

ORIGINAL ARTICLE OPEN ACCESS

Insights Into the Almond Domestication History

Stephane Decroocq¹ | Amandine Cornille^{2,3} | Naïma Dlalal⁴ | Henri Duval⁴ | David Tricon⁴ | Benedicte Quilot⁴ | Wisam K. Khalid^{1,5} | Aurélie Chague¹ | Iban Eduardo⁶ | Ignasi Batlle⁷ | Pavlina Drogoudi⁸ | Ayzin Küden⁹ | Bayram M. Asma¹⁰ | Tatiana Kostitsyna¹¹ | Véronique Decroocq¹

¹INRAE, Biologie du Fruit et Pathologie, UMR 1332, PrADAm, Université de Bordeaux, Villenave d'Ornon, France | ²INRAE, CNRS, AgroParisTech, GQE - IDEEV, Université Paris Saclay, Gif-sur-Yvette, France | ³Division of Science, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE | ⁴GAFL, INRAE, Montfavet, France | ⁵Department of Horticulture & Landscape Design, College of Agriculture & Forestry, University of Mosul, Mosul, Iraq | ⁶Centre de Recerca en Agrigenòmica (CRAG), CSIC-IRTA-UAB-UB, Cerdanyola del Vallès (Bellaterra), Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Barcelona, Spain | ⁷Mas Bové, Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Constantí, Tarragona, Spain | ⁸Department of Deciduous Fruit Trees, Institute of Plant Breeding and Genetic Resources, Hellenic Agricultural Organization (ELGO)—'Dimitra', Naoussa, Greece | ⁹Department of Horticulture, Faculty of Agriculture, Cukurova University (CU), Adana, Turkey | ¹⁰Department of Horticulture, Malatya Turgut Ozal, Inonu University, Malatya, Turkey | ¹¹International Higher School of Medicine, Bishkek, Kyrgyz Republic

Correspondence: Véronique Decroocq (veronique.decroocq@inrae.fr)

Received: 2 November 2024 | **Revised:** 25 July 2025 | **Accepted:** 15 August 2025

Funding: We acknowledge funding from the FP7 MCSA STONE, PRIMA (FREECLIMB ANR-18-PRIM-000, 10.13039/100014439), the Grant PCI2019-103670 funded by MCIN/AEI/10.13039/501100011033 and co-financed by the European Union, and the current support from Horizon Europe (FRUITDIV 101133964) project. European FP7 IFPT-PEOPLE-2009-PIRSES STONE (246795).

Keywords: almonds | domestication | fruit trees | gene flow | genetic resources | Mediterranean Basin | *Prunus* | sharka | virus

ABSTRACT

Understanding crop domestication offers crucial insights into the evolutionary processes that drive population divergence and adaptation. It also informs the identification of genetically diverse wild germplasm, which is essential for breeding and conservation efforts. While domestication has been extensively studied in many Mediterranean fruit trees, the evolutionary history of the almond (*Prunus dulcis*) remains comparatively underexplored. To address this, we analyzed 209 wild and cultivated almond accessions sampled across Eurasia and genotyped with 23 microsatellite markers. Using population genetics and coalescent-based inference, we reconstructed the domestication history of *P. dulcis* and its relationships with wild relatives. Bayesian clustering revealed four genetically distinct clusters of cultivated almonds: Turkish, Caucasian–Central Asian, Southern Spanish, and European/North American. These groups were differentiated from wild almond species—including *Prunus turcomanica*, *Prunus orientalis*, *Prunus fenzliana*, and *Prunus spinosissima*—each forming its gene pool across the Middle East and Central Asia. Approximate Bayesian Computation (ABC) supported a single domestication event in the Middle East, originating from either *P. orientalis* or *P. turcomanica*, with subsequent gene flow from *P. fenzliana* and *P. spinosissima* into the Turkish and Central Asian cultivated gene pools, respectively. We also inferred reciprocal introgression from cultivated almonds back into wild populations. Notably, sharka resistance—caused by plum pox virus (PPV)—was identified in three *P. dulcis* clusters and *P. fenzliana*, suggesting that resistance may have arisen independently or been maintained through crop–wild introgression. Together, our results highlight a complex and protracted domestication history for almond, shaped by contributions from multiple wild relatives and recurrent gene flow. These findings enhance our understanding of perennial crop evolution and underscore the value of wild germplasm in breeding programs aimed at increasing resilience in fruit trees.

Co-first authors: Stephane Decroocq, Amandine Cornille, and Naïma Dlalal contributed equally to this work.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Evolutionary Applications* published by John Wiley & Sons Ltd.

1 | Introduction

The study of crop domestication is crucial to unravel the evolutionary processes that shape population divergence (Gaut et al. 2015; Glémin and Bataillon 2009; Meyer et al. 2012) and to design future redomestication strategies (Gradziel 2020) in a context of rare archaeobotanical remains. Documentation on the domestication of fruit tree crops still lags behind that of annual crop plants (Burgarella et al. 2019; Gaut et al. 2015; Meyer et al. 2012; Miller and Gross 2011; Neale et al. 2017). Differences in life-history traits have likely led to significant differences in the mode and speed of evolution between trees and annuals (Gaut et al. 2015), particularly in the extent of gene flow during domestication and the absence of a bottleneck. Improving knowledge on the domestication of fruit tree crops is important for almond breeding and conservation of genetic resources. Due to long-term cultivation, fruit trees are increasingly susceptible to the ever-changing environments caused by climate change and pest outbreaks. In this context, resilient fruit production is believed to rely heavily on untapped wild genetic diversity (Fangning and Jacqueline 2020; Zhang et al. 2017). Such wild germplasm can occur in the center(s) of domestication of crop species, as for apricot (Groppi et al. 2021), peach (Bao et al. 2017), and apple (Cornille et al. 2012; Duan et al. 2017) trees in Central Asia, and for olive trees in the Mediterranean basin (Besnard et al. 2017; Gros-Balthazard et al. 2019). Wild genetic diversity can also be found and is endemic in additional areas out of the center of origin (Chen, Avia, et al. 2023; Chen, Cornille, et al. 2023; Cornille et al. 2012; Xiao et al. 2023).

The history of the domestication of the almond trees is far less documented than that of other Mediterranean fruit trees. The olive, grape, and date palm domestication histories have been extensively studied (Besnard et al. 2017; Dong et al. 2023; Flowers et al. 2019; Pérez-Escobar et al. 2021). The cultivated almond (*Prunus dulcis* (Mill.) D.A. Webb) is an iconic species in Mediterranean and Middle Eastern history, primarily grown for its nuts, which are commercialized worldwide. Indeed, the first step in almond domestication was the selection of almonds with sweet and large kernels, as opposed to the wild almonds with bitter and small kernels (Alioto et al. 2020; Sánchez-Pérez et al. 2019). This fruit tree crop is an economic pillar of Mediterranean countries and is considered one of the five founders of the Near East Horticulture era, together with olive, grapevine, fig, and date palm (Zohary et al. 2012). Almonds belong to the genus *Prunus* of the Rosaceae family and the Prunoideae subfamily. Among the *Prunus* species, the subgenus *Amygdalus* includes two sections: *Persica* (with the peach and peach-related species) and *Amygdalus* (with the almond and almond-related species). The section *Amygdalus* is a monophyletic group of up to 26 described almond species that occur as trees or shrubs in Central Asia, the Caucasus, the Middle East, and the Mediterranean basin (Browicz and Zohary 1996; Rahemi and Gradziel 2024). Wild almond species are distributed over a large area of over 6000 km in longitude and between 25° and 45° north latitude (Grasselly 1976), which includes severe environments. Adaptation to harsh environmental conditions in almond species is possible due to their genetic and associated developmental and physiological diversity, which is promoted by their typically self-sterile yet

interspecifically fertile compatibility. In almond and stone fruit-related species, most sources of resistance to sharka, powdery mildew, green peach aphid, brown rot, and root-knot nematodes have been identified in wild species (Esmenjaud et al. 2009; Marimon et al. 2020; Oliveira Lino et al. 2016; Pascal et al. 2017; Pascal, Pfeiffer, and Kervella 2002; Pascal, Pfeiffer, Kervella, Lacroze, et al. 2002). The high potential of the wild almond species as sources of resistance enhances the interest in almond germplasm in stone fruit breeding programs.

