Title : An additional area of apple domestication with crop-wild gene flow, and also cultivation of the local wild apple, in the Caucasus

Short title: An additional area of 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], Giraud Tatiana[5], Hossainpour Batol[6], Abdollahi Hamid[7], Gabrielyan Ivan[8], Nersesyan Anush[9], 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] Ecologie Systematique Evolution, Universite Paris-Saclay, CNRS, AgroParisTech, Orsay, France
[6] Iranian Research Organization for Science and Technology, Tehran, Iran
[7] Temperate Fruits Research Center, Horticultural Sciences Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Karaj, Iran
[8] 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
[9] Department of Palaeobotany, A. Takhtajyan Institute of Botany, Armenian National Academy of Sciences, Acharyan Str.1, 0040 Yerevan, Armenia

Corresponding authors: Amandine Cornille, amandine.cornille@gmail.com, hamed.yousefzadeh@modares.ir

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

Summary

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 from using 26 microsatellite markers amplified on 508 wild and cultivated samples. We found two specific Iranian cultivated populations that were differentiated from *Malus domestica*, the standard cultivated apple worldwide, suggesting a specific local domestication process in Iran. Some Iranian apple cultivars belonged to the Caucasian wild apple gene pools, indicating that farmers also use local wild apple for cultivation. Substantial wild-crop and crop-crop gene flow were also inferred. We identified seven genetically differentiated populations of wild apples (*Malus orientalis*) in the Caucasus. Niche modeling indicated that these populations likely resulted from range changes linked to the last glaciation. This study pinpoints Iran as a key region in the evolution and domestication of apple and further demonstrates the role of gene flow during fruit tree domestication as well as the impact of climate change on the natural divergence of a wild fruit tree. The results also provide a practical base for apple conservation and breeding programs in the Caucasus.

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 (Cornille *et al.* 2012, 2014; Diez *et al.* 2015; Gaut *et al.* 2015; Brandenburg *et al.* 2017; Besnard *et al.* 2018; Chen *et al.* 2019; Flowers *et al.* 2019). On the other hand, wild species allow the study of natural divergence over a longer timescale. Indeed, wild species have often undergone shifts in distribution following past climate changes associated with glacial periods, and range contraction has often led to population differentiation and divergence (Hewitt 1990, 1996; Petit *et al.* 2004; Schmitt 2007; Excoffier *et al.* 2009; Jezkova *et al.* 2011). 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 *devenir* 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) (Castañeda-Álvarez *et al.* 2016; Zhang *et al.* 2017; Bailey-Serres *et al.* 2019).

Fruit trees present several historical and biological features that make them fascinating models for investigating anthropogenic and natural divergence with gene flow. The origin of fruit trees is linked to the emergence of some of the most ancient civilizations (Zohary & Spiegel-Roy 1975; Vavilov 1992). Several wild-crop fruit tree complexes are now spread across the world and can be good model systems to study anthropogenic and natural divergence in trees. Fruit trees are also characterized by high levels of gene flow during divergence, which is expected considering the typical life history traits of trees (Petit & Hampe 2006; Oddou-Muratorio & Klein 2008; Cornille *et al.* 2013a, b). Population genetic 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 *et al.* 2004; Petit *et al.* 2004)) and animal-dispersed trees (Cornille *et al.* 2013b) demonstrated high levels of gene flow between populations as well as high dispersal capabilities. These studies also revealed the location of single (Tian *et al.* 2009; Bai & Spitkovsky 2010; Zeng *et al.* 2011) or multiple (Tian *et al.* 2009; Qiu *et al.* 2011) 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 *et al.* 2017). Population genetic and genomic 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; Myles *et al.* 2011; Cornille *et al.* 2012; Meyer *et al.* 2012a; Diez *et al.* 2015; Decroocq *et al.* 2016; Duan *et al.* 2017; Liu *et al.* 2019). These studies thus revealed that domestication of fruit trees involved a specific process that is different from that of annuals, and which can be explained by the long lifespan, long juvenile phase and self-incompatibility system of trees (Gaut *et al.* 2015; Fuller 2018). 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 economically important plants and their wild relatives, in particular wheat, rye, barley and also fruit trees including walnut, apricot and apple (Gabrielian & Zohary 2004; Yousefzadeh *et al.* 2012; Asanidze *et al.* 2014a). This region covers Georgia, Armenia, Azerbaijan, the North Caucasian part of the Russian Federation, the northern-eastern part of Turkey and the Hyrcanian Mixed Forests region in northwestern Iran (Nakhutsrishvili *et al.* 2015; Zazanashvili *et al.* 2020). Two refugia for temperate plants are recognized in this region (Yousefzadeh *et al.* 2012; Bina *et al.* 2016): the Colchis refugium in the catchment basin of the Black Sea, and the Hyrcanian refugium at the southern edge of the Caucasus. 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 Mixed Forests refugia. The geography of the Caucasus, with two parallel mountain chains separated by contrasted climatic zones makes this region a good model for investigating natural divergence processes associated with the last glacial maximum. Furthermore, it has 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 domestication center for apple, grape and apricot (Decroocq *et al.* 2016; Liang *et al.* 2019; Liu *et al.* 2019). However, up to now, inferences of the natural and anthropogenic divergence history of wild-crop fruit tree complexes in the Caucasus have been limited by the low number of samples (Decroocq *et al.* 2016; Liu *et al.* 2019) and/or genetic markers investigated (Vouillamoz *et al.* 2006; Gharghani *et al.* 2010; Myles *et al.* 2011; Cornille *et al.* 2013b; Asanidze *et al.* 2014b; Amirchakhmaghi *et al.* 2018; Volk & Cornille 2019).

The Caucasian crab apple, *Malus orientalis* Uglitzk., is an endemic wild apple species occurring in the Caucasus. More specifically, it is found in the southern part of Russia, northern Anatolia, Armenia, eastern Georgia, Turkey, the mountainous belt in northern Iran, the Hyrcanian Forests (Rechinger 1964; Büttner 2001) as well as in western, eastern and central Iran (Rechinger 1964; Browicz 1969). This species displays high phenotypic diversity across its distribution where it can be found 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 and color (Cornille *et al.* 2014). Fruit of *M. orientalis* are harvested across the Caucasus for stewing and processed as juice and other beverages, jelly, syrup, jam, wine and vinegar (Büttner 2001; Amirchakhmaghi *et al.* 2018). This has led some authors to suggest that some local apple cultivars from several regions of the Caucasus originated from *M. orientalis* (Langenfeld 1991; Forsline *et al.* 2003; Schmitt 2007). *Malus domestica*, the standard cultivated apple, is also currently grown in various regions of the Caucasus (Langenfeld 1991; Forsline *et al.* 2003; Schmitt 2007). So far, the relationships between *M. orientalis*, local Caucasian cultivars, *M. domestica* and its Central Asian progenitor, *Malus sieversii*, are still unknown. Two questions: i) Has *M. orientalis* contributed to the local cultivated Caucasian apple germplasm through wild-to-crop introgression, as has *M. sylvestris*, the European crab apple, to the *M. domestica* gene pool (Gharghani *et al.* 2009; Cornille *et al.* 2012)?; ii) Are cultivated apples in the Caucasus derived from the same domestication event as *M. domestica*? One study suggested a minor contribution of *M. orientalis* to Mediterranean *M. domestica* cultivars (Cornille *et al.* 2012), but without in-depth investigation. Reciprocally, the extent of crop-to-wild gene flow in apples in the Caucasus has just begun to be investigated. A population genetic 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.* 2013b). 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, respectively (Cornille *et al.* 2013b). At a smaller geographical scale, an east-west genetic structure was found across the Hyrcanian Forests in Iran, with five main populations showing admixture (Amirchakhmaghi *et al.* 2018). However, we still lack an integrated view of the genetic diversity and structure of *M. orientalis* across its distribution range to understand its natural divergence history. In addition, a study of local cultivars from the Caucasus will allow us to investigate the relationship 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 investigated anthropogenic and natural divergence processes in apples from the Caucasus, and the extent of gene flow during divergence. We used a comprehensive sample of 508 apple trees genotyped using 26 microsatellite markers, including local cultivated and wild apples from the Caucasus, as well *M. domestica* apple cultivars and *M. sieversii* samples. Note that *Malus baccata*, the Siberian wild apple, was used as an outgroup in certain analyses. With this extensive genetic dataset, combined with niche modeling approaches, we addressed the following questions: 1) Do Caucasian wild and cultivated apples, *M. domestica* and the Central Asian wild apple, *M. sieversii*, form distinct gene pools, and what is their genetic relationship? 2) What is the genetic diversity of cultivated apples in the Caucasus, *i.e.* can we detect a bottleneck in the gene pool of cultivated apples from this region? 3) Can we detect crop-crop and crop-wild gene flow in the Caucasus? 4) Can we detect the genetic consequences of the past range contraction and expansion associated with the last glacial maximum in *M. orientalis*? 5) Can we reconstruct the past habitats that were suitable for the Caucasian crab apple using ecological niche modeling methods?

