on chloroplast DNA polymorphisms. *Molecular Ecology*, *15*(12), 3707–3714. doi: 10.1111/j.1365-294X.2006.03049.x

Asanidze, Z., Akhalkatsi, M., Henk, A. D., Richards, C. M., & Volk, G. M. (2014a). Genetic relationships between wild progenitor pear (*Pyrus* L.) species and local cultivars native to Georgia, South Caucasus. *Flora: Morphology, Distribution, Functional Ecology of Plants*, *209*(9), 504–512. doi: 10.1016/j.flora.2014.06.013

Asanidze, Z., Akhalkatsi, M., Henk, A. D., Richards, C. M., & Volk, G. M. (2014b). Genetic relationships between wild progenitor pear (*Pyrus* L.) species and local cultivars native to Georgia, South Caucasus. *Flora - Morphology, Distribution, Functional Ecology of Plants*, *209*(9), 504–512. doi: https://doi.org/10.1016/j.flora.2014.06.013

Bacles, C. F. E., & Jump, A. S. (2011). Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science*, *16*(1), 13–18. doi: 10.1016/j.tplants.2010.10.002

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

Barbet-Massin, M., Jiguet, F., Albert, C. H., & Thuiller, W. (2012). Selecting pseudo-absence for species distribution models: How, where and how many? *Methods in Ecology and Evolution*, 3: *327–338*.

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Gharghani, Ali, Zamani, Z., Talaie, A., Oraguzie, N. C., Fatahi, R., Hajnajari, H., … Gardiner, S. E. (2009). Genetic identity and relationships of Iranian apple (Malus× domestica Borkh.) cultivars and landraces, wild Malus species and representative old apple cultivars based on simple sequence repeat (SSR) marker analysis. *Genetic Resources and Crop Evolution*, *56*(6), 829–842.

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

Parvizi, E., Keikhosravi, A., Naderloo, R., Solhjouy-Fard, S., Sheibak, F., & Schubart, C. D. (2019). Phylogeography of Potamon ibericum (Brachyura: Potamidae) identifies Quaternary glacial refugia within the Caucasus biodiversity hot spot. *Ecology and Evolution*, *9*(8), 4749–4759. doi: 10.1002/ece3.5078

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

Spengler, Robert N., Chang, C., & Tourtellotte, P. A. (2013). Agricultural production in the Central Asian mountains: Tuzusai, Kazakhstan (410–150 b.c.). *Journal of Field Archaeology*, *38*(1), 68–85. doi: 10.1179/0093469012Z.00000000037

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

Tian, F., Li, B., Ji, B., Zhang, G., & Luo, Y. (2009). Identification and structure-activity relationship of gallotannins separated from Galla chinensis. *LWT - Food Science and Technology*, *42*(7), 1289–1295. doi: 10.1016/j.lwt.2009.03.004

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

Vavilov, N. I. (1992). *Origin and geography of cultivated plants.* Cambridge: Cambridge University Press. CABDirect.

Vekemans, X., & Hardy, O. J. (2004a). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935. doi: 10.1046/j.1365-294X.2004.02076.x

Vekemans, X., & Hardy, O. J. (2004b). New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, *13*(4), 921–935.

Volk, G. M., & Cornille, A. (2019). Genetic Diversity and Domestication History in Pyrus. In *The Pear Genome* (pp. 51–62). Springer.

Volk, G. M., Richards, C. M., Reilley, A. A., Henk, A. D., Reeves, P. A., Forsline, P. L., & Aldwinckle, H. S. (2008). Genetic diversity and disease resistance of wild Malus

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

orientalis from Turkey and Southern Russia. *Journal of the American Society for Horticultural Science*, *133*(3), 383–389.

Vouillamoz, J. F., McGovern, P. E., Ergul, A., Söylemezoğlu, G., Tevzadze, G., Meredith, C. P., & Grando, M. S. (2006). Genetic characterization and relationships of traditional grape cultivars from Transcaucasia and Anatolia. *Plant Genetic Resources*, *4*(2), 144–158. doi: 10.1079/pgr2006114