Identifying the wild species that have contributed to the cultivated almond gene pool remains a topic of intense investigation. Hypotheses range from an origin in the Middle East or Central Asia to a complex scenario involving recurrent domestications and crop-wild gene flow. Domesticated almonds may have originated from the Middle East's wild forms of *P. dulcis*, which were once abundant in the Levantine countries (Lebanon, Israel, Jordan, Palestine, and Syria). Other authors argue that the forms of *P. dulcis* that grow spontaneously in the Middle East are wild types that have escaped cultivation, and that almond domestication occurred earlier in Central Asia, where the most significant number of wild almond species and the greatest number of hybrids are found (Browicz and Zohary 1996; Ladizinsky 1999). Another theory suggests that almond domestication was diffuse, with multiple domestication events involving recurrent wild-crop genetic exchanges from several wild species that contributed to the current domesticated almond genetic pool (Delplancke et al. 2013). The domestication scenarios of the almond are, therefore, still controversial. However, some wild almond species distributed in the Middle East, the Caucasus, and Central Asia are candidates (Browicz and Zohary 1996; Grasselly 1976). Populations of *Prunus orientalis* and *Prunus turcomanica* are described in the Southeast Anatolia region of Turkey, located in the Middle East (Ak et al. 2001). Frequent spontaneous genetic contributions of *P. orientalis* in the Middle East and *P. webbii* in Southern Europe to the cultivated *P. dulcis* suggest recurrent wild-to-crop introgressions (Delplancke et al. 2012; Gradziel and Martinez-Gomez 2013). From morphological features, it was hypothesized that the almond originated from *Prunus fenzliana* in the Caucasus (Ladizinsky 1999). Natural populations of *P. fenzliana* were identified along the eastern border of Turkey, Azerbaijan, and Iran (Tricon et al. 2023). Several wild almond species found in the Caucasus, the Middle East, and Central Asia likely contributed to the cultivated almond gene pool, either as initial progenitors or through recent wild-to-crop introgression (gene flow). However, a comprehensive sampling of wild species and *P. dulcis* across Eurasia is still necessary to fully understand the origins and processes of almond domestication. Understanding almond domestication and the genetic relationships between wild germplasm and the cultivated pool could also shed light on the origin of certain pests and diseases affecting almonds. For example, for sharka, one of the main viral diseases affecting *Prunus* cultivated species, different sources of resistance have been identified in *Amygdalus* germplasm (Pascal, Pfeiffer, and Kervella 2002; Tricon et al. 2023). However, it remains unclear whether they are phylogenetically related or if they represent distinct evolutionary origins.

In the current study, we investigated the domestication history of almonds using a comprehensive collection of almond cultivars

(*P. dulcis*) and their wild almond relatives (i.e., *P. fenzliana*, *P. orientalis*, *P. turcomanica*, *P. spinosissima*), which were genotyped for 23 microsatellite markers. Using population genetics, we explored (i) the population genetic diversity and structure, and genetic relationship among cultivated and wild almond relatives, (ii) the origin of the cultivated almonds and gene flow between crop and wild populations, and (iii) the variability of origin of resistance to sharka in almonds.

2 | Materials and Methods

2.1 | Plant Material

Wild almond samples were collected from adult trees between 2008 and 2012 in Azerbaijan, Kyrgyzstan, and Turkey (Eastern Anatolia along the Euphrates River and on the Akdamar Island of the Van Lake). *Prunus spinosissima* was sampled in 2019, after obtaining all local sampling and export authorizations. Almond cultivars used in this study are maintained at INRAE Bordeaux (UMR BFP), INRAE Avignon (GAFL), IRTA-Mas Bové research station (Constantí, Spain), ELGO-DIMITRA (Greece), and CU (Turkey).

In total, 209 samples of cultivated and wild almond species were used in this study (Table S1). For the cultivated almond, that is, *P. dulcis*, the 158 accessions originated from 14 countries as follows: Central Asia (Kyrgyzstan, Turkmenistan, and Uzbekistan, $N=8$), the Caucasus (Azerbaijan and Turkey, $N=43$), Russia ($N=2$), Middle East (Israel, $N=1$), Europe ($N=78$), North Africa (Tunisia, $N=1$), and North America (United States, $N=16$), plus nine accessions from advanced breeding programs (INRAE). A total of nine wild almond-related species were also sampled: *P. bucharica* ($N=1$), *P. communis* ($N=5$), *P. fenzliana* ($N=20$), *P. kuramica* ($N=1$), *P. orientalis* ($N=9$), *P. pedunculata* ($N=1$), *P. spinosissima* ($N=8$), *P. turcomanica* ($N=5$), and *P. webbii* ($N=1$) (Table S1). *P. communis*, *P. orientalis*, and *P. turcomanica* are native to Turkey; *P. fenzliana* is from Azerbaijan; *P. kuramica* is from Pakistan; *P. pedunculata* is from Mongolia, and *P. spinosissima* is from Kyrgyzstan. *P. bucharica* and *P. webbii* were provided by the ARS-USDA repository (Davis, USA), but their origin remains unknown.

2.2 | DNA Extraction and SSR Genotyping

Genomic DNA was extracted from leaves (Decroocq et al. 2016). Twenty-three microsatellite markers, distributed across the eight chromosomes of *P. dulcis*, *P. persica*, and *P. salicina*, were used for genotyping the 209 accessions (Table S2). Because some markers did not amplify well, the markers with more than 15% missing data across the collection were removed. SSR profiling was scored on an ABI sequencing platform (INRAE, GAFL Avignon). An M13 tail labeling strategy was used, and the fluorochrome LIZ 500 was used as an internal size marker. Raw data were transferred to Bordeaux, and the alleles were then determined and scored by GENEMAPPER v.3.7 software (Applied Biosystems). The bin sets (SSR profiles in Genemapper) for each marker were checked and adjusted according to the fingerprinting of the controls (R190, R269, R61, R40; see Table S1).

Additionally, the electropherograms of each sample and each marker were individually checked.

2.3 | Genetic Variation and Differentiation

Observed and expected heterozygosity (H_o and H_e , respectively) and the fixation index (F_{IS}) were calculated with GenAIEEx v6.503 (Peakall and Smouse 2012). The significance of pairwise genetic differentiation estimates (F_{ST} and Jost's D) was assessed with GenAIEEx v6.503. The allelic richness (A_r) and the private allelic richness (A_p) were calculated after adjustment for sample size differences among groups (set to four for the *P. dulcis* analysis and to five for the *P. dulcis* and wild-related species) through the rarefaction procedure implemented in Allelic Diversity Analyzer v1.0 (ADZE [Szpiech et al. 2008]).

2.3.1 | Inferences of Population Genetic Subdivision

From the genotyping data, clonemates were identified using GenoDive v.3.04 (Meirmans and Van Tienderen 2004) with a threshold of five. One individual per pair recognized as a clone or sibling was kept.

STRUCTURE v.2.3.4 (Pritchard et al. 2000) without a priori information was used to infer population structure and admixture among wild and cultivated almonds. STRUCTURE is based on Monte Carlo Markov Chain (MCMC) simulations, and it was used to deduce the proportion of ancestry of genotypes in K distinct clusters. Firstly, STRUCTURE was run with only cultivated almond trees (*P. dulcis*, after removal of clonemates and failed DNA extraction, $N=138$). Secondly, it was run with the above *P. dulcis*, together with wild relative species (*P. bucharica*, *P. kuramica*, *P. fenzliana*, *P. orientalis*, *P. spinosissima*, *P. turcomanica*, *P. webbii*, $N=51$). For each K , STRUCTURE runs consisted of 10 replicates of 10,000 burn-in steps followed by 100,000 MCMC iterations. The admixture model with correlated allele frequencies was selected. The resulting matrices of cluster membership coefficients were permuted with CLUMPP v.1.1.2 (Jakobsson and Rosenberg 2007), and bar plots were obtained with DISTRUCT v.1.1 (Rosenberg 2004). The ΔK was determined in the online post-processing software Structure Harvester (Earl and vonHoldt 2012) to determine the strongest level of genetic structure. However, it is possible that the optimal ΔK does not reflect the complete subdivision observed in STRUCTURE bar plots. It is, therefore, necessary to compare the obtained bar plots to choose the K value for which all clusters have well-assigned individuals, representing the finest genetic structure and not only the strongest, statistically supported by the ΔK (Puechmaillie 2016). The problem occurs when the program is forced to cluster individuals into an inappropriately small number of clusters (Cunningham et al. 2020; Puechmaillie 2016; Wang 2017). A more powerful method for identifying the best K value is to look at the bar plots and choose the highest K value for which no clusters are only represented by admixed individuals, indicating that we have reached the highest K value for which new genuine clusters could be delimited (Bina et al. 2022; Chen, Avia, et al. 2023; Chen, Cornille, et al. 2023; Cornille et al. 2014, 2012). The K value we therefore considered corresponded to the finest one, which can be higher than the K value of the strongest population structure level identified by ΔK .

Individuals with a membership coefficient $\geq 90\%$ to a *given* cluster were then assigned to a population (i.e., a group of individuals with a membership coefficient $\geq 90\%$ to a given cluster). Individuals with a membership coefficient < 0.90 were considered as admixed (putative hybrids between at least two genetic clusters).

We further explored genetic variation among populations defined above. A factorial correspondence analysis (FCA) was performed using the GENETIX v.4.05 (Belkhir et al. 2004) and then visualized with the “scatterplot3d” R package (Ligges and Mächler 2002). The software DARwin v.6.0.010 (Perrier and Jacquemoud-Collet 2006) explored genetic differentiation and relationships among samples using a neighbor-net based on Nei's standard genetic distance (D_{st}) and 30,000 bootstraps.

2.4 | Approximate Bayesian Computation to Reconstruct the Domestication History of the Almond

The almond domestication history was reconstructed using the approximate Bayesian computation (ABC) framework in combination with the coalescent-based inference simulator fastsimcoal2 v2.5.2.21 (Excoffier et al. 2021). We aimed to determine (1) whether gene flow occurred among wild populations and crop and wild populations during almond domestication, and (2) from which wild population(s) the cultivated almonds originated. The scenarios were established according to the results obtained from the STRUCTURE, FCA, neighbor-net tree, F_{ST} and Jost's D estimates. The populations were defined as those detected with STRUCTURE analyses. Putative hybrids (here referred to as “admixed”) were removed to trace historical gene flow among populations, while more recent admixture events were detected directly from the STRUCTURE analyses. The model parameters used were the divergence time between X and Y populations (T_{X-Y}), the effective population size of population X (N_{E-X}), and the migration rate per generation between X and Y populations (m_{X-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 listed in Table S3.