Materials and methods

Sampling, DNA extraction and microsatellite genotyping

Samples from the Caucasus comprised: 207 *M. orientalis* individuals from Turkey, Armenia and Russia (23 sites, Table S1) for which data was available (published in Cornille *et al.* (2013b)), four cultivars from Armenia, 40 cultivated apple *M. domestica* individuals (N = 40, Table S2); samples from Iran, which were collected for this study, comprised: 167 *M. orientalis* individuals from the Hyrcanian Forests and the Zagros region (Table S1) and 48 local Iranian apple cultivars from the *Seed and Plant Improvement Institute* (Karaj, Iran) (Table S2). We also included previously published data from 20 *M. sieversii* individuals from Kazakhstan (Cornille *et al.* 2012) and 22 *M. baccata* individuals from Russia (Cornille *et al.* 2012). Thus, a total of 508 individuals were analyzed, comprising 374 wild *M. orientalis*, 48 Iranian and four Armenian apple cultivars, 40 European apple cultivars belonging to *M. domestica*, 20 *M. sieversii* and 22 *M. baccata* (details are provided in Tables S1 and S2).

DNA from samples of wild and cultivated apples from Iran (*N* = 215, comprising 167 *M. orientalis* and 48 Iranian cultivars) 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.®) as previously described (Patocchi *et al.* 2009; Cornille *et al.* 2012) for 26 microsatellite markers. Note that on each DNA plate, we included two positive controls, *i.e.* one sample of *M. orientalis* and one of *M. domestica* used in a previous study (Cornille *et al.* 2013b). Genotypes of the positive 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 were missing. The suitability of the markers for population genetics analyses has been demonstrated in previous studies (Cornille *et al.* 2012, 2013b, c).

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 *et al.* 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 *et al.* 2005) as implemented in the online post-processing software Structure Harvester (Earl 2012). However, the *K* identified with the *ΔK* statistic often does not correspond to the finest biologically relevant population structure ; we therefore visually checked the bar plots and chose the *K* value for which all clusters had well-assigned individuals, while no further well-delimited and biogeographically relevant clusters could be identified for higher *K* values.

We ran STRUCTURE for the whole dataset (*N* = 508) to investigate population differentiation between 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 ≥ 0.9 were colored according to the respective color of each cluster, and admixed individuals (*i.e.* individuals with a membership coefficient to any given cluster < 0.9) were colored in gray. Second, we generated a neighbor-net tree with Splitstree v5 (Huson, 1998; Huson & Scornavacca, 2012), including only individuals with a membership coefficient to a given cluster ≥ 0.9. Third, we explored the relationship among populations identified with STRUCTURE (*i.e.* clusters of individuals with a membership coefficient ≥ 0.9 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 the shared allele distance (*Das*) (Jin & Chakraborty, 1994) 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.* cluster inferred by STRUCTURE, excluding admixed individuals with a membership coefficient < 0.9) and site (*i.e.* geographical location). We calculated allelic richness (A_R) and private allelic richness (A_P) 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 and 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). Recent events of effective population size reduction were investigated with BOTTLENECK v.1.2.02 (Piry *et al.*, 1999), which compares the expected heterozygosity estimated from allele frequencies with that estimated from the number of alleles and the sample size, and which should be identical for a neutral locus in a population at mutation-drift equilibrium.

Identification of crop-wild hybrids 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 and ran STRUCTURE as above (N = 466). We defined hybrids with crop-to-wild introgression as *M. orientalis* trees assigned to the *M. domestica* or to the Iranian or Armenian cultivated gene pools with a membership coefficient > 0.10. We defined hybrids with wild-to-crop introgression as cultivars assigned to any of the wild gene pools with a membership coefficient > 0.10.

Spatial pattern of genetic diversity and historical gene flow in the Caucasian crab apple

We investigated spatial patterns of diversity and genetic differentiation in the "pure" *M. orientalis*. To that 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 variability of "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 S1) with the geometry-based inverse distance weighted interpolation in QGIS (Quantum GIS, GRASS, SAGA GIS).

We also estimated the extent of historical gene flow in *M. orientalis* 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, we computed Nason's kinship coefficient for each population (F_{ij},(Loiselle *et al.*, 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 $ln(d_{ij})$. Low *Sp* implies low spatial population structure, which suggests high historical gene flow and/or high effective population size.

Species distribution modeling

The BIOMOD2 R package (Thuiller *et al.*, 2016) (R statistical software version 3.5.3) 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 the 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 interval spatial resolution from the CCSM4 global climate model was used (https://www.worldclim.org/data/worldclim21.html#). 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 S3). 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 S4). The model was run again using the selected variables with high predictive power.

Results

Clear genetic structure and differentiation between cultivated and wild apples

The ΔK statistic was highest at $K = 3$ (Figure S2 a, b). However, STRUCTURE identified eleven well-delimited clusters at $K = 11$ (Figures 1 a, b and S3), corresponding to species and/or geographic regions (Figure 1). We therefore considered these eleven clusters as the most relevant genetic structure.

Malus sieversii and *M. baccata* formed distinct gene pools. For *M. orientalis,* we identified several distinct genetic groups: a Western genetic group with samples from Russia, Turkey and northwestern Armenia (orange), a Central Armenian group (blue), a Southern Armenian group (brown), and four genetic groups in Iran corresponding respectively to the

Lorestan province (light green), the Kurdestan province (red), and two gene pools (pink and purple) spread across the Hyrcanian Forests (Figure 1 a, b).

The *M. domestica* apple cultivars formed a distinct 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 specific to cultivars (dark green), and another (purple) that included several wild *M. orientalis* individuals from the Hyrcanian Forests. We also detected Iranian cultivars that were highly admixed with the cultivated Iranian dark green cluster, the purple cluster, the yellow *M. domestica* genetic cluster, but also with two other clusters (red and orange), which included several wild *M. orientalis* individuals from the Kurdestan province in Iran and western Armenia, respectively (Figure 1 and S3). The four Armenian cultivars fell within the blue and orange clusters, which also included wild *M. orientalis* trees. The PCA retrieved the same clustering pattern as that inferred with STRUCTURE (Figure 1 c).

The STRUCTURE analysis therefore revealed three patterns. First, there is a specific cultivated cluster in Iran (dark green) that may result from a domestication event that may be distinct from the domestication of *M. domestica*. Second, the high level of admixture with wild *M. orientalis* in cultivated apple trees, and even sometimes their full membership to genetic clusters of wild *M. orientalis*, suggests that wild trees are grown in orchards for consumption without any strong domestication process, and/or that substantial wild-crop gene flow and/or feral individuals occur (*e.g.* the purple genetic cluster may represent a cultivated group that is also found in the wild). Third, we observed a spatial population structure of *M. orientalis* in the Caucasus that may have resulted from past range contraction and expansion associated with the last glacial maximum. We further investigated these three hypotheses (see below).

Two distinct cultivated apple groups in Iran with various levels of genetic diversity

Analyses were run without admixed individuals to better assess the genetic relationship among groups. The neighbor-net tree (Figure 1d) and the NJ tree (Figure 1e) showed that the apple cultivars (*M. domestica* and the purple and dark green Iranian cultivar clusters) grouped together and were distinct from the wild populations (except the wild Hyrcanian purple group, see below). However, the two cultivated apple populations from Iran were genetically highly differentiated (Table S5), were not sister groups (Figure 1d,e) and showed different levels of genetic diversity (Table 1). The dark green Iranian cultivated population was the most differentiated cultivated group while the purple cultivated apple population was the sister group to *M. domestica*. The two cultivated Iranian groups showed lower genetic diversity and fewer private alleles than *M. domestica* ($P < 0.03$, Table S6), the purple cultivated Iranian group displaying the lowest level of genetic diversity and the least number of private alleles. As expected from their STRUCTURE assignment, the wild and cultivated purple Iranian populations were sister groups and showed the signature of a recent bottleneck. The close relationship between trees sampled in the Hyrcanian Forests and cultivars from the purple gene pool (Figure 1d, e), and the footprint of a recent loss of genetic diversity suggest that the trees sampled in the Hyrcanian Forests assigned to the purple genetic group are feral and have originated from the purple cultivated Iranian population.

