New insights into the history of domesticated and wild apricots and its contribution to *Plum pox virus* resistance

STÉPHANE DECROOCQ,*† AMANDINE CORNILLE,‡[1] DAVID TRICON,*† SEVDA BABAYEVA,§
AURÉLIE CHAGUE,*† JEAN-PHILIPPE EYQUARD,*† RAUL KARYCHEV,¶
SVETLANA DOLGIKH,¶ TATIANA KOSTRITSYNA,**[2] SHUO LIU,††*† WEISHENG LIU,††
WENJUAN GENG,‡‡ KANG LIAO,‡‡ BAYRAM M. ASMA,§§ ZEYNAL AKPAROV,§
TATIANA GIRAUD¶¶ and VÉRONIQUE DECROOCQ*†

*UMR 1332 BFP, Equipe de virologie, INRA, 71 Avenue Edouard Bourlaux, 33882 Villenave d'Ornon, France, †UMR 1332 BFP, CS20032, Université de Bordeaux, 33882 Villenave d'Ornon, France, ‡Department of Ecology and Genetics, Evolutionary Biology Centre, Science for life Laboratory, Uppsala University, Uppsala, Sweden, §Genetic Resources Institute of ANAS, Azadlig ave. 155, AZ1106 Baku, Azerbaijan, ¶Kazakh Research Institute of Horticulture and Viticulture, 238-a Gagarin Avenue, 480060 Almaty, Kazakhstan, **Botanical Garden of National Academy of Sciences, Akhunbaeva street 1a, 720064 Bishkek, Kyrgyzstan, ††Liaoning Institute of Pomology, Tiedong Street, Xiongyue Town, Bayuquan District, Yingkou City, Liaoning 115009, China, ‡‡ College of Horticulture & Forestry Sciences, Xinjiang Agricultural University, 311 NongDaDong Road, 830052 Urumqi City, Xinjiang, China, §§Department of Horticulture, Inonu University, Malatya 44210, Turkey, ¶¶Ecologie Systematique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

Abstract

Studying domesticated species and their wild relatives allows understanding of the mechanisms of population divergence and adaptation, and identifying valuable genetic resources. Apricot is an important fruit in the Northern hemisphere, where it is threatened by the *Plum pox virus* (PPV), causing the sharka disease. The histories of apricot domestication and of its resistance to sharka are however still poorly understood. We used 18 microsatellite markers to genotype a collection of 230 wild trees from Central Asia and 142 cultivated apricots as representatives of the worldwide cultivated apricot germplasm; we also performed experimental PPV inoculation tests. The genetic markers revealed highest levels of diversity in Central Asian and Chinese wild and cultivated apricots, confirming an origin in this region. In cultivated apricots, Chinese accessions were differentiated from more Western accessions, while cultivated apricots were differentiated from wild apricots. An approximate Bayesian approach indicated that apricots likely underwent two independent domestication events, with bottlenecks, from the same wild population. Central Asian native apricots exhibited genetic subdivision and high frequency of resistance to sharka. Altogether, our results contribute to the understanding of the domestication history of cultivated apricot and point to valuable genetic diversity in the extant genetic resources of wild apricots.

Keywords: fruit tree, pathogen, population structure, *Prunus armeniaca*, virus, wild progenitor

Received 28 February 2016; revision received 14 July 2016; accepted 25 July 2016

Introduction

Domestication is an evolutionary process by which humans produce, from wild species, populations with modified traits, by selecting individuals most suited to cultivation and consumption (Gerbault *et al.* 2014). Domestication often involves a loss of genetic diversity in crops relative to their wild progenitors (Glemin &

Correspondence: Véronique Decroocq, Fax: +33 557122384; E-mail: decroocq@bordeaux.inra.fr and Tatiana Giraud, Fax: +33 169154697; E-mail: tatiana.giraud@u-psud.fr
[1]Present address: Center for Adaptation to a Changing Environment, ETH Zürich, 8092 Zürich, Switzerland
[2]Present address: International School of Medicine, 1 f, Street Intergelpo, 720054 Bishkek, Kyrgyzstan

Bataillon 2009; Miller & Gross 2011). Many traits desired by breeders, such as resistances to major crop pest/diseases, may thus be lacking in the cultivated germplasm while present in wild relatives, for example mildew resistance in grape germplasm (Riaz *et al.* 2013). A comprehensive analysis of the genetic and phenotypic diversity in both crop- and wild-related species can help to understand domestication history and to identify valuable genetic resources in wild populations for further crop breeding.