A newly developed ABC method was used for model selection and parameter estimation using a machine learning tool called “Random Forest” (ABC-RF). This approach enabled us to disentangle complex demographic models (Pudlo et al. 2015) by comparing groups of scenarios with a specific type of evolutionary event to other groups with different types of evolutionary events (instead of considering all scenarios separately) (Estoup et al. 2018), which we will hereafter refer to as “ABC rounds.” Such a grouping approach in scenario choice is more powerful than testing all scenarios individually to disentangle the main evolutionary events that characterize speciation (Estoup et al. 2018).

Three nested sets of ABC analyses were conducted. Within each set, two rounds were performed: a first round to test the modalities of gene flow among populations and a second round to investigate the domestication history of the cultivated almonds (Figure S1). For the first ABC set, the modalities of gene flow

were as follows: (1) no gene flow, (2) gene flow between crop and wild populations, and (3) gene flow between crop and wild populations only. For the two other ABC sets, there were (1) no gene flow or (2) gene flow between the crop and wild populations only. This nested ABC approach avoids the comparison of complex models with numerous populations and parameters (Estoup et al. 2018).

We used ABCtoolbox (Wegmann et al. 2010) with fastsimcoal 2.5 (Excoffier and Foll 2011) to simulate the datasets. For all models, the microsatellite datasets were simulated for 23 markers. We assumed a generalized stepwise model of microsatellite evolution (Slatkin 1995). Mutation rates were allowed to vary across loci with locus-specific mutation rates, where μ is the mutation rate per generation, with a log-uniform prior distribution for μ (10^{-4} , 10^{-3}) (Table S3). The other model parameters were drawn from the prior distributions (Table S1). We performed 10,000 simulations per scenario. For each simulation, we calculated three summary statistics per population with arlsumstats v3.5 (Excoffier and Lischer 2010): H , the mean heterozygosity across loci, NGW , the adjusted mean Garza-Williamson statistic over populations (Garza and Williamson 2001), the pairwise F_{ST} (Weir and Cockerham 1984) and genetic distances $(\delta\mu)^2$ (Goldstein et al. 1995) between pairs of populations.

We used the abcrf v.1.7.0 R statistical package (Pudlo et al. 2015) to conduct the ABC-RF analysis. This analysis provides a classification vote representing the number of times a scenario is selected as the best among the 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 the best group of scenarios among 500 classification trees (Breiman 2001). We computed the posterior probabilities and prior error rates (i.e., the probability of choosing an incorrect group of scenarios when drawing the 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 visually verified that the simulated models were compatible with the observed dataset by projecting both datasets onto the first two linear discriminant analysis (LDA) axes (Pudlo et al. 2015) and checking that the observed dataset fell within the cloud of the simulated datasets. We then calculated parameter inferences using the final selected model. The median and 90% confidence intervals (CI, 5%–95%) are given for each model parameter estimate. Note that the ABC-RF approach includes the model-checking step performed a posteriori in previous ABC methods.

3 | Results

Before analysis, four individuals (Turc_114, Vairo from the French and Greek collections, and Occhiorosso-2) were removed as they showed more than 50% missing data. SSR fingerprinting also revealed that the US143 accession was polyploid (and was thus discarded); it belongs to the *P. pedunculata* species from Mongolia, where almond species displaying a chromosomal number exceeding the expected $2n = 16$ have been described (up to $2n = 88$) (Ekimova et al. 2012). Thirty-two pairs of clones of cultivated almonds were identified by Genodive analysis. Only

one of the two (or three) individuals with high genetic relatedness was retained for analysis. In total, 186 individuals (89%) were retained for the subsequent analyses, including 138 cultivated and 48 wild almonds.

Out of the initial 35 microsatellite markers tested in this study, two (UDP98-408 and CPDCT035) were filtered out because of too high a rate of missing data, null alleles, or unclear genotype peak patterns (Table S2). The neutrality test implemented in Arlequin v3.5 detected no markers that significantly deviated from a neutral equilibrium model. In particular, marker CPDCT035 did not amplify in *P. fenzliana* samples. Although a majority of the initial genetic markers (23 out of 25, Table S2) were derived from *Amygdalus* or *Persica* species, the incidence of missing data or null alleles did not exhibit a significant correlation with the marker's taxonomic origin. This is evidenced by the robust performance of the two Simple Sequence Repeat (SSR) markers originating from *P. salicina* (plum) species (Table S2). In total, we retained 23 microsatellite markers for subsequent analyses of neutral genetic variation and differentiation (Table S2).

3.1 | Genetic Differentiation and Relationships Among *P. dulcis* and Its Related Species

To characterize the diversity and genetic subdivision among the cultivated almond and related species, we first ran STRUCTURE on the cured dataset ($N=186$, Table S1bis) composed of cultivated (*P. dulcis*, $N=138$) and wild almond-related species, *P. bucharica* ($N=1$), *P. communis* ($N=5$), *P. fenzliana* ($N=19$), *P. kuramica* ($N=1$), *P. orientalis* ($N=8$), *P. spinosissima* ($N=8$), *P. turcomanica* ($N=5$), and *P. webbii* ($N=1$), that were genotyped with 23 microsatellite markers. The change rate in the log-likelihood between successive K values (ΔK) inferred with STRUCTURE revealed three peaks at $K=2$, $K=4$, and $K=7$, respectively (Figure S2). At $K=4$, *P. fenzliana* formed a distinct cluster. The same was true for the *P. dulcis* samples from Akdamar Island (purple color in Figure 1 and Figure S3). A large proportion of the cultivated almonds also formed a distinct cluster (Figure S3). But based on the different bar plots obtained from $K=2$ to $K=10$ (Figure S3), we deduced that the most relevant value of ΔK is $K=7$. For $K < 7$, some clusters appeared admixed, while they appeared non-admixed and well-delimited at $K=7$. Further increasing K above 7 did not reveal well-defined new clusters, except for those distinguishing *P. orientalis* and *P. turcomanica* ($K=8$). However, this differentiation was not stable, as at $K > 8$, the two species were once again indistinguishable (Figure S3). This altogether suggested that $K=7$ corresponded to the most relevant K value for our sampling. The STRUCTURE analysis revealed ambiguous assignments for specific genotypes, particularly those represented by only one or two individuals (e.g., *P. bucharica*, *P. webbii*). In other cases, such as the grouping of *P. orientalis* and *P. turcomanica*, the observed patterns may reflect either taxonomic misidentification or resolution limitations due to the use of only 23 SSR loci. Additionally, insufficient within-population sampling can hinder the reliable detection of genetic clusters or gradients (see Puechmaillie 2016). For these reasons, we excluded such genotypes from subsequent analyses to ensure more robust and interpretable results.

At $K=7$, the cultivated almonds split into four distinct clusters (Figure 2A), similar to the ones described above (Figure S3)

(pink, purple, light, and dark blue) and the wild-related species formed two main, non-admixed clusters, *P. fenzliana* and *P. spinosissima* (yellow and green, respectively in Figures 1A and 2B). The other six wild-related species were either non-distinguishable from cultivated *P. dulcis* (*P. communis*, thus indicating a feral origin), admixed between a wild species and *P. dulcis* (*P. webbii*, *P. bucharica*, *P. kuramica*, thus indicating a crop-to-wild gene flow), or admixed between *P. spinosissima* and another, still-unknown cluster (*P. orientalis* and *P. turcomanica*, in gray Figures 1A and 2B). Bayesian clustering thus showed that individuals were mainly classified according to their species and/or origin (for the *P. dulcis* clusters). We confirmed the four genetic clusters of cultivated almonds ($N=138$) in a separate Bayesian analysis (Data S1, Table S4, Figure S4), and pairwise Jost's D among the four cultivated almond clusters were all significant (Table 1).

Subsequently, to estimate the genetic distance between the different almond and almond-related populations, admixed individuals were removed from the dataset using a 90% assignment threshold to at least one of the seven genetic clusters. At the 90% threshold, a total of 131 non-admixed accessions were retained as follows: *P. communis* ($N=4$), *P. dulcis* ($N=89$), *P. fenzliana* ($N=19$), *P. orientalis* ($N=6$), *P. spinosissima* ($N=8$), and *P. turcomanica* ($N=5$). The distribution of the almond and almond-related clusters inferred by STRUCTURE (Figure 2B) confirmed a clear spatial differentiation between the four *P. dulcis* clusters and between *P. dulcis* and its wild-related species following an East to West axis (Figure 1B).