The occurrence of two distinct Iranian cultivated gene pools that are differentiated from *M. domestica* suggests that specific domestication events may have occurred in Iran.

Geographic population structure in the Caucasian crab apple

We ran analyses without admixed individuals (*i.e.*, individuals assigned with a membership coefficient < 0.9 to a given cluster, Table 1) to better assess the genetic relationships among groups. *Malus orientalis* did not form a monophyletic group. Wild *M. orientalis* populations from the Western Caucasus, and central and southeastern Armenia grouped together (Figure 1e), with the Central and Western populations as sister groups. The wild Iranian apple populations (Hyrcanian (pink), Kurdestan (red) and Lorestan (light green)) formed distinct groups that were genetically close to the wild Armenian *M. orientalis* and *M. sieversii* (Table S5, Figure 1d,e), the Hyrcanian (pink) population being the closest to the latter two. *Malus sieversii*, the Central Asian wild apple, formed a distinct group in the neighbor-net tree (Figure 1d), and was genetically close to the wild Armenian populations, as well as the Hyrcanian (pink) population, in the NJ tree (Figure 1e).

Malus orientalis in Iran has low genetic diversity and Armenia and Turkey are hotspots of genetic diversity

The level of allelic richness was significantly lower in the Lorestan (light green) and Kurdestan (red) wild populations than in the other wild populations (Tables 1 and S6), which is the signature of a recent bottleneck (Table S7). We also found that the Western (orange) population had the highest level of allelic richness (Tables 1 and S6 and Figure 2). We found a significant positive correlation between longitude and allelic richness (Figure 2 and S5, average adjusted *R-squared* = 0.66, $P < 0.0001$) and a significant negative correlation between latitude and allelic richness (Figure 2 and S5, average adjusted R-squared = -0.43, $P < 0.001$). The Western Caucasus may therefore have been a glacial refugium in the past. In addition to there being a high level 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 2) suggesting potential glacial refugia in those mountainous regions.

Range expansion and contraction associated with the last glacial maximum (LGM) of *M. orientalis* in the Caucasus

Ecological niche modeling indicated past contraction and expansion of the *M. orientalis* range. Model performance as assessed with AUC and TSS was high (Table S8), indicating that the ANN, GLM and GAM algorithms fitted the data well (Monserud & Leemans, 1992; Fieldings & Bell, 1997; Allouche *et al.*, 2006). 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 recalibrate the models to predict the past and present distribution of *M. orientalis*. The MIROC model (Figure 3) predicted that the areas suitable for *M. orientalis* during the LGM contracted to the western Lesser Caucasus and northeast 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 2). The climatic model therefore suggested that populations of the Caucasian wild apple *M. orientalis* may have been maintained in at least two glacial refugia.

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

The second STRUCTURE analysis centered on the Caucasus revealed the same genetic clustering for the wild apples and *M. domestica* at $K = 9$ (Figure S4) as in the previous analysis ($K = 11$) (Figure 1). At $K = 9$, 150 apple genotypes could be considered hybrids (*i.e.* 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). The Iranian cultivars showed the highest proportion of hybrids (67%), mostly admixed with 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 among populations.

We removed the 150 hybrids and all apple cultivars (Tables 1 and S1) 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 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.

Discussion

Our study provides insights into the natural and anthropogenic divergence history of apples in a hotspot of crop diversity, the Caucasus. We showed that apple cultivars from this region belonged to the Caucasian wild apple gene pools suggesting that local farmers used wild apple for cultivation. We also detected substantial wild-crop and crop-crop gene flow, as has been previously found in apples in Europe (Cornille *et al.*, 2012, 2014). We also identified two specific Iranian clusters of cultivated apple trees, which are differentiated from the European cultivated apple *M. domestica*, revealing a history of apple domestication that is even more complex than previously thought. The « pure » wild apple in this region, *M. orientalis*, showed a clear spatial genetic structure with seven populations spread across Turkey, Russia, Armenia and Iran, with the Iranian populations having the lowest levels of genetic diversity and strongest population structure. The combination of niche modeling and population genetics approaches suggested that these populations resulted from range contraction and expansion associated with the last glaciation. This study reveals that apples underwent a specific domestication process in the Caucasus, and pinpoints Iran as a key center in the evolution and domestication of apple. We also provided new insights into processes of natural divergence in a wild fruit in an emblematic diversity hotspot.

Iran is an additional area of apple domestication: two specific cultivated gene pools with low genetic diversity

The occurrence of two specific cultivated populations in Iran suggests that Iran is an additional center of apple domestication. The two Iranian gene pools had lower genetic diversity than *M. domestica*, but exhibited weaker signatures of bottlenecks than cereals (Glémin & Bataillon, 2009; Meyer *et al.*, 2012b; Cornille *et al.*, 2014, 2019; Gaut *et al.*, 2015). Note that we observed that *M. domestica* showed the footprint of a recent bottleneck, but the inbreeding coefficient was not significant, suggesting that this loss of genetic diversity is very recent and is likely to have resulted from recent breeding selection methods (Cornille *et al.*, 2019; Peace *et al.*, 2019).

The specific genetic diversity and differentiation of the two Iranian cultivated populations, compared with *M. domestica*, suggest a complex domestication history. There were no obvious qualitative differences in morphology or uses between the two gene pools (pers. Comm. H. Hamid). The monophyly of the two Iranian cultivated groups and the European *M. domestica* suggests that they diverged following a single domestication event in Central Asia, or represented independent domestication events from the same progenitor. We attempted to infer more precisely the domestication history using coalescent-based methods combined with approximate Bayesian computation, but these methods lacked power and we could not infer robust scenarios.

A specific apple domestication event in the Caucasus is a plausible hypothesis. Caravans carrying seeds in bags from Central Asia along the trade routes passing through Iran and Southern Caucasus may have facilitated the introduction of the cultivated apple into Europe (Gharghani *et al.*, 2010), making Iran and Southern Caucasus important hubs for the cultivated apples. The Armenian and Iranian human populations then became historically isolated from European countries with the Persian-Ottoman wars 500 years ago (Haber *et al.*, 2016), at a time when Bronze Age civilizations in the Eastern Mediterranean were collapsing with major cities destroyed or abandoned and most trade routes disrupted. This isolation have caused isolation from their surroundings, sustained by the cultural/linguistic/religious distinctiveness that persists to this day (Haber *et al.*, 2016). The isolation of Armenia and Iran, after the considerable commercial exchanges associated with the trade routes passing through northern Iran, may have led to a distinct domestication event in this region and thus to specific cultivated apple gene pools sheltered from the massive introgression of *M. sylvestris* into the cultivated gene pool in Western Europe (Cornille *et al.*, 2012). This isolation is still maintained as only a handful of international *M. domestica* cultivars are currently grown in the Caucasus. This is not surprising as local apple cultivar varieties from the Caucasus display interesting agronomic traits such as fruit size and resistance to several diseases and drought (Volk *et al.*, 2008; Höfer *et al.*, 2013; Amirchakhmaghi *et al.*, 2018). The finding that one of the two specific Iranian cultivated apple gene pools is closer to *M. domestica* than the other suggests there may have been two distinct domestication events in Iran, perhaps at different times or with different secondary progenitors.

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

We found evidence for substantial wild-crop and crop-crop gene flow in the Caucasus. Indeed, we found a substantial number of Iranian cultivars that were introgressed by local wild apple gene pools or were an admixture of two cultivated gene pools. Reciprocally, many wild Iranian and Armenian individuals were introgressed by local cultivated gene pools. This extensive wild-crop and crop-crop gene flow is strikingly similar to the pattern documented in apples in Europe. Substantial crop-to-wild gene flow has been reported from *M. domestica* to *M. sylvestris* (Cornille *et al.*, 2015). Reciprocally, *M. sylvestris* has been shown to be a significant contributor to the *M. domestica* gene pool through recurrent and recent hybridization and introgression events starting when the cultivated apple was introduced in Europe by the Greeks around 1,500 years ago (Cornille *et al.*, 2012). Similarly, extensive gene flow was found during the domestication of other fruit trees (Arroyo-García *et al.*, 2006; Myles *et al.*, 2011; Cornille *et al.*, 2012; Meyer *et al.*, 2012a; Diez *et al.*, 2015; Decroocq *et al.*, 2016; Duan *et al.*, 2017; Liu *et al.*, 2019).