Prunus armeniaca L. (also called *Armeniaca vulgaris*) (Lingdi & Bartholomew 2003) corresponds to both the domesticated apricot, cultivated worldwide, and wild populations, now only growing in Central Asia. Mid-twentieth-century studies based on morphological features suggested that the domesticated apricot originated from Asia about 5000 years ago, possibly with two successive domestication events, one in western Central Asia (Fergana valley, at the border of Uzbekistan, Tajikistan and Kyrgyzstan) and one in China (Vavilov 1951). The apricot would have then been transferred to the Irano-Caucasian area, where it would have undergone further improvement. These domestication events might have been associated with bottlenecks (Bourguiba *et al.* 2012). Later on, apricot was brought to Europe during Alexander the Great's incursions in Asia (356–323 BC) or via the Silk Roads (Janik 2005; Ugurtan Yilmaz & Gurcan 2012). More recently, it spread with the Spaniards from Europe to North America and the rest of the world. Key features distinguish domesticated apricots from wild populations: leaf and fruit size, cold hardiness and, to some extent, reproductive system. Part of the Mediterranean cultivars are indeed self-compatible while wild trees are strictly self-incompatible (Halász *et al.* 2007). Wild populations of *P. armeniaca* once occurred on vast areas in Fergana Valley, Almaty region down to Pakistan and Afghanistan (Kostina 1946), but they are now found within a much more restricted area in Central Asia, on the slopes of the Tien Shan ranges, likely leading to a reduction in valuable apricot genetic resources (Dzhangaliev *et al.* 2003).

Studies based on germplasm acquisition trips in Central Asia by the Nikitsky Botanical Garden in Yalta (Crimea) in the late 1950s distinguished two main ecogeographic groups within locally cultivated and wild *P. armeniaca*, mainly based on fruit size*:* (i) a group in Central Asia (including the western Chinese province of Xinjiang, Pakistan, Afghanistan and Northern India) and (ii) a 'Dzhungar-Zailig' group (corresponding mostly to the Dzhungarsky and Zailisky national parks located in the Tien Shan, between Kazakhstan, Kyrgyzstan and the northern part of the Chinese province of Xinjiang) (Kostina 1969). In addition, two other ecogeographic groups were described that corresponded to cultivated apricots in the Irano-Caucasian region (i.e. Iran, Iraq, Syria, Azerbaijan, Georgia, Armenia, eastern Anatolia in Turkey) and in Europe, respectively. Previous studies aiming at evaluating genetic diversity in apricot focused either on cultivated European apricots and a few samples from botanical collections (Romero *et al.* 2003; Zhebentyayeva *et al.* 2003, 2008; Maghuly *et al.* 2005; Pedryc *et al.* 2009), or on Chinese apricots (He *et al.* 2007; Yuan *et al.* 2007; Zhang *et al.* 2014). The present study aimed at elucidating the domestication history of the cultivated apricot and characterizing the genetic diversity and population structure of the remaining wild *P. armeniaca* populations in their natural habitat in Central Asia using more comprehensive collections. Our aim was to gain fundamental knowledge on a key evolutionary process, domestication, as well as on sources of diversity in wild apricots, in particular regarding resistance to diseases.

One of the most devastating diseases on apricots is the sharka disease, caused by a potyvirus, the *Plum Pox Virus* (PPV hereafter), to which all European cultivars are susceptible (Dosba *et al.* 1991), and that affects more generally all stone fruit species (i.e. *Prunus* sp.). The disease was first identified about one century ago, in Bulgaria, on European plums (*P. domestica*) and described officially in 1933 (Atanasoff 1933). Meanwhile, the virus spread on other *Prunus* species, apricot included (Atanasoff 1935). Trade activity, including the exchange of grafted plants, facilitated the rapid spread of PPV across Europe (Rimbaud *et al.* 2015), and more recently in America and Asia (EPPO 2014). In Central Asia, sharka was first detected only 10 years ago (Spiegel *et al.* 2004), with no indication of earlier occurrence of sharka in this region. Yet, the resistance to sharka discovered in a few North American apricot cultivars has been suggested to result from an introgression from Asian apricots (Zhebentyayeva *et al.* 2008; Pedryc *et al.* 2009). There may therefore be further sources of resistance to sharka in Asia. Previous studies pointed to one major locus responsible for resistance, *PPVres* (Lambert *et al.* 2007; Marandel *et al.* 2009; Pilařová *et al.* 2010; Dondini *et al.* 2011; Vera Ruiz *et al.* 2011). Markers linked to this locus in the North American apricot cultivars have been developed (Soriano *et al.* 2012; Decroocq *et al.* 2014). However, selection based exclusively on the *PPVres* locus was not sufficient to unambiguously select PPV-resistant apricot cultivars (Decroocq *et al.* 2014; Rubio *et al.* 2014), suggesting that recombination can occur between the markers and the resistance locus or that this locus is not sufficient by itself to determine resistance to sharka. In fact, a second locus involved in resistance has been recently identified, *MetaPPV1b* (Mariette *et al.* 2016).