Further analyses of genetic differentiation between populations (excluding admixed individuals) revealed clear separation between *P. dulcis* and its wild-related species, as well as among the *P. dulcis* populations (after removing non-*P. dulcis* samples). In the FCA (Figure 2B) and neighbor-joining tree (Figure 3), the wild accessions separated from *P. dulcis*. In the NJ tree, the wild relatives, *P. orientalis* and *P. turcomanica*, were monophyletic but had relatively low bootstrap support (Figure S5). The different wild species clustered together, with a distinct clade for *P. fenzliana*. The overall structure of the dendrogram also aligned with the clustering indicated by STRUCTURE, except for the Caucasian/Central Asian *P. dulcis* cluster (shown in pink in the NJ tree, Figure 3), which split into four main clades.

3.1.1 | Genetic Diversity in Wild and Cultivated Almond Populations

As we showed that *P. communis* individuals were most probably feral *P. dulcis*, we grouped the two species, *P. communis* and *P. dulcis*, for subsequent analyses. We then computed population genetic diversity (Table 2) for the five remaining species, excluding *P. webbii*, *P. kuramica*, and *P. bucharica*, as they were represented by a low number of individuals ($N=1$ or 2 for each). The highest fixation index was that of *P. dulcis* ($F_{IS}=0.164$, ± 0.032 , Table 2). However, this value was difficult to compare with the values of the other species due to a significantly unbalanced sample size (i.e., $N=92$ for *P. dulcis* but $N=5$ for *P. turcomanica*). Using ADZE software, we determined the allelic richness of the different populations weighted by the sample size. As a result, *P. fenzliana* not only showed significantly higher allelic

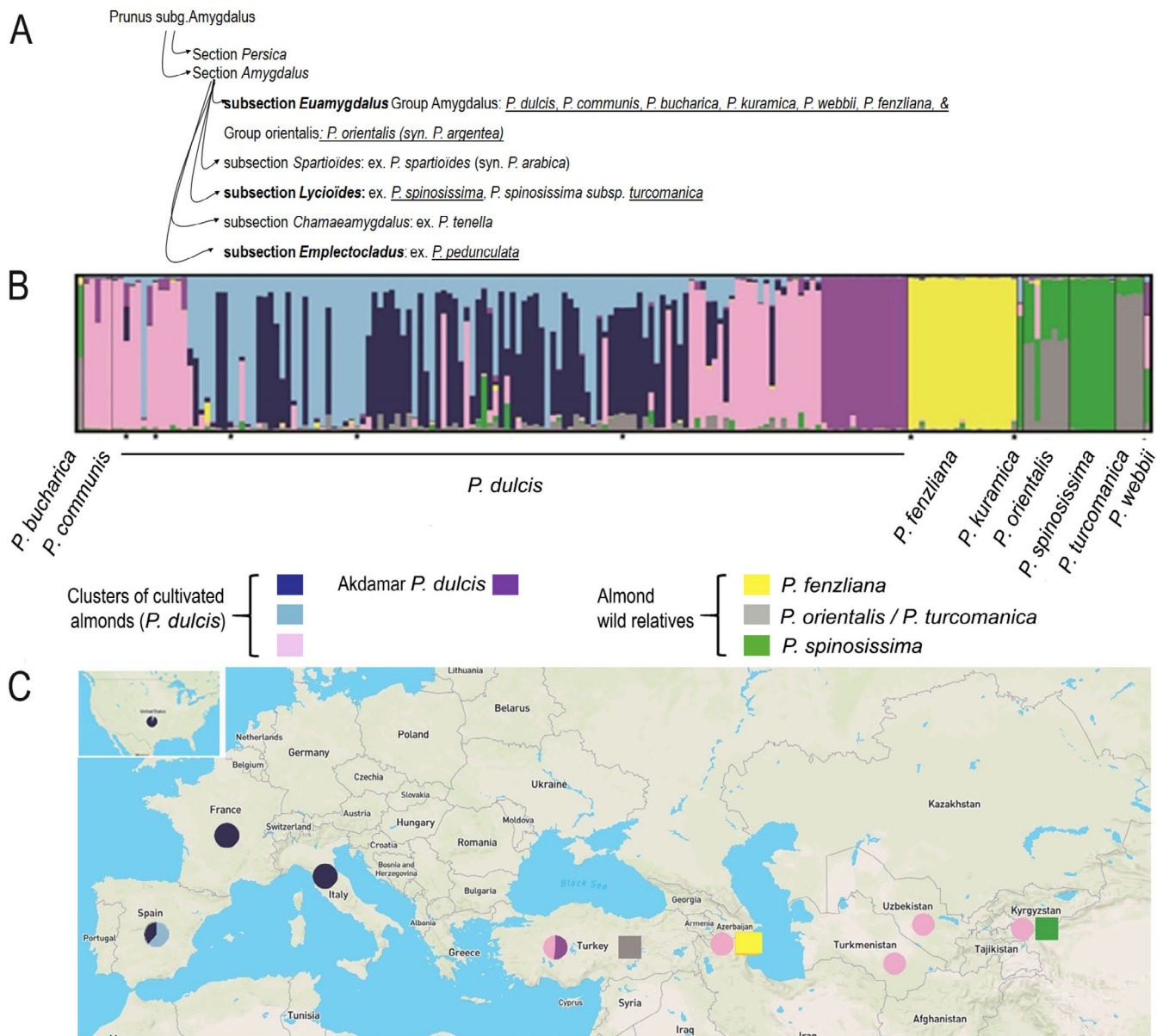


FIGURE 1 | Population genetic structure of *Prunus dulcis* ($N=89$) and its wild relative species (*Prunus fenzliana*, $N=19$, *Prunus orientalis*, $N=6$, *Prunus spinosissima*, $N=8$, and *Prunus turcomanica*, $N=5$) inferred with STRUCTURE at $K=7$. (A) Classification of almond and almond-related species. Adapted from Browicz and Zohary (1996) and Grasselly (1976). Underlined, species included in this study. (B) Bayesian clustering results inferred with STRUCTURE at $K=7$. Underneath the figure are depicted the names of the crop and wild Amygdalus populations as described in Table 2. Each individual is represented by a vertical bar, divided into K segments representing the amount of ancestry in its genotype corresponding to K clusters. (*) refers to Plum Pox virus-resistant samples as listed in Table S1. Species names are provided primarily as contextual information to relate our genetic clusters to previously identified taxonomic classifications. (C) Spatial population genetic structure of Amygdalus clusters and its wild relatives along the Northern hemisphere. The world map can be downloaded under a free license at <https://www.vecteezy.com/vector-art/10961532-world-map-vector-illustration-isolated-on-grey-background-flat-earth-globe-or-world-map>.

richness ($A_r = 4.118 \pm 0.138$, Table 2) than the other four species, but also the lowest value of unbiased expected heterozygosity ($uH_e = 0.710 \pm 0.051$, Table 2).

3.1.2 | Approximate Bayesian Computation Analyses for Testing Demographic Scenarios

We defined the populations used in the ABC framework from the clusters detected with STRUCTURE at $K=7$ (Figure 1) for a total of 131 wild and cultivated almond accessions, excluding

admixed individuals (i.e., with a membership coefficient <0.90 to any given cluster, as recent gene flow can easily be seen from visual inspection of the bar plots). We included *P. spinosissima*, *P. orientalis*, and *P. fenzliana*, two Turkish almond populations (purple and pink), and northwestern cultivated almonds (Spanish and French/North American). The other species were not retained further in the analysis because of their too low number of samples (one or two representatives) or because they were too much admixed and could not be assigned to any population. We also simulated an ancestral population. Eight populations were therefore simulated (Figure S1).

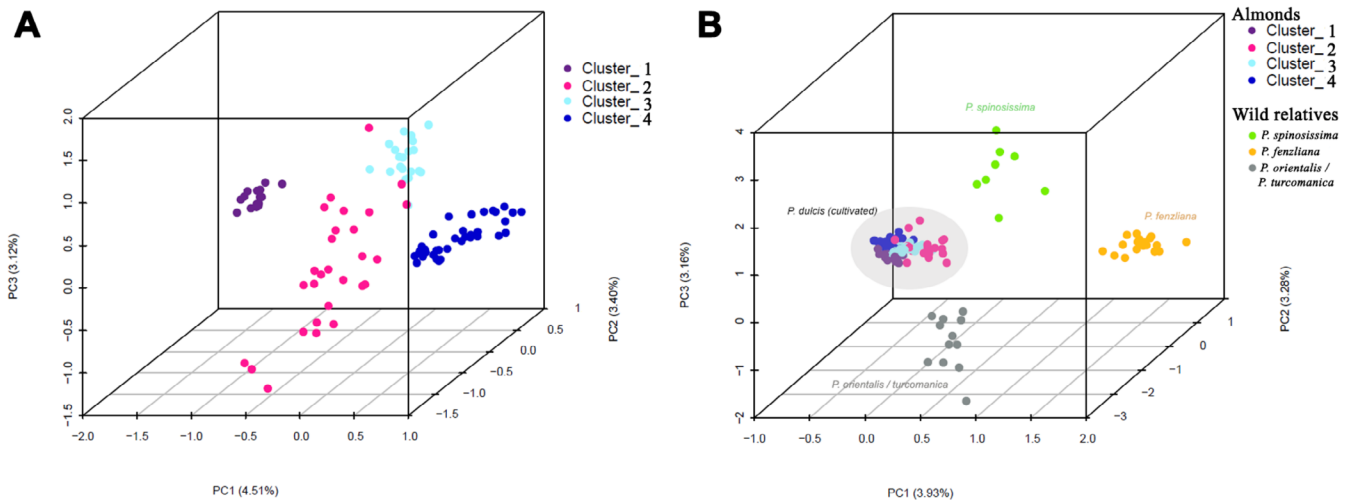


FIGURE 2 | Factorial correspondence analyses (FCAs). (A) Including *Prunus dulcis* individuals ($N=98$), colored as in Figure 1. (B) Including *P. dulcis* and related species individuals ($N=131$) with membership coefficient $\geq 90\%$ to a cluster in the STRUCTURE analysis at $K=7$, colored as in Figure S5. The cluster numbers correspond to (1–4) *P. dulcis*, (5) *P. fenzliana*, (6) *Prunus orientalis* and *Prunus turcomanica*, and (7) *Prunus spinosissima*. Note that individuals with a probability of assignment to one of the seven clusters $< 90\%$ were removed.