Despite the spread of the cultivated apple *M. domestica* along the Silk Trade Routes that crossed Iran as well as South Caucasus to reach Turkey (Canepa, 2010; Spengler, 2019), it seems that local farmers often used the local wild apples rather than *M. domestica*. Indeed, Armenian cultivars shared their gene pools with the Western and Central Caucasian wild apple populations, and Iranian cultivars were highly admixed with two of the four wild apple populations identified in Iran. This is not surprising as this wild species can grow in mountainous areas, is highly resistant to pests and diseases (Büttner, 2001) and has high-quality fruit that are intermediate between *M. sylvestris* and *M. sieversii* apples (Cornille *et al.*, 2014). The use of the local wild apple has also been documented in Europe for specific purposes at different times in history (Tardío *et al.*, 2020).

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 found in Turkey, Russia and northwestern Armenia, two in Armenia (a southern and a northern population) and four in Iran, including one in the Kurdestan province, one in 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 Armenia, Turkey and Russia, and of several local hotpots of genetic diversity in the Lesser Caucasus Mountains in Armenia and in the Hyrcanian Forests. Ecological niche modeling further supported the existence of strong contractions in the range of *M. orientalis* in the Turkey and the Lesser Caucasus, as well as in some local parts of the Hyrcanian Forests. Ecological niche modeling also suggested that the Caucasian wild apple restricted its range to the Colchis region and the Higher Caucasus Mountains, bordering the Black Sea, and the Eastern Caucasus close to Azerbaijan, but we lacked samples from these regions to further confirm this hypothesis. These glacial refugia have been described in relation to other species (Parvizi *et al.*, 2019). Indeed, in the Caucasus, two refugia are recognized (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 northeast 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 extreme West and East Caucasus is now 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.*, 2013a). 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 (Vekemans & Hardy, 2004b; Cornille *et al.*, 2013a,b). 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. The wild apples can thus spread over kilometers (Cornille *et al.*, 2015; Feurtey *et al.*, 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, which had the signature of a very recent and strong bottleneck. 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 *devenir* 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

In this study, we revealed the existence of an additional center of apple domestication, which seems to have followed strikingly similar processes as those observed in Europe, *i.e.* substantial wild-crop and crop-crop gene flow. We also found that wild apples are used for cultivation in the Caucasus. *Malus orientalis* therefore appears to be an additional contributor to the history of apple domestication in the Caucasus. The specificity of the Iranian cultivated apple gene pools indicate a more complex domestication history than previously thought, with specific domestication events in this region. Our landscape genetics study also provides insights into the processes of natural divergence of this emblematic wild fruit and contributor to apple domestication in this region. However, further investigations and sampling of the huge diversity of *M. orientalis* in the Caucasian ecoregion are now necessary. A better understanding of the properties of functional genetic diversity and ecological relationships of wild apples in their ecosystem is also needed for developing and implementing effective conservation genetic strategies in this region (Teixeira & Huber, 2021).

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 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 modelling 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, FS, HA, IG, AN, AC sampled the material; AV, CR, AR performed the molecular work; AC, HB analyzed the data. The manuscript was written by AC and HB with critical inputs from other co-authors.

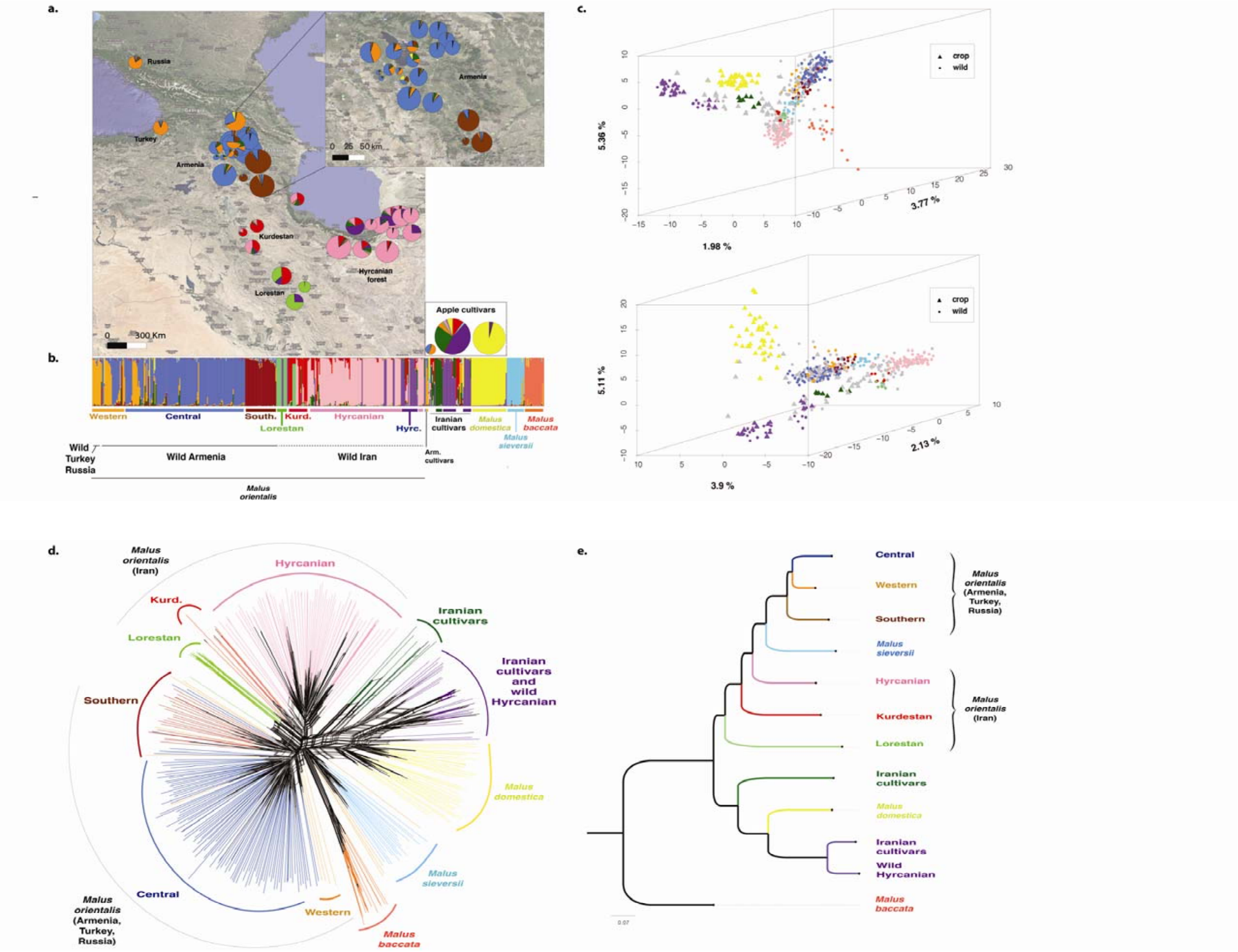

a.
Russia
Turkey
Armenia
Armenia
0 25 50 km
Kurdestan
Hyrcanian forest
Lorestan
0 300 Km
Apple cultivars
b.
Western
Central
South.
Kurd.
Lorestan
Hyrcanian
Hyrc.
Iranian cultivars
Malus domestica
Malus sieversii
Malus baccata
Wild Turkey Russia
Wild Armenia
Wild Iran
Arm. cultivars
Malus orientalis
c.
crop
wild
5.36 %
1.98 %
3.77 %
5.11 %
3.9 %
2.13 %
d.
Malus orientalis (Iran)
Hyrcanian
Kurd.
Lorestan
Southern
Iranian cultivars
Iranian cultivars and wild Hyrcanian
Malus domestica
Malus sieversii
Malus baccata
Western
Central
Malus orientalis (Armenia, Turkey, Russia)
e.
Central
Western
Southern
Malus orientalis (Armenia, Turkey, Russia)
Malus sieversii
Hyrcanian
Kurdestan
Lorestan
Malus orientalis (Iran)
Iranian cultivars
Malus domestica
Iranian cultivars
Wild Hyrcanian
Malus baccata
0.07

Figure 1. Population genetic structure and differentiation between cultivated and wild apple from the Caucasus, *Malus domestica*, *Malus sieversii* and *Malus baccata* (based on 26 microsatellite markers). a. Spatial population genetic structure inferred with STRUCTURE at $K = 11$ ($N = 508$); 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*, $N = 40$). Pie chart size is proportional to the number of samples per site. **b.** STRUCTURE bar plot ($N = 508$) at $K = 11$ showing eleven 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 apple cultivars are grouped according to their origin: Armenia ($N = 4$), Iran ($N = 48$) and *M. domestica* ($N = 40$), *M. sieversii* and *M. baccata*. c. Principal component analysis (PCA) of 508 individuals (above), and after removing the outgroup *M. baccata* (below, $N = 488$), with the respective total variation explained by each component. Colors correspond to the genetic groups inferred at $K = 11$; crop samples are shown as triangles and wild samples as circles. d. Neighbor-net representing the genetic relationships among the wild and cultivated apple populations (*i.e.* genetic groups comprising individuals with a membership coefficient > 0.9 to a given cluster) inferred with STRUCTURE at $K = 11$. f. Neighbor-joining tree representing the distance among the eleven populations inferred with STRUCTURE at $K = 11$ and rooted with *M. baccata*. For e. and f. each branch is coloured according to the population inferred with STRUCTURE at $K = 11$.