Here, we investigated genetic diversity and spatial population structure in cultivated and wild apricots as a basis to then infer the history of apricot domestication and assess its genetic variability in Asia, including regarding apricot sharka resistance. For these goals, we collected a comprehensive collection of wild and cultivated apricot trees from Central Asia and Caucasus, with also representatives of the main western European and North American cultivated phylogroups (Bourguiba *et al.* 2012). This collection was genotyped using 18 molecular markers, 15 being putatively neutral (Bourguiba *et al.* 2012) and four being linked in cultivated apricots to the identified loci conferring resistance to sharka (Soriano *et al.* 2012; Decroocq *et al.* 2014; Mariette *et al.* 2016). We also experimentally tested the actual resistance to sharka using inoculations on a subset sample of wild apricots. The purpose was to trace back the geographic origin of resistance to sharka, check its linkage with the identified resistance loci in apricot wild populations and assess the resistance variability in wild populations.

More specifically, our questions were as follows: (i) To what extent are wild and cultivated apricots genetically differentiated? (ii) Is there any population structure within wild and cultivated apricot populations, respectively? (iii) What is the most likely scenario of domestication: a single event, two independent events from different wild populations, or two successive events from the same wild population? (iv) Have bottlenecks been associated with apricot domestication? (v) What is the frequency of the resistance to the sharka disease in wild apricots, and does this inform on PPV origin?

Material and methods

Sample collection

The apricot collection in total included 372 individuals with 142 cultivated and 230 wild accessions. Wild accessions as referred hereafter were considered as 'wild' when sampled in natural forest mountains, and cultivated accessions were considered as samples of recent varieties, landraces, and ancient local, possibly feral, varieties. Wild apricot trees in Central Asia are usually growing, away from the cultivation areas, in mountainous forests at an altitude ranging from 1500 to 2000 m high, together with the wild apple (*Malus sieversii*) and walnut trees.

The 142 cultivated accessions had different origins: cultivars come from orchards or breeding germplasm of the INRA and the ARS-USDA national repositories (Tables 1 and S1, Supporting information); landraces were collected in private gardens and do not originate

from breeding programmes; ancient local, possibly feral, varieties correspond to isolated trees, along roads or in abandoned human settlements (such as the Turk former settlements in the former Armenia/Kurdistan area, in Turkey and Azerbaijan), in areas where no forest with apricots is present. This collection included cultivated apricots from: (i) the putative primary domestication centre (Kazakhstan, China, Kyrgyzstan and Uzbekistan) (*n* = 75, CULT03 to 10 and CULT13), (ii) the secondary putative domestication centre (Azerbaijan, eastern Anatolia in Turkey) (*n* = 37, CULT01 and CULT11), (iii) at least one representative of each cultivated Mediterranean/North American phylogroup (Bourguiba *et al.* 2012), that is 'Bakour', 'Bergeron', 'Canino', 'Currot', 'Goldrich', 'Luizet', 'Moniqui', 'Stark Early Orange', 'Stella' (CULT02, *n* = 9), and (iv) Central Asian accessions from Pakistan, Afghanistan and Turkmenistan (CULT12, *n* = 21).

Additionally, a total of 230 'wild' accessions were collected across 15 Central Asian natural sites, in mountainous forests of Kazakhstan, Kyrgyzstan, Uzbekistan and western China between 2011 and 2014 (Table 1, WILD01 to WILD15). This corresponds to the range of distribution of the Dzungar-Zailig ecogeographic group as described by Kostina (1946). Leaf material was retrieved from adult trees; when this was not possible due to degradation of the collected leaves, collected seeds were germinated and leaves were sampled on the resulting seedlings (one single seedling per mother plant) (Table S1, Supporting information). The latitude and longitude coordinates of the plant material were recorded (Fig. S1, Supporting information).

DNA extraction and molecular genotyping

Genomic DNA was extracted using a modified previously published protocol (Doyle & Doyle 1987). Each apricot leaf sample (8–16 cm^2) was grinded in 5 mL of buffer 1 (0.2 M Tris–HCl pH8, 0.07 M EDTA, 2 M NaCl, 0.02 M Sodium metabisulfite). A total of 500 μL of the first extract was completed with 450 μL of buffer 2 (2% HATMAB, 1.4 M NaCl, 0.02 M EDTA pH8, 0.1 M Tris–HCl pH 8) and incubated for 30 min at 65 °C. After Chloroform: IAA (isoamylic alcohol) extraction and isopropanol precipitation, supplemented with 10 M ammonium acetate, DNA was resuspended in 200 μL of 0.1× TE buffer (Tris–HCl/EDTA).