TABLE 1 | Pairwise population matrix of Jost's estimate of differentiation (Jost's D) among the *Prunus dulcis* clusters inferred with GenAlEx V.6.503.

Jost's D			Cluster_1	Cluster_2	Cluster_3	Cluster_4
Cluster_1		Akdamar island	—	0.001	0.001	0.001
Cluster_2		Caucasus/Central Asia	0.480	—	0.001	0.001
Cluster_3		South Europe	0.489	0.257	—	0.001
Cluster_4		EU/North-America	0.552	0.260	0.267	—

Note: Only the individuals assigned to a genetic cluster with a membership proportion greater than or equal to 90%, were retained. Jost's D values are displayed below the diagonal and the probability above diagonal. All pairwise Jost's D values were significant ($p < 0.05$, number of permutations = 999). Colors correspond to genetic clusters identified in Figure 1B.

To keep the number of scenarios tractable, we fixed the divergence histories of the wild almond populations and that of the North American/French from the Spanish cultivated almonds as previously known and inferred from the neighbor-joining tree (Figure S1), F_{ST} estimates, and PCA. In the first ABC set, we built three scenarios to test whether the purple Turkish cultivated almond population diverged from one of the three wild populations (Figure S1). These three scenarios were simulated under three alternative hypotheses of gene flow: (1) no gene flow, (2) gene flow between wild populations, and (3) gene flow between crop and wild populations. Thus, nine scenarios (three gene flow hypotheses $\times$ three divergence scenarios) were simulated (Figure S1). Once the most likely scenario was chosen, it was used as a backbone for the second ABC set, which tested the origin of the pink Turkish cultivated almond population from one of the three wild populations and the purple Turkish cultivated almond population (three divergence histories, Figure S1). These four scenarios were simulated under two alternative hypotheses of gene flow: (1) no gene flow and (2) gene flow between crop and wild populations only. Once the most likely scenario was chosen, it served as a backbone for the third ABC set, which tested the origin of the Northwestern cultivated almond populations (Spanish

and North American/French) from one of the three wild almond populations (three scenarios, Figure S1). Once again, these three scenarios were simulated under two alternative hypotheses of gene flow: (1) no gene flow and (2) gene flow between crop and wild populations only.

For each round 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 (Figures S6–S8), forming different clouds for each scenario or group of scenarios. Visual inspection of the LDA plots indicated that we had enough power to discriminate and select scenarios; the results were subsequently validated by the ABC-RF inferences presented below.

In the first ABC set, which tested the divergence history of the purple Turkish cultivated almond population, all 10 replicate ABC-RF analyses supported the occurrence of gene flow only among wild and crop populations (Figure S1, Table S5, an average of 453 votes out of the 500 RF trees; posterior probabilities = 94%, prior error rate = 5.8%) and a sister relationship

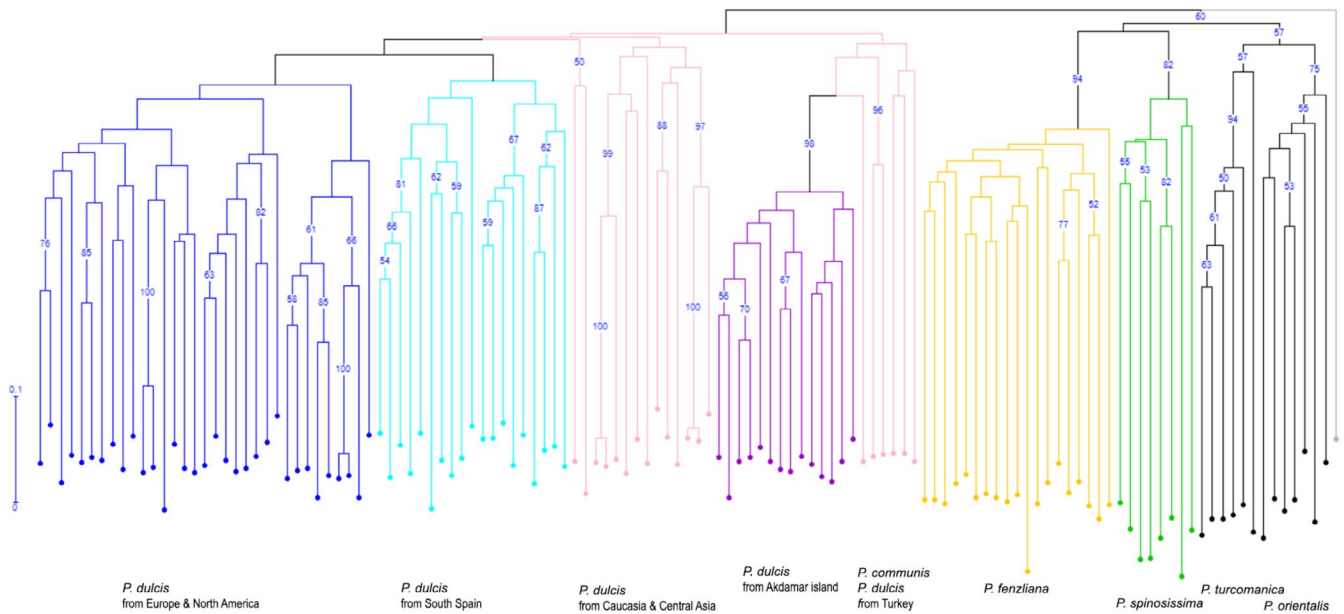


FIGURE 3 | Rooted weighted neighbor-joining (NJ) tree among cultivated and wild almond species performed with DARwin. The 131 non-admixed samples belong to six different *Prunus* species including *P. communis* ($N=4$) in pink and *P. dulcis* ($N=89$) in pink, light blue, and dark blue, *P. fenzliana* ($N=19$) in yellow, *P. orientalis* ($N=6$) in gray, *P. spinosissima* ($N=8$) in green, and *P. turcomanica* ($N=5$) in dark gray. In light gray (far right), *P. kuramica* which was used to root the tree. The NJ tree was built with DARwin, bootstrap support values were obtained from 30,000 repetitions. Bootstrap values when greater than 50% are shown above the branches. Species names are provided primarily as contextual information to relate our genetic clusters to previously identified taxonomic classifications.

TABLE 2 | Genetic diversity estimated for the four cultivated almond populations (*Prunus dulcis*) and three wild species (*Prunus spinosissima*, *Prunus fenzliana*, *Prunus orientalis/turcomanica*) inferred at $K=7$ with STRUCTURE.

Crop or wild	Species	Populations	N	H_o	uH_e	F_{IS}	A_p	A_r
Crop	<i>P. dulcis</i>	Akdamar island, (purple)	15	0.53	0.56	0.045 NS	0.47 (± 0.09)	3.36 (± 0.26)
		Caucasus/Central Asia (pink)	23	0.72	0.81	0.11	1.16 (± 0.18)	6.06 (± 0.40)
		South Europe (light blue)	20	0.73	0.74	0.011 NS	0.63 (± 0.10)	4.85 (± 0.28)
		EU/North-America (dark blue)	36	0.66	0.75	0.12	0.45 (± 0.09)	4.47 (± 0.22)
Wild	<i>P. spinosissima</i>	Central Asia (green)	8	0.78	0.85	0.08	2.43 (± 0.29)	6.18 (± 0.34)
	<i>P. fenzliana</i>	Caucasus (yellow)	19	0.64	0.71	0.10	1.79 (± 0.27)	4.87 (± 0.34)
	<i>P. orientalis/turcomanica</i>	Middle East (gray)	12	0.70	0.81	0.13	2.15 (± 0.25)	5.83 (± 0.36)

Note: Only individuals assigned to a genetic cluster with a membership proportion greater than or equal to 90% were retained, that is, 131 accessions. The observed heterozygosity (H_o), the unbiased expected heterozygosity (uH_e), and the fixation index (F_{IS}) were calculated thanks to GenAlEx v.6.503, and add-in off Excel. The allelic richness (A_r) and the private allelic richness (A_p) were calculated thanks to ADZE v.1.0. The standard deviations are depicted in brackets. Species names are provided primarily as contextual information to relate our genetic clusters to previously identified taxonomic classifications.

between the purple Turkish cultivated and *P. orientalis* (Figure S1, Table S6, an average of 292 votes out of the 500 RF trees; posterior probabilities=68%, prior error rate=29.49%, Table S9). In the second ABC set, all 10 replicate ABC-RF analyses supported the occurrence of gene flow among the wild and crop populations (Figure S1, Table S7, an average of 270 out of the 500 RF trees; posterior probabilities=63%, prior error rate=14.6%) and a sister relationship between the pink Turkish cultivated and *P. orientalis* (Figure S1, Table S8, an average of 307 votes out of the 500 RF trees; posterior

probabilities=66%, prior error rate=17.66%). In the third ABC set, all 10 replicate ABC-RF analyses supported the absence of gene flow among wild and European crop populations (Figure S1, Table S9, an average of 258 out of the 500 RF trees; posterior probabilities=71%, prior error rate=2%) and a sister relationship between the northwestern cultivated almonds and *P. orientalis* (Figure S1, Table S10, an average of 357 votes out of the 500 RF trees; posterior probabilities=75%, prior error rate=39.21%). This nested ABC approach avoids comparing too many complex models with too many parameters

The detection of sharka resistance in several *P. dulcis* clusters and in *P. fenzliana* further underscores the potential value of wild gene pools for breeding. Overall, our results offer a foundation for further investigations into Turkey's role in almond diversification and domestication, and contribute to the growing body of evidence supporting complex domestication histories involving multiple wild taxa and recurrent introgression in perennial crops.