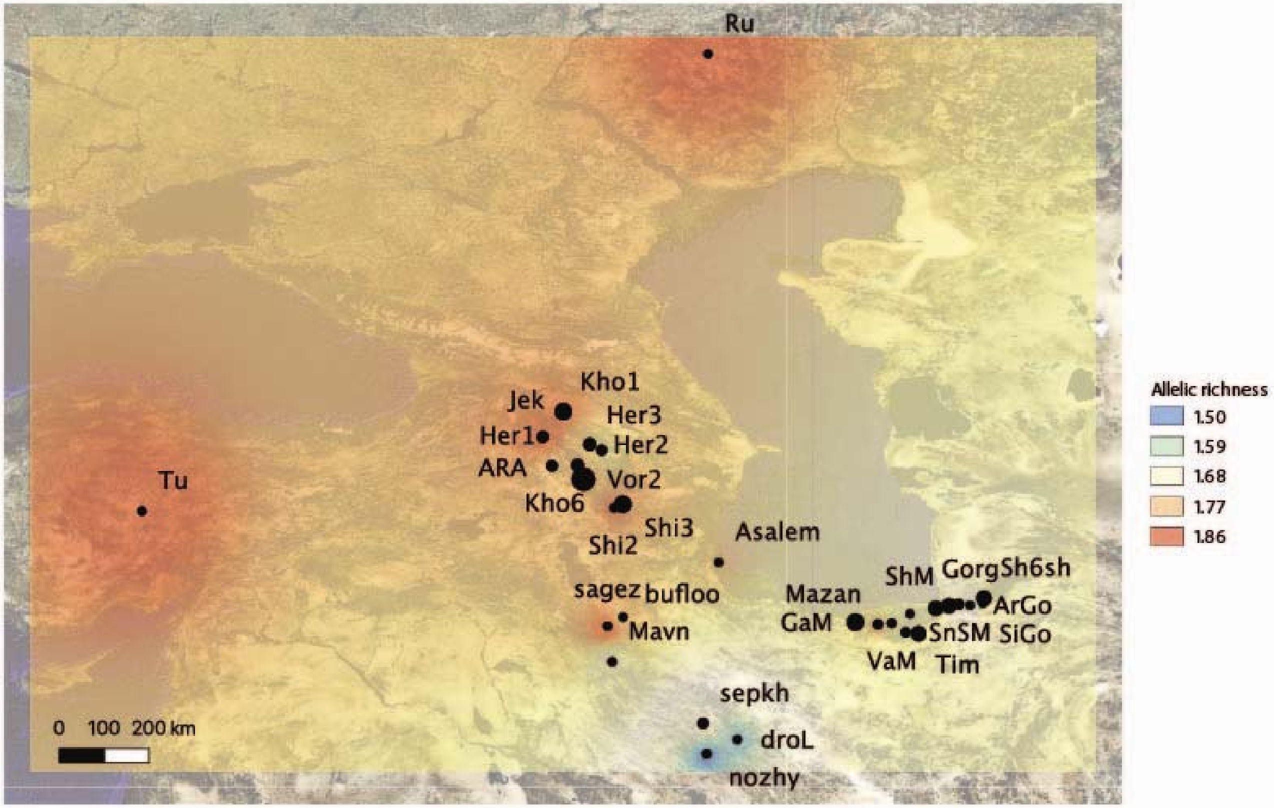

Ru
Tu
Kho1
Jek
Her3
Her1
Her2
ARA
Vor2
Kho6
Shi3
Shi2
Asalem
sagez
bufloo
Mavn
Mazan
GaM
ShM
GorgSh6sh
ArGo
SnSM
SiGo
VaM
Tim
sepkh
droL
nozhy
0 100 200 km
Allelic richness
1.50
1.59
1.68
1.77
1.86

Figure 2. Spatial genetic diversity (allelic richness) in *Malus orientalis* across the Caucasus (36 sites).

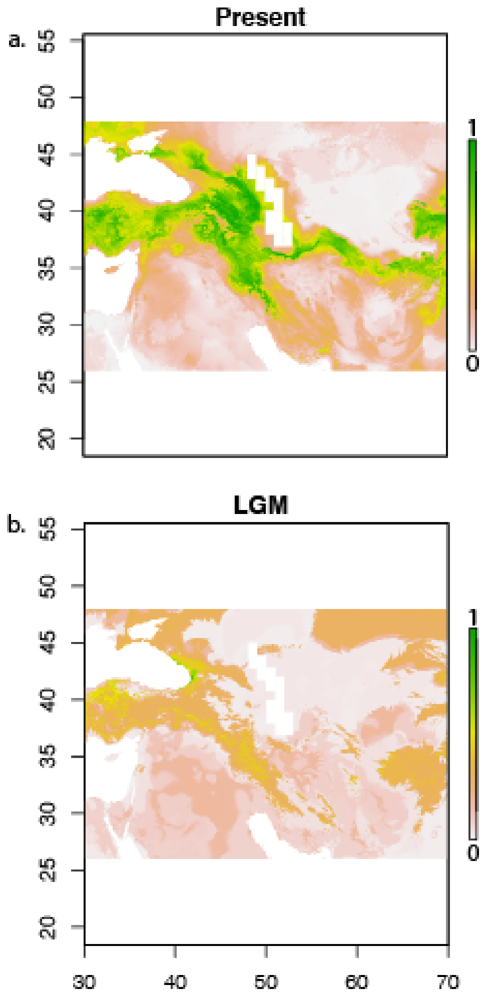

Present
a.
55
50
45
40
35
30
25
20
1
0
LGM
b.
55
50
45
40
35
30
25
20
1
0
30
40
50
60
70

Figure 3. 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 *Malus orientalis*. The probabilities of being a suitable habitat are given in the legend.

Table 1. Genetic diversity estimates for wild and cultivated apple populations (*i.e.* individuals with a membership coefficient < 0.9 to any given cluster were excluded from the analysis, N = 345) **detected with STRUCTURE at K = 11.** Note that the purple cluster was split between cultivated and wild samples. Thus samples were partitioned into 12 populations, including nine wild and three cultivated apple populations.

Wild or cultivated	Species	Country of origin	Population	N	H_O	H_E	F_{IS}	Bottleneck (P-value)	A_R (G=6)	A_P (G=6)
wild	*Malus orientalis*	Armenia	Western (orange)	17	0.83	0.87	**0.05****	0.85 ^{NS}	4.55±0.11	1.03±0.11
			Central (blue)	97	0.78	0.79	0.01 ^{NS}	0.91 ^{NS}	3.86±0.00	0.63±0.09
			Southern (brown)	25	0.77	0.80	0.03 ^{NS}	0.06 ^{NS}	3.88±0.14	0.64±0.12
		Iran	Lorestan (light green)	10	0.85	0.46	**-0.83 *****	**0**	2.02±0.09	0.23±0.06
			Kurdestan (red)	7	0.78	0.72	**-0.08 ***	**0.01**	3.38±0.14	0.61±0.10
			Hyrcanian (pink)	77	0.72	0.77	**0.07 *****	0.94 ^{NS}	3.77±0.15	0.73±0.11
			Hyrcanian (purple)	19	0.72	0.60	**-0.20 *****	**0.01**	2.90±0.12	0.22±0.03
	Malus sieversii	Kazahkst	(light blue)	16	0.70	0.73	**0.05 ***	**0**	3.40±0.12	0.40±0.0

		an								7
	Malus baccata	Russia	(light red)	15	0.50	0.53	0.06 ^{NS}	0.90 ^{NS}	2.62±0.19	1.14±0.15
cultivated	*Malus domestica*	Mostly Europe	European cultivars	38	0.79	0.78	-0.01 ^{NS}	**0.01**	3.77±0.09	0.88±0.10
	?	Iran	**Iranian cultivars (purple)**	17	0.84	0.58	**-0.47** ***	0.15 ^{NS}	2.71±0.11	0.11±0.02
			Iranian cultivars (dark green)	7	0.71	0.65	**-0.09** *	0.77 ^{NS}	3.06±0.13	0.55±0.09
			TOTAL	**345**						

N: number of individuals assigned to a focal cluster with a membership coefficient > 0.90; 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 6; A_P: number of private alleles, corrected by the rarefaction method, estimated for a sample size of 6; *P-value* two-phase mutation model: *P-values* were obtained with the one tail test for heterozygosity excess implemented in BOTTLENECK from the two-phased mutation model; *: 0.05<*P*<0.01; ***: *P*<0.001 ; NS : non-significant.