Microsatellite markers were amplified using previously published multiplex PCR protocols (Bourguiba *et al.* 2012; Decroocq *et al.* 2014). We initially used 23 microsatellite loci distributed across the eight chromosomes (Table S2, Supporting information) and one SSLP (simple sequence length polymorphism) marker (ZP002) (Decroocq *et al.* 2014). See details in Supporting

Table 1 List of apricot, cultivated and wild groups sampled in this study

Code*	Type of material†	Habitat	Origin	Site of collection	Ecogeographic group
Breeding varieties, local and ancient cultivars and landraces					
Cult01_AZE	Landraces	Cultivated	Azerbaijan	Azerbaijan_landraces	Irano-caucasian
Cult02_EUR	Cultivars	Cultivated	EU, USA	Occidental_cultivars (EU, USA)	European
Cult03_CHN	Local cultivars	Cultivated	China	Xiongyue_apricot_repository‡	Chinese
Cult04_KAZ	Landraces and wild	Cultivated	Kazakhstan	Almaty_Pomological_garden	Dzhungar-Zailig
Cult05_KAZ	Landraces	Semi-wild	Kazakhstan	Ak-Kain	Dzhungar-Zailig
Cult06_KAZ	Landraces/local cultivars	Cultivated	Kazakhstan	Chymkent_Dendro_Park‡	Central Asian
Cult07_KAZ	Landraces	Semi-wild	Kazakhstan	Sayram	Dzhungar-Zailig
Cult08_KAZ	Landraces	Semi-wild	Kazakhstan	Wine_yard_irrigation_canal	Dzhungar-Zailig
Cult09_KAZ	Landraces and cultivars	Cultivated	Kazakhstan	Almaty_market	Dzhungar-Zailig
Cult10_UZB	Landraces	Cultivated	Uzbekistan	Boukhara	Central Asian
Cult11_TUR	Landraces	Cultivated	Turkey	Eastern_Anatolian_landraces	Irano-caucasian
Cult12_PAK	Local cultivars	Cultivated	Pakistan	Pakistan_landraces	Central Asian
Cult13_KGZ	Landraces	Cultivated	Kyrgyzstan	Arslan_Bob	Central Asian
P. armeniaca natural populations					
Wild01_CHN	Wild	Montane forest	China	Ily_valley	Dzhungar-Zailig
Wild02_KAZ	Wild	Montane forest	Kazakhstan	Belbulak_Canyon	Dzhungar-Zailig
Wild03_KAZ	Wild	Montane forest	Kazakhstan	Esik_Lake	Dzhungar-Zailig
Wild04_KAZ	Wild	Montane forest	Kazakhstan	Medeu_valley	Dzhungar-Zailig
Wild05_KAZ	Wild	Montane forest	Kazakhstan	Turgen_Valley	Dzhungar-Zailig
Wild06_KAZ	Wild	Montane forest	Kazakhstan	Kaskelen_Canyon	Dzhungar-Zailig
Wild07_KAZ	Wild	Montane forest	Kazakhstan	Big_Almaty_Lake	Dzhungar-Zailig
Wild08_KAZ	Wild	Montane forest	Kazakhstan	Aksu_Zhabagyly_National_Park	Dzhungar-Zailig
Wild09_UZB	Wild	Montane forest	Uzbekistan	Chimgan_Beldersay	Central Asian
Wild10_KGZ	Wild	Montane forest	Kyrgyzstan	Issyk_Kul_Urukty_river	Dzhungar-Zailig
Wild11_KGZ	Wild	Montane forest	Kyrgyzstan	Issyk_Kul_Anan'Yevo_village	Dzhungar-Zailig
Wild12_KGZ	Wild	Montane forest	Kyrgyzstan	Issyk_Kul_Orto_Byrosun_river	Dzhungar-Zailig
Wild13_KGZ	Wild	Montane forest	Kyrgyzstan	Chuy_River_/_Boom_canyon	Dzhungar-Zailig
Wild14_KGZ	Wild	Montane forest	Kyrgyzstan	Ala_Archa_National_Park	Dzhungar-Zailig
Wild15_KGZ	Wild	Montane forest	Kyrgyzstan	Sary_Chelek_National_Park	Central Asian

*Cult# refers to geographic groups of apricot landraces and cultivars, followed by country codes; Wild# refers to natural populations sampled in Kazakhstan (KAZ), China (CHN), Uzbekistan (UZB) and Kyrgyzstan (KGZ).