4.1 | Turkey as a Diversification Center and the Role of Sampling in Resolving Almond Phylogeny

After the divergence between peach and almond that occurred millions of years ago in Central Asia, the almond species dispersed westwards and established itself, particularly in harsh environments, from Central Asia to the Mediterranean area, including the Caucasus and the Middle East (Velasco et al. 2016; Yazbek and Al-Zein 2014). Previous publications have shown the existence of several almond wild relatives around the Mediterranean Basin; however, none have investigated the relationship between cultivated almonds and their wild relatives due to the lack of a proper sampling design. Indeed, in most cases, authors would compare collections of almond cultivars with only one or two samples per related species (and most often of dubious origin). Lately, we showed in the case of another *Prunus* species, *P. brigantina*, the importance of sample size and sampling design that should encompass sufficient genetic diversity at the species level to result in reliable phylogenetic studies (Liu et al. 2021).

Here, we significantly expanded the sampling of *Amygdalus* species, which is still limited in some wild species compared to cultivated almonds. It includes worldwide almond cultivars, as well as natural populations of wild relatives from Turkey, the Caucasus, and Central Asia. We identified four genetically differentiated clusters of cultivated almonds (*P. dulcis*) with distinct geographical distributions, ranging from Central Asia to Europe and North America. Based on the genotyping of 152 almond landraces with a 60K SNP array, de Pérez los Cobos et al. (2023) confirmed our clustering, but they also identified further subdivisions within the European/North American group (de Pérez los Cobos et al. 2023). In our study, Turkey emerged as a diversification center for cultivated almonds, characterized by two distinct genetic clusters that may have resulted from population isolation and island speciation, as observed in the case of the Akdamar population. This late population would have thus diverged due to local adaptation or reproductive isolation (Westram et al. 2022). Indeed, the case of the Akdamar population must be considered with caution, as its isolation may affect pairwise differentiation and coalescent models. Furthermore, gene flow between the different trees on the island would induce an increase in homozygosity and allele fixation, which could be the cause of the genetic differentiation of this population from the other *P. dulcis* clusters (Halász et al. 2019) reported comparable findings, demonstrating through microsatellite genotyping at 15 loci that Turkish almond accessions resolved into two distinct genetic clusters: one associated with Akdamar Island and another closely related to other Central Asian cultivars. Their

study also indicated allele fixation and elevated homozygosity within the Akdamar population. However, contrary to their findings, our investigation did not confirm the presence of Akdamar alleles in the US cultivars.

4.2 | Geographic Dispersal and Wild-Crop Gene Flow

Besides those two genetic populations in Turkey, we observed two other ones in Europe. One of the two European clusters occurred specifically in Spanish almond cultivars; it may therefore correspond to the Middle Eastern and North African populations (Delplancke et al. 2013). Indeed, after the Muslim conquest of the Iberian Peninsula (711–726 AD), southern Spain has long been a hub of trade between Arabic and European countries (Delplancke et al. 2013). A close relationship between the Southern Spanish and North African gene pools has already been documented for another stone fruit species, apricot (Bourguiba et al. 2012), and more recently for almond (de Pérez los Cobos et al. 2023). It was recently shown that Moroccan almonds present private alleles, most probably resulting from hybridization with *P. arabica*, and thus indicating ancient dissemination routes through Northern Africa toward Spain (Delplancke et al. 2013; Halász et al. 2019). The phylogeographic origin of South Spanish almond cultivars, specifically their potential derivation from a North African gene pool, requires further empirical validation. This necessitates a comparative genetic analysis of both European and North African almond germplasm.

The second European population of cultivated almonds encompasses accessions from North Spain, France, Italy, and Eastern Europe. The identification of a similar sub-clade of European almonds in prior studies suggests a secondary route for the introduction of *P. dulcis* into Europe, potentially from the East or via ancient Phoenician and Greek trade roads (Delplancke et al. 2013; Halász et al. 2019). For North American individuals, it was not surprising to find them closely related to one of the two European genetic clusters, since all American cultivars originated either directly or after hybridization from more ancient European varieties (de Pérez los Cobos et al. 2021).

We performed Bayesian analysis, highlighting the differentiation between cultivated almond trees and their wild related species. Cluster analysis using SSRs as valuable markers allowed for the discrimination and estimation of relationships among accessions. When wild-related species are included in the study, *P. dulcis* is still differentiated into four clusters, while *P. communis* samples grouped with the Turkish, Caucasian, and Central Asian *P. dulcis* cultivars and landraces. Moreover, in the phylogenetic tree, *P. communis* individuals were placed in the same clade as *P. dulcis*. We can therefore deduce that the species *P. communis* is a synonym of *P. dulcis*, as previously proposed (Rahemi et al. 2012). Concerning *P. fenzliana*, two independent studies have proposed this species as the wild ancestor of almond (Ladizinsky 1999; Zeinalabedini et al. 2010). According to our phylogenetic analysis, *P. fenzliana* appears as a distinct species, separate from the other *Amygdalus* species, with limited evidence of admixture. Our phylogenetic study did not support an exclusive relationship between *P. dulcis* and *P. fenzliana*. It showed a kinship between a set of *P. dulcis* cultivars (from