Table 2. Distribution of hybrids (i.e., individual with a membership coefficient < 0.90 to any given genetic cluster, as inferred for K = 9 with STRUCTURE) for the 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.

Genepool		Cultivated apple			Wild apple			
		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		**0.13**	0.66	0.04	0.02		
	N (% over the total)		**11 (23%)**	2 (5%)	4 (2%)	5 (2.9%)		
Wild and cultivated Hyrcanian (purple)	mean introgression rate		**0.36**	0.33		**0.16**		
	N (%)		**18 (37.5%)**	2 (5%)		**13 (7.8%)**		
Cult. Iran (dark_green)	mean introgression rate		**0.17**		0.05	**0.06**		
	N (%)		**14 (30%)**		11 (5.6%)	**8 (4.7%)**		
Hyrcanian (pink)	mean introgression rate		**0.05**			**0.41**		
	N (%)		**6 (12.5%)**			**34 (20.3%)**		
Lorestan (light green)	mean introgression rate		0.01			0.04		
	N (%)		2 (4%)			6		
Wild and cultivated Kurdestan (red)	mean introgression rate		**0.17**			**0.27**		
	N (%)		**12 (35%)**			**30 (17.9%)**		
Western (orange)	mean introgression rate	0.65	0.08		**0.28**	0.02	0.49	0.84
	N (%)	1 (33%)	3 (6.2%)		**33 (16.8%)**	4 (2.3%)	1 (20%)	2 (33%)

Southern (brown)	mean introgression rate		0.02		**0.17**	0.02	0.35			
	N (%)		2 (4.1%)		**13 (6.6%)**	2 (1.1%)	1 (20%)			
Central (blue)	mean introgression rate	0.32	0.02		**0.46**	0.01		0.11		
	N (%)	1 (33%)	1 (2%)		**46 (23.4%)**	1 (0.5%)		1 (16.6%)		
	Total number of hybrids (and %)	**1**	**32 (66%)**	**2 (5%)**	**58 (28%)**	**54 (32%)**	**1 (20%)**	**2 (33%)**	**150 (32%**	

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**: 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**: 1255–1268.

Arroyo-García R, Ruiz-García L, Bolling L, Ocete R, López MA, Arnold C, Ergul A, Söylemezoğlu G, Uzun HI, Cabello F, *et al.* **2006**. Multiple origins of cultivated grapevine (Vitis vinifera L. ssp. sativa) based on chloroplast DNA polymorphisms. *Molecular Ecology* **15**: 3707–3714.

Asanidze Z, Akhalkatsi M, Henk AD, Richards CM, Volk GM. **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**: 504–512.

Asanidze Z, Akhalkatsi M, Henk AD, Richards CM, Volk GM. **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**: 504–512.

Bacles CFE, Jump AS. **2011**. Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science* **16**: 13–18.

Bai XN, Spitkovsky A. **2010**. Uncertainties of modeling gamma-ray pulsar light curves using vacuum dipole magnetic field. *Astrophysical Journal* **715**: 1270–1281.

Bailey-Serres J, Parker JE, Ainsworth EA, Oldroyd GED, Schroeder JI. **2019**. Genetic strategies for improving crop yields. *Nature* **575**: 109–118.

Barbet-Massin M, Jiguet F, Albert CH, Thuiller W. **2012**. Selecting pseudo-absence for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **in press**.

Besnard G, Terral JF, Cornille A. **2018**. On the origins and domestication of the olive: A review and perspectives. *Annals of Botany* **121**: 385–403.

Bina H, Yousefzadeh H, Ali SS, Esmailpour M. **2016**. Phylogenetic relationships, molecular taxonomy, biogeography of Betula, with emphasis on phylogenetic position of Iranian populations. *Tree Genetics and Genomes* **12**.

Brandenburg JT, Mary-Huard T, Rigaill G, Hearne SJ, Corti H, Joets J, Vitte C, Charcosset A, Nicolas SD, Tenaillon MI. **2017**. Independent introductions and admixtures have contributed to adaptation of European maize and its American counterparts. *PLoS Genetics* **13**: 1–30.

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.

Castañeda-Álvarez NP, Khoury CK, Achicanoy HA, Bernau V, Dempewolf H, Eastwood RJ, Guarino L, Harker RH, Jarvis A, Maxted N, ***et al.*** **2016**. Global conservation priorities for crop wild relatives. *Nature Plants* **2**: 16022.

Chen J, Li L, Milesi P, Jansson G, Berlin M, Karlsson B, Aleksic J, Vendramin GG, 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**: 1539–1551.

Cornille A, Antolín F, Garcia E, Vernesi C, Fietta A, Brinkkemper O, Kirleis W, Schlumbaum A, Roldán-Ruiz I. **2019**. A Multifaceted Overview of Apple Tree Domestication. *Trends in Plant Science* **24**: 770–782.

Cornille A, 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**: 373–384.

Cornille A, Giraud T, Bellard C, Tellier A, Le Cam B, Smulders MJM, Kleinschmit J, Roldan-Ruiz I, Gladieux P. **2013a**. Post-glacial recolonization history of the European crabapple (Malus sylvestris Mill.), a wild contributor to the domesticated apple. *Molecular Ecology* **22**: 2249–63.

Cornille A, Giraud T, Smulders MJM, Roldán-Ruiz I, Gladieux P. **2014**. The domestication and evolutionary ecology of apples. *Trends in Genetics* **30**: 57–65.

Cornille A, 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**: 737–748.

Cornille A, Gladieux P, Giraud T. **2013c**. Crop-to-wild gene flow and spatial genetic structure in the closest wild relatives of the cultivated apple. *Evolutionary Applications* **6**: 737–748.

Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, ***et al.*** **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**.

Decroocq S, Cornille A, Tricon D, Babayeva S, Chague A, Eyquard JP, Karychev R, Dolgikh S, Kostritsyna T, Liu S, ***et al.*** **2016**. New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. *Molecular ecology* **25**: 4712–4729.

Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS. **2015**. Olive domestication and diversification in the Mediterranean Basin. *New Phytologist* **206**: 436–447.

Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, Wang X, Jiao C, Legall N, Mao L, *et al.* **2017**. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nature Communications* **8**.

Earl DA. **2012**. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics resources* **4**: 359–361.

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

Excoffier L, Foll M, Petit RJ. **2009**. Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics* **40**: 481–501.

Feurtey A, Cornille A, Shykoff JA, Snirc A, Giraud T. **2017**. Crop□to□wild gene flow and its fitness consequences for a wild fruit tree: Towards a comprehensive conservation strategy of the wild apple in Europe. *Evolutionary Applications* **10**: 180–188.

Fieldings AH, Bell JF. **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**: 157–167.

Flowers JM, Hazzouri KM, Gros-Balthazard M, Mo Z, Koutroumpa K, Perrakis A, Ferrand S, Khierallah HSM, Fuller DQ, Aberlenc F, *et al.* **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**: 1651–1658.

Forsline PL, Aldwinckle HS, Dickson EE, Luby JJ, Hokanson SC. **2003**. *of Wild Apples of Central Asia*.

Fuller DQ. **2018**. Long and attenuated: comparative trends in the domestication of tree fruits. *Vegetation History and Archaeobotany* **27**: 165–176.

Gabrielian ET, Zohary D. **2004**. Wild relatives of food crops native to Armenia and Nakhichevan. *Flora Mediterranea* **14**: 5–80.

Gaut BS, Díez CM, Morrell PL. **2015**. Genomics and the Contrasting Dynamics of Annual and Perennial Domestication. *Trends in Genetics* **31**: 709–719.

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

Gharghani A, Zamani Z, Talaie A, Oraguzie NC, Fatahi R, Hajnajari H, Wiedow C, Gardiner SE. **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**: 829–842.

Giesecke T, Brewer S, Finsinger W, Leydet M, Bradshaw RHW. **2017**. Patterns and dynamics of European vegetation change over the last 15,000 years. *Journal of Biogeography* **44**: 1441–1456.