†In semi-wild are included landraces or local, ancient selections. They can also correspond to feral wild apricot trees.

‡The CULT03_CHN group is a repository of apricot accessions collected all over eastern and western China. Chinese provinces where apricots were collected are displayed in Table S1 (Supporting information). Similarly, CULT06_KAZ includes samples from Dzhungar-Zailig and old, local varieties from the Central Asian ecogeographic group (Tajikistan, Uzbekistan, etc.) which were collected in Soviet times. The ecogeographic groups correspond to the classification of Kostina (1969).

information online. Alleles were scored with the GeneMapper® software (Applied Biosystem).

Genetic variation and differentiation

The observed heterozygosity (H_0), the unbiased expected heterozygosity (H_E) and the inbreeding coefficient F_{IS} were calculated using SPAGEDI 1.3 (Hardy & Vekemans 2002) and verified with GENETIX v4.05 (Belkhir *et al.* 2004). Allelic richness (A_R) and private allelic richness (A_{PR}) were computed with ADZE software to adjust for sample size differences (Szpiech *et al.* 2008).

We further explored the genetic differentiation and relationships among samples using an unweighted Neighbor-Joining tree constructed using simple matching dissimilarity indices of Jaccard's coefficient method and bootstrap values over 2000 replicates as implemented in the DARWIN software package v6.0.010 (Perrier & Jacquemoud-Collet 2006). Among-population F_{ST} and Nei's indices were estimated using ARLEQUIN v3.5 (Excoffier & Lischer 2010). The significance of F_{ST} was assessed by random resampling of the genotypic data through 1000 permutations. Jost's D estimates among populations (Jost 2008) were computed using GENALEX (Peakall & Smouse 2012), and their significance was established by bootstrapping over 1000 replicates. Mean differences between Jost's D values were tested using Student's *t*-tests with JMP v.7.0 (SAS Institute).

Within the wild apricot sampling (n = 260 individuals), we tested for an isolation-by-distance pattern using a Mantel's test, as implemented in SPAGEDI (http://ebe.ulb.ac.be/ebe/SPAGeDi.html) with 10 000 permutations (Rousset 2008).

Population structure

Representation of the genetic relationships among individuals was explored with a factorial correspondence analysis (FCA) performed with GENETIX v4.05 (Belkhir *et al.* 2004). We also used the individual-based Bayesian clustering method implemented in STRUCTURE 2.3.3 (Pritchard *et al.* 2000) to investigate population subdivision. We ran STRUCTURE from K = 2 to K = 10 using admixture and correlated allele frequencies assuming no prior information. Burn-in and number of Markov chain Monte Carlo iterations were set to 10 000 and 100 000, respectively. Ten independent runs were carried out for each K, and outputs were processed with CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007). STRUCTURE barplots were displayed using DISTRUCT 1.1 (Rosenberg 2004). For all these analyses, the markers linked to sharka resistance were removed. We examined the distribution of ΔK, plotted with STRUCTURE harvester (http://taylor0.biology.ucla.edu/structureHarvester/).

Approximate bayesian computation

We used ABCTOOLBOX (Wegmann *et al.* 2010) with fast-simcoal 2.5 to compare domestication scenarios. See details in Supporting information online.

Frequency of resistance to sharka

Two loci have been identified as conferring resistance to sharka so far, *PPVres* (Soriano *et al.* 2012) and *MetaPPV1b* (Mariette *et al.* 2016), and markers linked to these loci in cultivated apricots have been identified. We screened these markers in our sample to check whether they were also linked to resistance in wild apricots. PGS 1.21 and PGS 1.24 were used as markers, being microsatellite loci flanking the *PPVres* resistance locus, with alleles specific to resistance in cultivated apricots (Soriano *et al.* 2012), although they are ~141 kb distant from *PPVres*. We also used the SSLP marker ZP002, designed within the best candidate gene for the *PPVres* locus (Zuriaga *et al.* 2013; Decroocq *et al.* 2014). In addition, we developed a new marker, SSRLG1_11 m52, designed within the second locus involved in PPV resistance, *MetaPPV1b* (Mariette *et al.* 2016).

Sharka resistance phenotyping

Fifty wild apricots were chosen at random in our sample for resistance phenotyping, using both visual inspection of symptoms and ELISA measure of virus load. Stems were collected from 2-year-old apricot seedlings and buds were grafted onto 6-month-old GF305 peach rootstocks. Five plants were grafted per sample, of which four were inoculated. The rootstocks were beforehand inoculated by chip-budding using bark pieces from PPV-M20-infected peach tree showing typical symptoms of sharka. Once the scion started growing, both the rootstock and the scion were pruned and the plants were placed in the cold chamber at 5 °C for at least 3 months. Plants without sharka symptoms on the shoots growing from the chip-buds or rootstocks, and with a negative ELISA reaction (enzyme-linked immunosorbent assay) for both the inoculum (infected chip-bud as described above) and rootstock, were eliminated from the final scoring, as they represented failed inoculation treatments.