- and Recommended Conservation Strategies." *PLoS One* 12, no. 11: e0188685. <https://doi.org/10.1371/journal.pone.0188685>.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2004. "GENETIX4. 05, Logiciel sous Windows TM pour la génétique des Populations." Laboratoire Génome, Populations, Interactions, CNRS UMR 5000 Université de Montpellier II, 5000.
- Besnard, G., J.-F. Terral, and A. Cornille. 2017. "On the Origins and Domestication of the Olive: A Review and Perspectives." *Annals of Botany* 121, no. 3: 385–403. <https://doi.org/10.1093/aob/mcx145>.
- Bina, H., H. Yousefzadeh, A. Venon, et al. 2022. "Evidence of an Additional Centre of Apple Domestication in Iran, With Contributions From the Caucasian Crab Apple *Malus orientalis* Uglitzk. To the Cultivated Apple Gene Pool." *Molecular Ecology* 31, no. 21: 5581–5601. <https://doi.org/10.1111/mec.16667>.
- Bourguiba, H., J. M. Audergon, L. Krichen, et al. 2012. "Loss of Genetic Diversity as a Signature of Apricot Domestication and Diffusion Into the Mediterranean Basin." *BMC Plant Biology* 12: 49. <https://doi.org/10.1186/1471-2229-12-49>.
- Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. <https://doi.org/10.1023/A:1010950718922>.
- Browicz, K., and D. Zohary. 1996. "The Genus *Amygdalus* L. (Rosaceae): Species Relationships, Distribution and Evolution Under Domestication." *Genetic Resources and Crop Evolution* 43, no. 3: 229–247. <https://doi.org/10.1007/bf00123275>.
- Burgarella, C., A. Barnaud, N. A. Kane, et al. 2019. "Adaptive Introgression: An Untapped Evolutionary Mechanism for Crop Adaptation." *Frontiers in Plant Science* 10: 4. <https://doi.org/10.3389/fpls.2019.00004>.
- Chen, X., K. Avia, A. Forler, et al. 2023. "Ecological and Evolutionary Drivers of Phenotypic and Genetic Variation in the European Crabapple [*Malus sylvestris* (L.) Mill.], a Wild Relative of the Cultivated Apple." *Annals of Botany* 131, no. 6: 1025–1037. <https://doi.org/10.1093/aob/mcad061>.
- Chen, X., A. Cornille, N. An, et al. 2023. "The East Asian Wild Apples, *Malus baccata* (L.) Borkh and *Malus hupehensis* (Pamp.) Rehder., Are Additional Contributors to the Genomes of Cultivated European and Chinese Varieties." *Molecular Ecology* 32, no. 18: 5125–5139. <https://doi.org/10.1111/mec.16485>.
- Cornille, A., T. Giraud, M. J. Smulders, I. Roldán-Ruiz, and P. Gladieux. 2014. "The Domestication and Evolutionary Ecology of Apples." *Trends in Genetics* 30, no. 2: 57–65. <https://doi.org/10.1016/j.tig.2013.10.002>.
- Cornille, A., P. Gladieux, M. J. M. Smulders, 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, no. 5: e1002703. <https://doi.org/10.1371/journal.pgen.1002703>.
- Cunningham, C. I., J. M. Miller, R. M. Peery, et al. 2020. "Confidently Identifying the Correct K Value Using the ΔK Method: When Does $K=2$?" *Molecular Ecology* 29, no. 5: 862–869. <https://doi.org/10.1111/mec.15374>.
- Dal Martello, R., M. von Baeyer, M. Hudson, et al. 2023. "The Domestication and Dispersal of Large-Fruiting *Prunus* spp.: A Metadata Analysis of Archaeobotanical Material." *Agronomy* 13, no. 4: 1027.
- de Pérez los Cobos, F., E. Coindre, N. Dlalal, et al. 2023. "Almond Population Genomics and Non-Additive GWAS Reveal New Insights Into Almond Dissemination History and Candidate Genes for Nut Traits and Blooming Time." *Horticulture Research* 10, no. 10: uhad193. <https://doi.org/10.1093/hr/uhad193>.
- de Pérez los Cobos, F., P. J. Martínez-García, A. Romero, et al. 2021. "Pedigree Analysis of 220 Almond Genotypes Reveals Two World Mainstream Breeding Lines Based on Only Three Different Cultivars." *Horticulture Research* 8, no. 1: 11. <https://doi.org/10.1038/s41438-020-00444-4>.
- Decroocq, S., A. Cornille, D. Tricon, et al. 2016. "New Insights Into the History of Domesticated and Wild Apricots and Its Contribution to Plum Pox Virus Resistance." *Molecular Ecology* 25, no. 19: 4712–4729. <https://doi.org/10.1111/mec.13772>.
- Delplancke, M., N. Alvarez, L. Benoit, A. Espindola, S. Neuenschwander, and N. Arrigo. 2013. "Evolutionary History of Almond Tree Domestication in the Mediterranean Basin." *Molecular Ecology* 22, no. 4: 1092–1104. <https://doi.org/10.1111/mec.12129>.
- Delplancke, M., N. Alvarez, A. Espindola, et al. 2012. "Gene Flow Among Wild and Domesticated Almond Species: Insights From Chloroplast and Nuclear Markers." *Evolutionary Applications* 5, no. 4: 317–329. <https://doi.org/10.1111/j.1752-4571.2011.00223.x>.
- Dong, Y., S. Duan, Q. Xia, et al. 2023. "Dual Domestications and Origin of Traits in Grapevine Evolution." *Science* 379, no. 6635: 892–901. <https://doi.org/10.1126/science.add8655>.
- Duan, N., Y. Bai, H. Sun, et al. 2017. "Genome Re-Sequencing Reveals the History of Apple and Supports a Two-Stage Model for Fruit Enlargement." *Nature Communications* 8, no. 1: 249. <https://doi.org/10.1038/s41467-017-00336-7>.
- Earl, D. A., and B. M. vonHoldt. 2012. "STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method." *Conservation Genetics Resources* 4, no. 2: 359–361. <https://doi.org/10.1007/s12686-011-9548-7>.
- Ekimova, N., E. Muratova, and P. Silkin. 2012. "The Role of Polyploidy in Adaptation and Settling of Steppe Shrubs in Central Asia." *Russian Journal of Genetics: Applied Research* 2: 105–109. <https://doi.org/10.1134/S2079059712020037>.
- Esmenjaud, D., R. Voisin, C. Ghelder, et al. 2009. "Genetic Dissection of Resistance to Root-Knot Nematodes *Meloidogyne* spp. in Plum, Peach, Almond, and Apricot From Various Segregating Interspecific *Prunus* Progenies." *Tree Genetics & Genomes* 5: 279–289. <https://doi.org/10.1007/s11295-008-0173-x>.
- Estoup, A., L. Raynal, P. Verdu, and J.-M. Marin. 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 Société Française de Statistique* 159: 167–190.
- Excoffier, L., and M. Foll. 2011. "Fastsimcoal: A Continuous-Time Coalescent Simulator of Genomic Diversity Under Arbitrarily Complex Evolutionary Scenarios." *Bioinformatics* 27, no. 9: 1332–1334. <https://doi.org/10.1093/bioinformatics/btr124>.
- Excoffier, L., and H. E. L. Lischer. 2010. "Arlequin Suite ver 3.5: A New Series of Programs to Perform Population Genetics Analyses Under Linux and Windows." *Molecular Ecology Resources* 10, no. 3: 564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>.
- Excoffier, L., N. Marchi, D. A. Marques, R. Matthey-Doret, A. Gouy, and V. C. Sousa. 2021. "fastsimcoal2: Demographic Inference Under Complex Evolutionary Scenarios." *Bioinformatics* 37, no. 24: 4882–4885. <https://doi.org/10.1093/bioinformatics/btab468>.
- Fangning, Z., and B. Jacqueline. 2020. "Exploring the Application of Wild Species for Crop Improvement in a Changing Climate." *Current Opinion in Plant Biology* 56: 218–222. <https://doi.org/10.1016/j.pbi.2019.12.013>.
- Flowers, J. M., K. M. Hazzouri, M. Gros-Balthazard, 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, no. 5: 1651–1658. <https://doi.org/10.1073/pnas.1817453116>.
- Garza, J. C., and E. G. Williamson. 2001. "Detection of Reduction in Population Size Using Data From Microsatellite Loci." *Molecular*