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

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

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

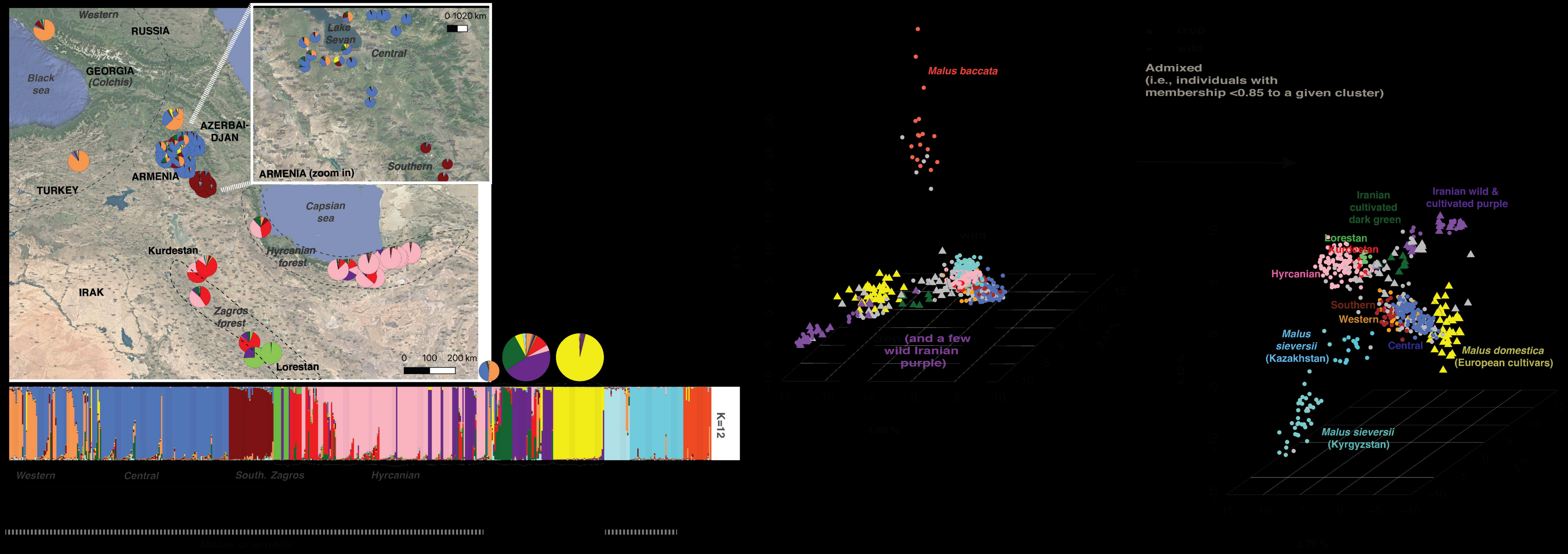
Zazanashvili, N., Sanadiradze, G., Garforth, M., Bitsadze, M., Manvelyan, K., Askerov, E., … Devranoğlu Tavsel, S. (2020). *Ecoregional Conservation Plan (ECP) For The Caucasus 2020 Edition*. WWF, KfW, Tbilisi.

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

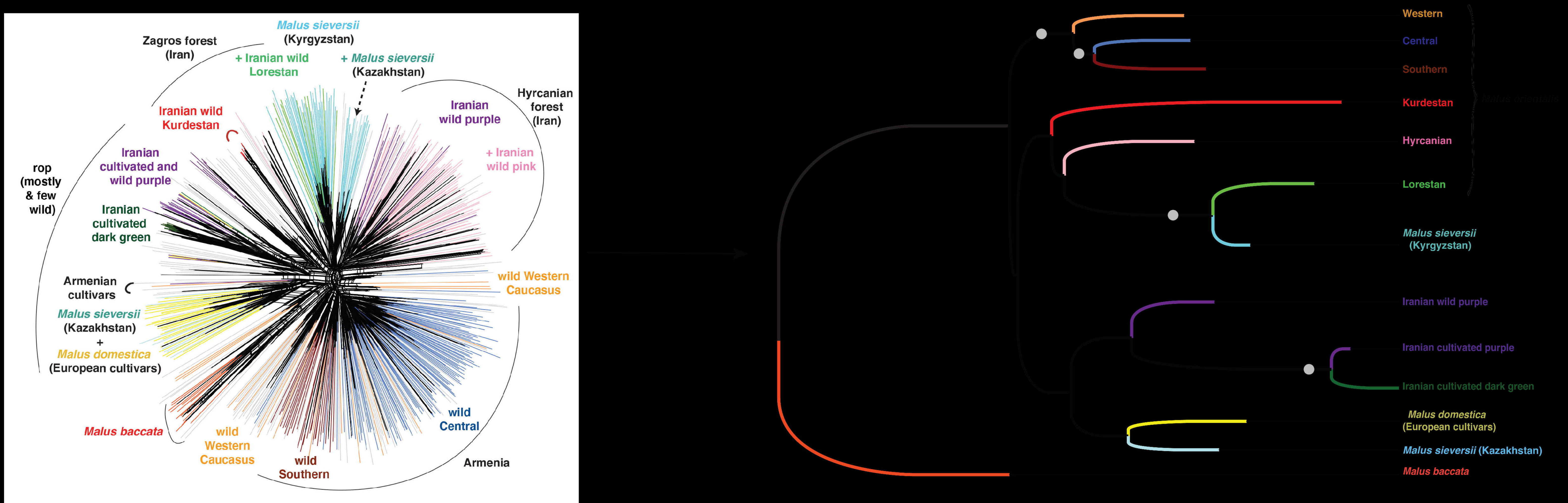
bioRxiv preprint doi: https://doi.org/10.1101/2021.03.28.437401; this version posted August 31, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