Responses of the grafted scions to PPV infection were scored for at least two vegetative cycles, starting from the first bud burst after the initial cold treatment. One vegetative cycle corresponds to a combination of 3 months of dormancy in the cold chamber followed by 4 months of vegetative growth in the greenhouse. Symptoms of sharka were visually evaluated on apricot leaves of the rootstocks, the scions and the inoculums if the chip-buds were still growing. ELISAs were performed to check the presence of the virus in leaves of the scions tested for susceptibility to sharka. Two scorings, 3 weeks apart, were performed per vegetative cycle, including symptom observations and ELISA detection of the virus. After each 4-month growth period in the high-confinement greenhouse, plants were placed back in the cold chamber for 3 months to overgo bud dormancy. Pruning was performed at the beginning of each growth period to induce vigorous new shoots used for symptom scoring.

Viral accumulation was estimated for each individual plant from double antibody sandwich ELISAs. Optical densities (OD) were normalized using PPV-M20-infected, 'GF305' indicator plants used as a positive controls on each ELISA plate that was set at 100. Two serological assays were carried out per vegetative cycle, at 4-week interval, and the first test was performed 3 weeks after bud burst. The final viral accumulation value (ACC) is the average of normalized measurements from all PPV-infected replicates of each apricot genotype, over two vegetative cycles. Analyses of the phenotypic data were performed under R v2.15.0 (http://www.R-project.org). Both visual inspection and serological tests were used for phenotypic scoring and

for assessing which plants were resistant, partially resistant or susceptible. Because PPV is a quarantine pathogen and every infected plant has to be eradicated, regardless of symptom expression, we considered as valuable genitors for resistance breeding only the individuals with no viral particle detected by serological tests (normalized OD ≤ negative control).

Results

Genetic variability and population differentiation among cultivated and wild apricots

After having filtered out markers that were linked one to each other or deviated from neutrality, and individuals that were likely clonemates or siblings (see Supporting information), 15 microsatellites markers and 372 apricots, including 230 wild and 142 cultivated trees, were retained for analyses of genetic diversity and population structure (Tables S1 and S2, Supporting information). The apricot sample geographic distribution is depicted in Fig. S1 (Supporting information). The 372 apricot genotypes displayed from six to 22 alleles per marker, and the allelic richness ranged from 2.48 to 3.55 (Table S2, Supporting information). The average observed and expected heterozygosities across markers were $H_0 = 0.72$ (min–max: 0.57–0.81) and $H_E = 0.84$ (min–max: 0.66–0.92), respectively (Table S2, Supporting information).

The change rate in the log-likelihood between successive K values (ΔK) inferred with STRUCTURE revealed three peaks at $K = 2$, $K = 5$ and $K = 9$, respectively (Fig. S2A, B, Supporting information). For $K < 9$, some clusters appeared admixed, while they appeared nonadmixed and well-delimited at $K = 9$ (e.g. the blue Q_3 cluster and the pink W_4 cluster at $K = 9$, Fig. 1A). Further increasing K above 9 did not reveal well-delimited new cluster. This altogether suggested that $K = 9$ corresponded to the most relevant K value for our sampling. At $K = 9$ (Fig. 1A), the cultivated apricots split into four distinct clusters (named Q_1–Q_4) and the wild apricots formed five clusters, distinct from the cultivated ones

(named W_1–W_5). Many wild trees showed footprints of admixture between the W_1, W_2 and W_4 clusters, while little admixture was found among the other clusters (Fig. 1A, Table S3, Supporting information). We used STRUCTURE membership coefficients inferred at $K = 9$ to define the populations used in subsequent analyses. For analyses hereafter, genotypes were assigned to a given population if their membership coefficient for that population were ≥0.80 (Table S3, Supporting information; $n = 196$ individuals, excluding admixed genotypes).