Glémin S, Bataillon T. **2009**. A comparative view of the evolution of grasses under domestication. *New Phytol.* **183**: 273–290.

Haber M, Mezzavilla M, Xue Y, Comas D, Gasparini P, Zalloua P, Tyler-Smith C. **2016**. Genetic evidence for an origin of the Armenians from Bronze Age mixing of multiple populations. *European Journal of Human Genetics* **24**: 931–936.

Hardy OJ, Vekemans X. **2002**. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular ecology notes* **2**: 618–620.

Hewitt GM. **1990**. Divergence and speciation as viewed from an insect hybrid zone. *Canadian Journal of Zoology* **68**: 1701–1715.

Hewitt GM. **1996**. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**: 247–276.

Hewitt GM. **2004**. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**: 183–195.

Höfer M, Flachowsky H, Hanke M-V, Semënov V, Šlâvas A, Bandurko I, Sorokin A, Alexanian S. **2013**. Assessment of phenotypic variation of Malus orientalis in the North Caucasus region. *Genetic Resources and Crop Evolution* **60**: 1463–1477.

Huson DH. **1998**. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England)* **14**: 68–73.

Huson DH, Scornavacca C. **2012**. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic biology* **61**: 1061–1067.

Jezkova T, Olah-Hemmings V, Riddle BR. **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**: 3486–3502.

Jin L, Chakraborty R. **1994**. Estimation of genetic distance and coefficient of gene diversity from single-probe multilocus DNA fingerprinting data. *Molecular biology and evolution* **11**: 120–127.

Jombart T, Ahmed I. **2011**. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**: 3070–3071.

Kalinowski ST. **2011**. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity* **106**: 625–632.

Langenfeld WT. **1991**. Apple trees. *Morphological evolution, phylogeny, geography and systematics. Riga (Zinatne)* **232**.

Lascoux M, Palmé AE, Cheddadi R, Latta RG. **2004**. Impact of Ice Ages on the genetic structure of trees and shrubs. *Philosophical Transactions of the Royal Society B: Biological Sciences* **359**: 197–207.

Leroy B, Bellard C, Dubos N, Colliot A, Vasseur M, Courtial C, Bakkenes M, Canard A, Ysnel F. **2014**. Forecasted climate and land use changes, and protected areas: the contrasting case of spiders. *Diversity and Distributions* **20**: 686–697.

Liang Z, Duan S, Sheng J, Zhu S, Ni X, Shao J, Liu C, Nick P, Du F, Fan P, *et al.* 2019. Whole-genome resequencing of 472 Vitis accessions for grapevine diversity and demographic history analyses. *Nature Communications* **10**: 1190.

Liu S, Cornille A, Decroocq S, Tricon D, Chague A, Eyquard J, Liu W, Giraud T, Decroocq V. **2019**. *The complex evolutionary history of apricots: species divergence, gene flow and multiple domestication events*.

Loiselle BA, Sork VL, Nason J, Graham C. **1995**. Spatial genetic structure of a tropical understory shrub, Psychotria officinalis (Rubiaceae). *American journal of botany* **82**: 1420–1425.

Meyer RS, Duval AE, Jensen HR. **2012a**. Patterns and processes in crop domestication: An historical review and quantitative analysis of 203 global food crops. *New Phytologist* **196**: 29–48.

Meyer RS, DuVal AE, Jensen HR. **2012b**. Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytologist* **196**: 29–48.

Monserud RA, Leemans R. **1992**. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling* **62**: 275–293.

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, *et al.* 2011. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 3530–3535.

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.

Oddou-Muratorio S, Klein EK. **2008**. Comparing direct vs. indirect estimates of gene flow within a population of a scattered tree species. *Molecular Ecology* **17**: 2743–2754.

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

Patocchi A, Frei A, Frey JE, 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**: 337–347.

Peace CP, Bianco L, Troggio M, van de Weg E, Howard NP, Cornille A, Durel C-E, Myles S, Migicovsky Z, Schaffer RJ, *et al.* **2019**. Apple whole genome sequences: recent advances and new prospects. *Horticulture Research* **6**: 59.

Petit RJ, 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**: 117–137.

Petit RJ, Hampe A. **2006**. Some Evolutionary Consequences of Being a Tree. *Annual Review of Ecology, Evolution, and Systematics* **37**: 187–214.

Piry S, Luikart G, Cornuet JM. **1999**. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of heredity* **90**: 502–503.

Pritchard JK, Stephens M, Donnelly P. **2000**. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Puechmaille SJ. **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**: 608–627.

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**: 6489–6491.

Raymond M, Rousset F. **1995**. An exact test for population differentiation. *Evolution* **49**: 1280–1283.

Rechinger KH. **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**: 103–106.

Schmitt T. **2007**. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* **4**: 1–13.

Soofi M, Ghoddousi A, Zeppenfeld T, Shokri S, Soufi M, Jafari A, Ahmadpour M, Qashqaei AT, Egli L, Ghadirian T, *et al.* **2018**. Livestock grazing in protected areas and its effects on large mammals in the Hyrcanian forest, Iran. *Biological Conservation* **217**: 377–382.

Spengler RN. **2019**. *Fruit from the Sands: The Silk Road Origins of the Foods We Eat*. University of California Press.

Szpiech ZA, Jakobsson M, Rosenberg NA. **2008**. ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**: 2498–2504.

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*.

Teixeira JC, Huber CD. **2021**. The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences* **118**: e2015096118.

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**: 1289–1295.

Vavilov NI. **1992**. *Origin and geography of cultivated plants*. Cambridge: Cambridge University Press.

Vekemans X, Hardy OJ. **2004a**. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**: 921–935.

Vekemans X, Hardy OJ. **2004b**. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* **13**: 921–935.

Volk GM, Cornille A. **2019**. Genetic Diversity and Domestication History in Pyrus. In: The Pear Genome. Springer, 51–62.

Volk GM, Richards CM, Reilley AA, Henk AD, Reeves PA, Forsline PL, Aldwinckle HS. **2008**. Genetic diversity and disease resistance of wild Malus orientalis from Turkey and Southern Russia. *Journal of the American Society for Horticultural Science* **133**: 383–389.

Vouillamoz JF, McGovern PE, Ergul A, Söylemezoğlu G, Tevzadze G, Meredith CP, Grando MS. **2006**. Genetic characterization and relationships of traditional grape cultivars from Transcaucasia and Anatolia. *Plant Genetic Resources* **4**: 144–158.