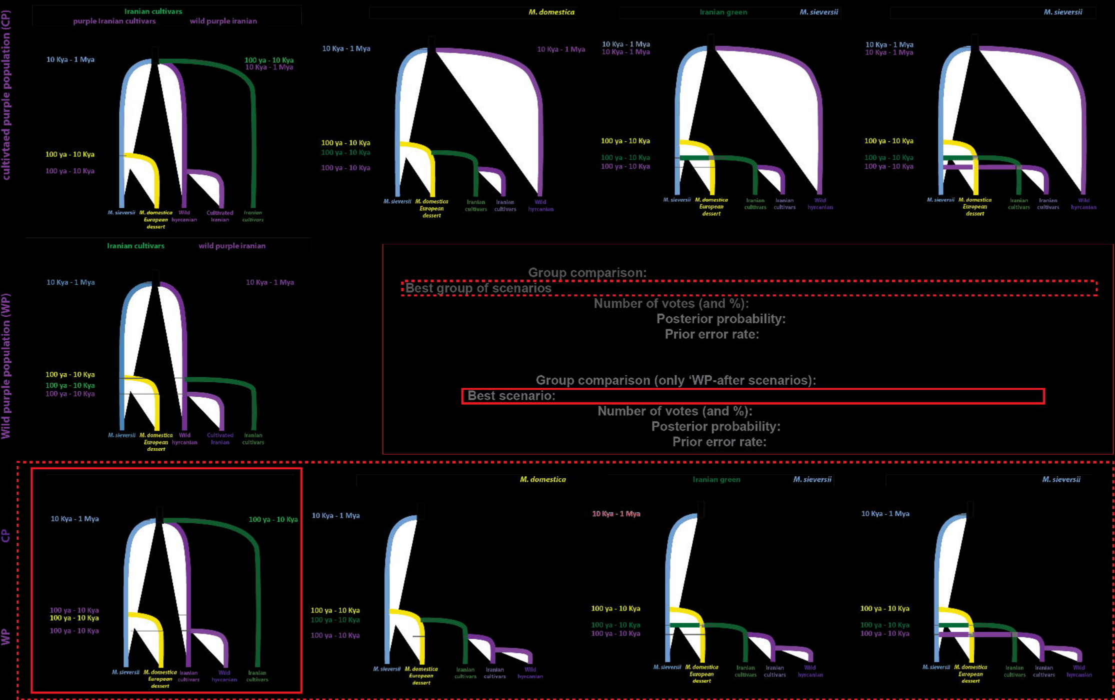
Zhang, H., Mittal, N., Leamy, L. J., Barazani, O., & Song, B. H. (2017). Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications*, *10*(1), 5–24. doi: 10.1111/eva.12434

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


Western
RUSSIA
GEORGIA
(Colchis)
Black
sea
AZERBAI-
DJAN
ARMENIA
TURKEY
Lake
Sevan
Central
0 10 20 km
ARMENIA (zoom in)
Southern
Capsian
sea
Hyrcanian
forest
Kurdestan
IRAK
Zagros
forest
Lorestan
0 100 200 km
K=12
Western
Central
South. Zagros
Hyrcanian
Malus orientalis
Malus baccata
crop
wild
Admixed
(i.e., individuals with
membership <0.85 to a given cluster)
wild
crop (and a few
wild Iranian
purple)
1.86 %
5.1 %
Iranian
cultivated
dark green
Iranian wild &
cultivated purple
Lorestan
Kurdestan
Hyrcanian
Southern
Western
Central
Malus
sieversii
(Kazakhstan)
Malus domestica
(European cultivars)
Malus sieversii
(Kyrgyzstan)
5.12 %
3.79 %

https://doi.org/10.1101/2021.03.28.437401





Group comparison:

Best group of scenarios

Number of votes (and %):

Posterior probability:

Prior error rate:

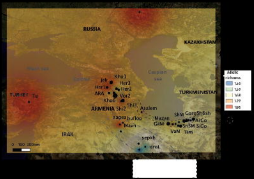
Group comparison (only ‘WP-after scenarios):

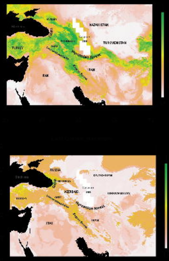
Best scenario:

Number of votes (and %):

Posterior probability:

Prior error rate:


RUSSIA
KAZAKHSTAN
TURKMENISTAN
TURKEY
Tu
ARMENIA
IRAK
Black sea
Caspian sea
Kho1
Jek
Her3
Her1
Her2
ARA
Vor2
Kho6
Shi3
Shi2
Asalem
sapez
bufloo
Mavn
Mazan
GaM
ShM
GorgSh6sh
ArGo
SnSM
SiGo
VaM
Tim
sepkh
droL
0 100 200 km
Allelic richness
1.50
1.59
1.68
1.77
1.86


RUSSIA
KAZAKHSTAN
TURKMENISTAN
TURKEY
IRAK
IRAN
AZERBAID.
Caspian sea
Hyrcanian forest
Zagros forest