Within cultivated apricots, a west–east geographic differentiation was found across Eurasia. Q_1 included essentially individuals from Europe, Turkey and Azerbaijan (CULT02_EUR, CULT11_TUR and CULT01_AZE), while cultivated apricots from China (CULT03_CHN) split into two clusters, Q_2 and Q_3, the later also including genotypes from southern Central Asia (CULT12_PAK from Pakistan, Turkmenistan, Afghanistan). Q_1 would thus correspond to the Irano-Caucasian ecogeographic group previously defined (Kostina 1969). Q_4 only included cultivated apricot trees from the Almaty Kazakh region that geographically overlaps with the Kostina's Dzhungar-Zailig ecogeographic group. Some individuals appeared as hybrids between wild and cultivated apricots; they mainly belonged to CULT05_KAZ to CULT07_KAZ. They were initially classified as 'cultivated' (Tables 1 and S1, Supporting information) because they had been obviously planted by humans, in germplasm repositories or along roads. Furthermore, STRUCTURE assignment suggests that some recent Central Asian collections (i.e. CULT04_KAZ) and plantations of cultivars in fact correspond to wild genepools, being assigned to W_2 with the wild accessions from Kazakhstan.

Regarding the wild genetic clusters, two main clusters, W_3 and W_5, clearly separated from the other wild clusters already from $K = 5$; W_3 is located in Kazakhstan (Wild03_KAZ, Esik Lake) and W_5 in Kyrgyzstan (Wild15_KGZ, Sary-Chelek national park). At $K = 9$, the remaining wild individuals formed three additional genetic clusters, although showing high levels of admixture. W_1 and W_2 mostly corresponded to Kazakh

Fig. 1 Population structure analysis of the cultivated and wild apricots inferred using the model-based program STRUCTURE from $K = 2$ to $K = 9$. (A) Proportions of ancestry of cultivated and wild *Prunus armeniaca* accessions ($n = 372$) inferred with STRUCTURE for $K = 2$, $K = 3$, $K = 5$, $K = 7$ and $K = 9$. Each individual is represented by a vertical bar, partitioned into coloured segments in proportion of the estimated membership in the different genetic clusters inferred with STRUCTURE. Above the figure are depicted the names of the cultivated and wild apricot sample locations as described in Table 1. The two stars (*) point to the two admixed North American, PPV-resistant cultivars, 'SEO' and 'Stella'. (B) Neighbor-Joining dendrogram based on DICE dissimilarity indices showing the relationships among the nonadmixed 196 apricot individuals (i.e. individuals assigned to one cluster at $K = 9$ with a membership coefficient > 0.80). Genotypes were coloured according to their assignment to the different genetic clusters Q_1–Q_4 and W_1–W_5, as inferred by STRUCTURE. Branch length is proportional to the distance between nodes. (C) Microsatellite distance-based Neighbor-Joining (NJ) trees on apricot populations defined according to STRUCTURE inferences. The left- and right-rooted NJ trees were based on shared allele distance D_{SA} (Chakraborty & Jin 1993) or Nei's D_A distance (Nei *et al.* 1983) and were computed as described in Vercken *et al.* (2010). The bootstrap values above 40% are shown.

(A)

(B)