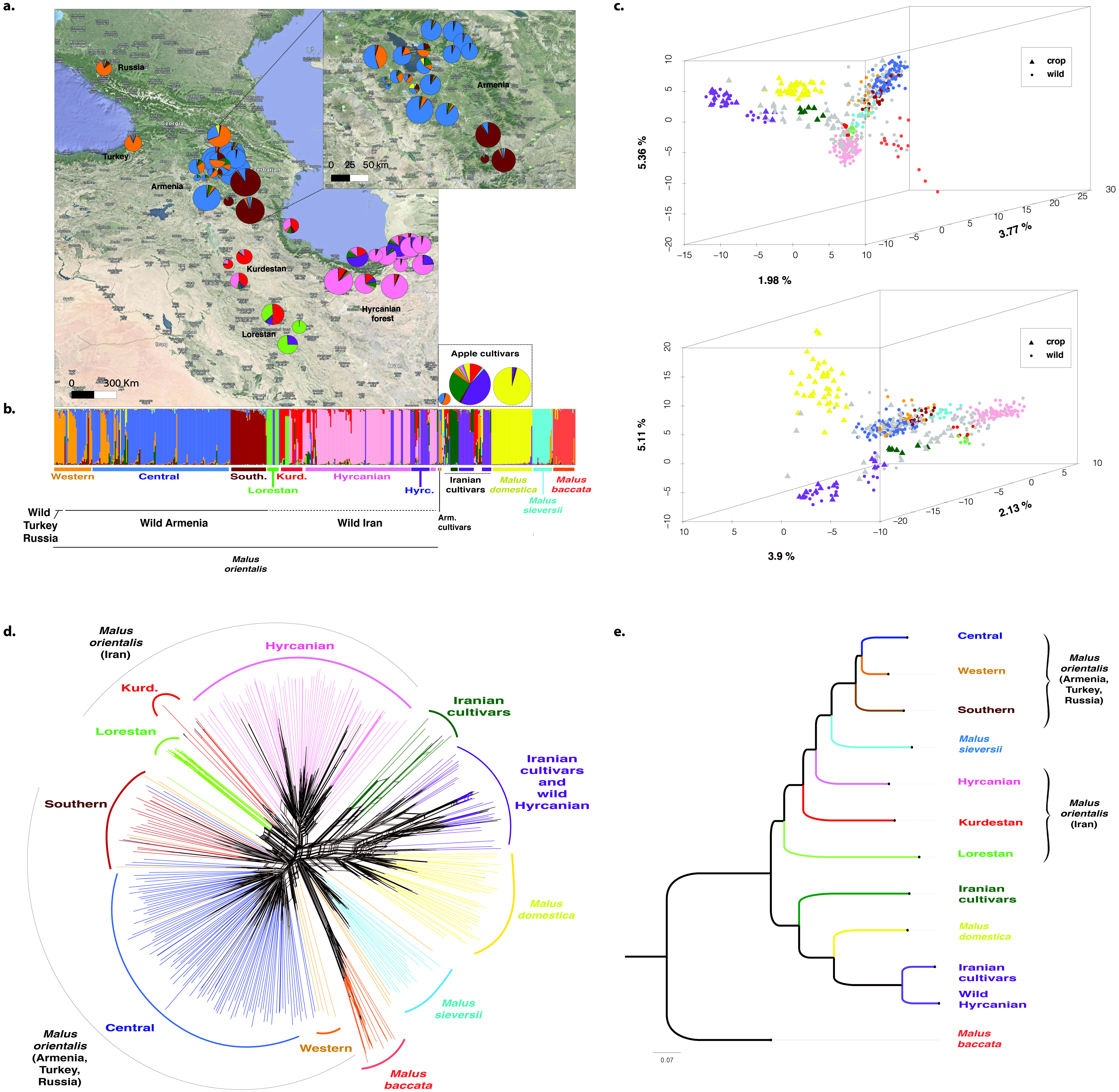
Yousefzadeh H, Hosseinzadeh Colagar A, Tabari M, Sattarian A, Assadi M. **2012**. Utility of ITS region sequence and structure for molecular identification of Tilia species from Hyrcanian forests, Iran. *Plant Systematics and Evolution* **298**: 947–961.

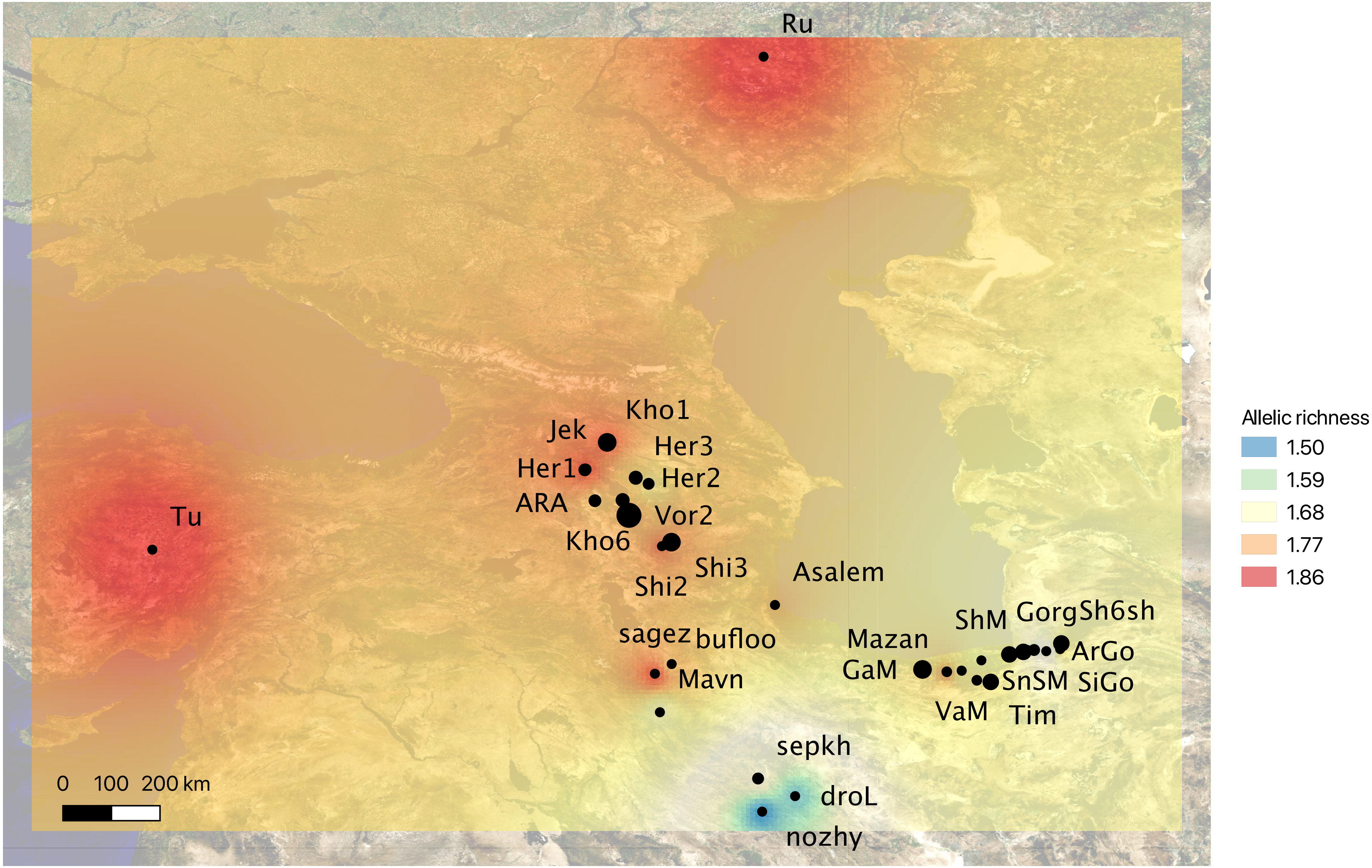
Zazanashvili N, Sanadiradze G, Garforth M, Bitsadze M, Manvelyan K, Askerov E, Mousavi M, Krever V, Shmunk V, Kalem S, *et al.* 2020. Ecoregional Conservation Plan (ECP) For The Caucasus 2020 Edition.

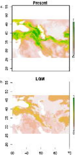
Zeng G, Zhang J, Chen Y, Yu Z, Yu M, Li H, Liu Z, Chen M, Lu L, Hu C. **2011**. Relative contributions of archaea and bacteria to microbial ammonia oxidation differ under different conditions during agricultural waste composting. *Bioresource Technology* **102**: 9026–9032.

Zhang H, Mittal N, Leamy LJ, Barazani O, Song BH. **2017**. Back into the wild—Apply untapped genetic diversity of wild relatives for crop improvement. *Evolutionary Applications* **10**: 5–24.

Zohary D, Spiegel-Roy P. **1975**. Beginnings of fruit growing in the Old World. *Science*: 319–327.


a.
Russia
Georgia
Turkey
Armenia
Azerbaijan
Kurdestan
Hyrcanian forest
Lorestan
Iraq
Syria
Iran
0 300 Km
0 25 50 km
Armenia
Apple cultivars
b.
Western
Central
South.
Lorestan
Kurd.
Hyrcanian
Hyrc.
Iranian cultivars
Malus domestica
Malus sieversii
Malus baccata
Wild Turkey Russia
Wild Armenia
Wild Iran
Arm. cultivars
Malus orientalis
c.
crop
wild
5.36 %
1.98 %
3.77 %
5.11 %
3.9 %
2.13 %
d.
Malus orientalis (Iran)
Hyrcanian
Kurd.
Lorestan
Iranian cultivars
Iranian cultivars and wild Hyrcanian
Southern
Malus domestica
Malus sieversii
Central
Western
Malus baccata
Malus orientalis (Armenia, Turkey, Russia)
e.
Central
Western
Southern
Malus orientalis (Armenia, Turkey, Russia)
Malus sieversii
Hyrcanian
Kurdestan
Lorestan
Malus orientalis (Iran)
Iranian cultivars
Malus domestica
Iranian cultivars
Wild Hyrcanian
Malus baccata
0.07


Ru
Kho1
Jek
Her3
Her1
Her2
ARA
Vor2
Tu
Kho6
Shi3
Shi2
Asalem
ShM
Gorg
Sh6sh
sagez
bufloo
Mazan
ArGo
GaM
SnSM
SiGo
Mavn
VaM
Tim
sepkh
droL
nozhy
0
100
200 km
Allelic richness
1.50
1.59
1.68
1.77
1.86


Present
LGM