- Ecology* 10, no. 2: 305–318. <https://doi.org/10.1046/j.1365-294X.2001.01190.x>.
- Gaut, B. S., C. M. Diez, and P. L. Morrell. 2015. “Genomics and the Contrasting Dynamics of Annual and Perennial Domestication.” *Trends in Genetics* 31, no. 12: 709–719. <https://doi.org/10.1016/j.tig.2015.10.002>.
- Glémin, S., and T. Bataillon. 2009. “A Comparative View of the Evolution of Grasses Under Domestication.” *New Phytologist* 183, no. 2: 273–290. <https://doi.org/10.1111/j.1469-8137.2009.02884.x>.
- Goldstein, D. B., A. Ruiz Linares, L. L. Cavalli-Sforza, and M. W. Feldman. 1995. “An Evaluation of Genetic Distances for Use With Microsatellite Loci.” *Genetics* 139, no. 1: 463–471. <https://doi.org/10.1093/genetics/139.1.463>.
- Gradziel, T. M. 2020. “Redomesticating Almond to Meet Emerging Food Safety Needs.” *Frontiers in Plant Science* 11: 778. <https://doi.org/10.3389/fpls.2020.00778>.
- Gradziel, T., and P. Martinez-Gomez. 2013. “Almond Breeding.” In *Plant Breeding Reviews*, edited by J. Janick, vol. 37, 207–258. Wiley-Blackwell.
- Grasselly, C. 1976. “Les espèces sauvages d’amandiers.” In *L’amandier*, vol. 32, 28–43. CIHEAM.
- Groppi, A., S. Liu, A. Cornille, et al. 2021. “Population Genomics of Apricots Unravels Domestication History and Adaptive Events.” *Nature Communications* 12, no. 1: 3956. <https://doi.org/10.1038/s41467-021-24283-6>.
- Gros-Balthazard, M., G. Besnard, G. Sarah, et al. 2019. “Evolutionary Transcriptomics Reveals the Origins of Olives and the Genomic Changes Associated With Their Domestication.” *Plant Journal* 100, no. 1: 143–157. <https://doi.org/10.1111/tpl.14435>.
- Halász, J., O. Kodad, G. M. Galiba, et al. 2019. “Genetic Variability Is Preserved Among Strongly Differentiated and Geographically Diverse Almond Germplasm: An Assessment by Simple Sequence Repeat Markers.” *Tree Genetics & Genomes* 15, no. 1: 12. <https://doi.org/10.1007/s11295-019-1319-8>.
- Jakobsson, M., and N. A. Rosenberg. 2007. “CLUMPP: A Cluster Matching and Permutation Program for Dealing With Label Switching and Multimodality in Analysis of Population Structure.” *Bioinformatics* 23, no. 14: 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>.
- Kester, D. E., and T. M. Gradziel. 1996. *Almonds*, edited by J. Janick and J. N. Moore, vol. III. Wiley J and Sons.
- Kester, D. E., T. M. Gradziel, and C. Grasselly. 1990. “Almonds (*Prunus*).” *Acta Horticulturae* 290: 699–758.
- Ladizinsky, G. 1999. “On the Origin of Almond.” *Genetic Resources and Crop Evolution* 46, no. 2: 143–147. <https://doi.org/10.1023/a:1008690409554>.
- Ligges, U., and M. Mächler. 2002. “Scatterplot3d—An R Package for Visualizing Multivariate Data.” *Journal of Statistical Software* 8: 1–20. <https://doi.org/10.17877/DE290R-15194>.
- Liu, S., S. Decroocq, E. Harte, et al. 2021. “Genetic Diversity and Population Structure Analyses in the Alpine Plum (*Prunus brigantia* Vill.) Confirm Its Affiliation to the Armeniaca Section.” *Tree Genetics & Genomes* 17: 2. <https://doi.org/10.1007/s11295-020-01484-6>.
- Marimon, N., J. Luque, P. Arús, and I. Eduardo. 2020. “Fine Mapping and Identification of Candidate Genes for the Peach Powdery Mildew Resistance Gene Vr3.” *Horticulture Research* 7, no. 1: 175. <https://doi.org/10.1038/s41438-020-00396-9>.
- Meirmans, P. G., and P. H. Van Tienderen. 2004. “Genotype and Genodive: Two Programs for the Analysis of Genetic Diversity of Asexual Organisms.” *Molecular Ecology Notes* 4, no. 4: 792–794. <https://doi.org/10.1111/j.1471-8286.2004.00770.x>.
- Meyer, R. S., A. E. DuVal, and H. R. Jensen. 2012. “Patterns and Processes in Crop Domestication: An Historical Review and Quantitative Analysis of 203 Global Food Crops.” *New Phytologist* 196, no. 1: 29–48. <https://doi.org/10.1111/j.1469-8137.2012.04253.x>.
- Miller, A. J., and B. L. Gross. 2011. “From Forest to Field: Perennial Fruit Crop Domestication.” *American Journal of Botany* 98, no. 9: 1389–1414. <https://doi.org/10.3732/ajb.1000522>.
- Neale, D. B., P. J. Martínez-García, A. R. D. L. Torre, S. Montanari, and X.-X. Wei. 2017. “Novel Insights Into Tree Biology and Genome Evolution as Revealed Through Genomics.” *Annual Review of Plant Biology* 68, no. 1: 457–483. <https://doi.org/10.1146/annurev-arplant-042916-041049>.
- Oliveira Lino, L., I. Pacheco, V. Mercier, et al. 2016. “Brown Rot Strikes Prunus Fruit: An Ancient Fight Almost Always Lost.” *Journal of Agricultural and Food Chemistry* 64, no. 20: 4029–4047. <https://doi.org/10.1021/acs.jafc.6b00104>.
- Pascal, T., R. Aberlenc, C. Confolent, et al. 2017. “Mapping of New Resistance (Vr2, Rm1) and Ornamental (Di2, pl) Mendelian Trait Loci in Peach.” *Euphytica* 213: 132. <https://doi.org/10.1007/s10681-017-1921-5>.
- Pascal, T., F. Pfeiffer, and J. Kervella. 2002. “Preliminary Observations on the Resistance to Sharka in Peach and Related Species.” *Acta Horticulturae* 592: 699–704. <https://doi.org/10.17660/ActaHortic.2002.592.98>.
- Pascal, T., F. Pfeiffer, J. Kervella, J. P. Lacroze, M. H. Sauge, and W. E. Weber. 2002. “Inheritance of Green Peach Aphid Resistance in the Peach Cultivar ‘Rubira.’” *Plant Breeding* 121, no. 5: 459–461. <https://doi.org/10.1046/j.1439-0523.2002.734333.x>.
- Peakall, R., and P. E. Smouse. 2012. “GenAlEx 6.5: Genetic Analysis in Excel. Population Genetic Software for Teaching and Research—An Update.” *Bioinformatics* 28, no. 19: 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>.
- Pérez-Escobar, O. A., S. Bellot, N. A. S. Przelomska, et al. 2021. “Molecular Clocks and Archeogenomics of a Late Period Egyptian Date Palm Leaf Reveal Introgression From Wild Relatives and Add Timestamps on the Domestication.” *Molecular Biology and Evolution* 38, no. 10: 4475–4492. <https://doi.org/10.1093/molbev/msab188>.
- Perrier, X., and J. P. Jacquemoud-Collet. 2006. “DARwin software.” <http://darwin.cirad.fr/>.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. “Inference of Population Structure Using Multilocus Genotype Data.” *Genetics* 155, no. 2: 945–959. <https://doi.org/10.1093/genetics/155.2.945>.
- Pudlo, P., J.-M. Marin, A. Estoup, J.-M. Cornuet, M. Gautier, and C. P. Robert. 2015. “Reliable ABC Model Choice via Random Forests.” *Bioinformatics* 32, no. 6: 859–866. <https://doi.org/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, no. 3: 608–627. <https://doi.org/10.1111/1755-0998.12512>.
- Rahemi, A., R. Fatahi, A. Ebadi, et al. 2012. “Genetic Diversity of Some Wild Almonds and Related Prunus Species Revealed by SSR and EST-SSR Molecular Markers.” *Plant Systematics and Evolution* 298, no. 1: 173–192. <https://doi.org/10.1007/s00606-011-0536-x>.
- Rahemi, A., and T. M. Gradziel. 2024. “Almond Species Classification.” In *The Almonds and Related Species: Identification, Characteristics and Uses*, edited by T. M. Gradziel, 9–48. Springer International Publishing.
- Rosenberg, N. A. 2004. “Distruct: A Program for the Graphical Display of Population Structure.” *Molecular Ecology Notes* 4, no. 1: 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>.
- Rubio, M., P. Martinez-Gomez, F. Dicenta, and W. Weber. 2008. “Resistance of Almond Cultivars to Plum Pox Virus (Sharka).” *Plant Breeding* 122: 462–464. <https://doi.org/10.1046/j.1439-0523.2003.00872.x>.

- Sánchez-Pérez, R., S. Pavan, R. Mazzeo, et al. 2019. "Mutation of a bHLH Transcription Factor Allowed Almond Domestication." *Science* 364, no. 6445: 1095–1098. <https://doi.org/10.1126/science.aav8197>.
- Slatkin, M. 1995. "A Measure of Population Subdivision Based on Microsatellite Allele Frequencies." *Genetics* 139, no. 1: 457–462. <https://doi.org/10.1093/genetics/139.1.457>.
- Szpiech, Z. A., M. Jakobsson, and N. A. Rosenberg. 2008. "ADZE: A Rarefaction Approach for Counting Alleles Private to Combinations of Populations." *Bioinformatics* 24, no. 21: 2498–2504. <https://doi.org/10.1093/bioinformatics/btn478>.
- Tricon, D., J. Faivre d'Arcier, J.-P. Eyquard, et al. 2023. "Allele Mining of Eukaryotic Translation Initiation Factor Genes in Prunus for the Identification of New Sources of Resistance to Sharka." *Scientific Reports* 13, no. 1: 15247. <https://doi.org/10.1038/s41598-023-42215-w>.
- Velasco, D., J. Hough, M. Aradhya, and J. Ross-Ibarra. 2016. "Evolutionary Genomics of Peach and Almond Domestication." *G3 (Bethesda)* 6, no. 12: 3985–3993. <https://doi.org/10.1534/g3.116.032672>.
- Wang, J. 2017. "The Computer Program Structure for Assigning Individuals to Populations: Easy to Use but Easier to Misuse." *Molecular Ecology Resources* 17, no. 5: 981–990. <https://doi.org/10.1111/1755-0998.12650>.
- Wegmann, D., C. Leuenberger, S. Neuenschwander, and L. Excoffier. 2010. "ABCtoolbox: A Versatile Toolkit for Approximate Bayesian Computations." *BMC Bioinformatics* 11, no. 1: 116. <https://doi.org/10.1186/1471-2105-11-116>.
- Weir, B., and C. Cockerham. 1984. "Estimating F-Statistics for the Analysis of Population-Structure." *Evolution* 38: 1358–1370. <https://doi.org/10.2307/2408641>.
- Westram, A. M., S. Stankowski, P. Surendranadh, and N. Barton. 2022. "What Is Reproductive Isolation?" *Journal of Evolutionary Biology* 35, no. 9: 1143–1164. <https://doi.org/10.1111/jeb.14005>.
- Xiao, H., Z. Liu, N. Wang, et al. 2023. "Adaptive and Maladaptive Introgression in Grapevine Domestication." *Proceedings of the National Academy of Sciences of the United States of America* 120, no. 24: e2222041120. <https://doi.org/10.1073/pnas.2222041120>.
- Yazbek, M. M., and M. S. Al-Zein. 2014. "Wild Almonds Gone Wild: Revisiting Darwin's Statement on the Origin of Peaches." *Genetic Resources and Crop Evolution* 61, no. 7: 1319–1328. <https://doi.org/10.1007/s10722-014-0113-6>.
- Zeinalabedini, M., M. Khayam-Nekoui, V. Grigorian, T. M. Gradziel, and P. Martinez-Gomez. 2010. "The Origin and Dissemination of the Cultivated Almond as Determined by Nuclear and Chloroplast SSR Marker Analysis." *Scientia Horticulturae* 125: 593–601. <https://doi.org/10.1016/j.scienta.2010.05.007>.
- Zhang, H., N. Mittal, L. J. Leamy, O. Barazani, and B.-H. Song. 2017. "Back Into the Wild—Apply Untapped Genetic Diversity of Wild Relatives for Crop Improvement." *Evolutionary Applications* 10, no. 1: 5–24. <https://doi.org/10.1111/eva.12434>.
- Zohary, D., E. Weiss, and M. Hopf. 2012. "Fruit Trees and Nuts." In *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*, edited by D. Zohary, M. Hopf, and E. Weiss. Oxford University Press.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** eva70150-sup-0001-Figures.docx. **Table S1:** eva70150-sup-0002-Tables.xlsx.