(C)
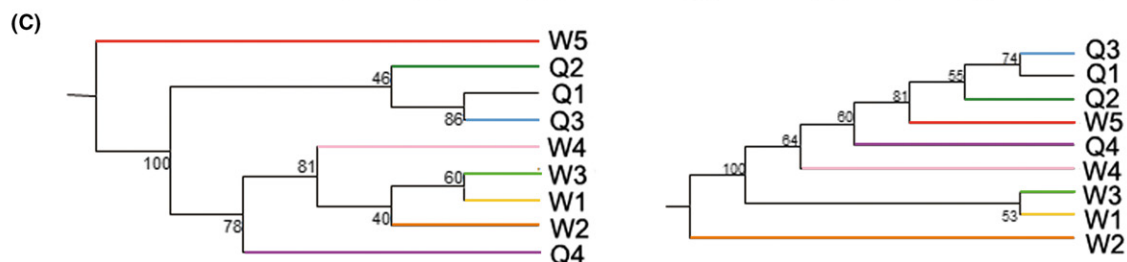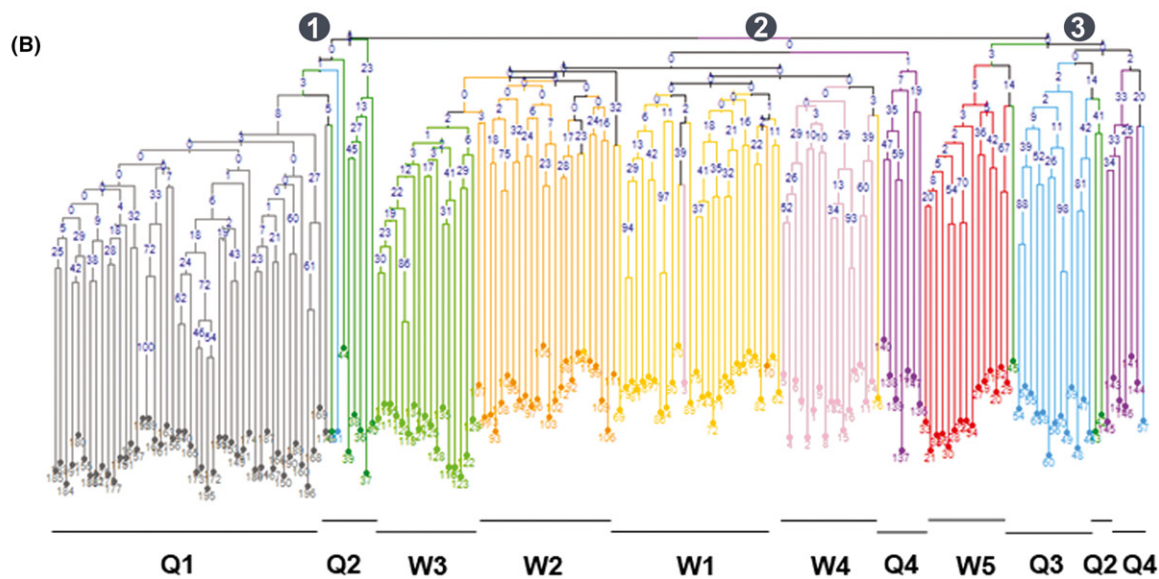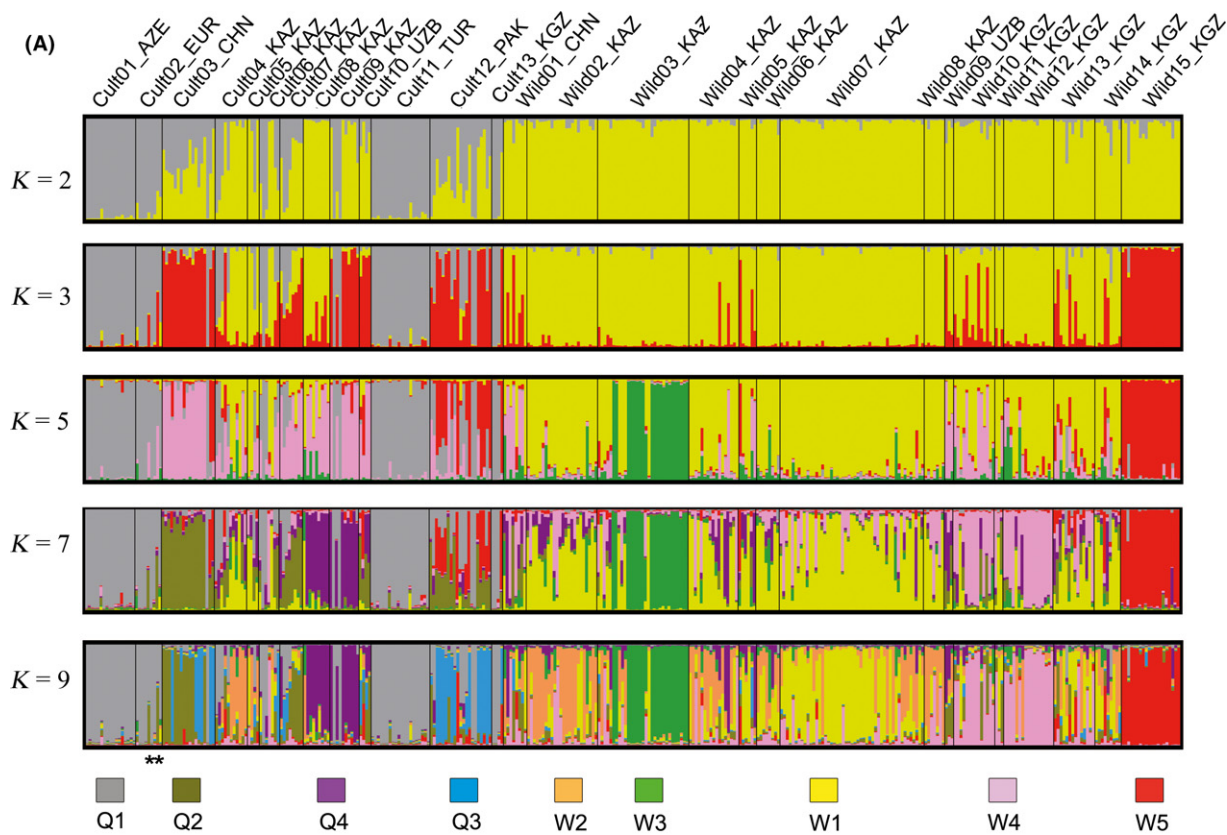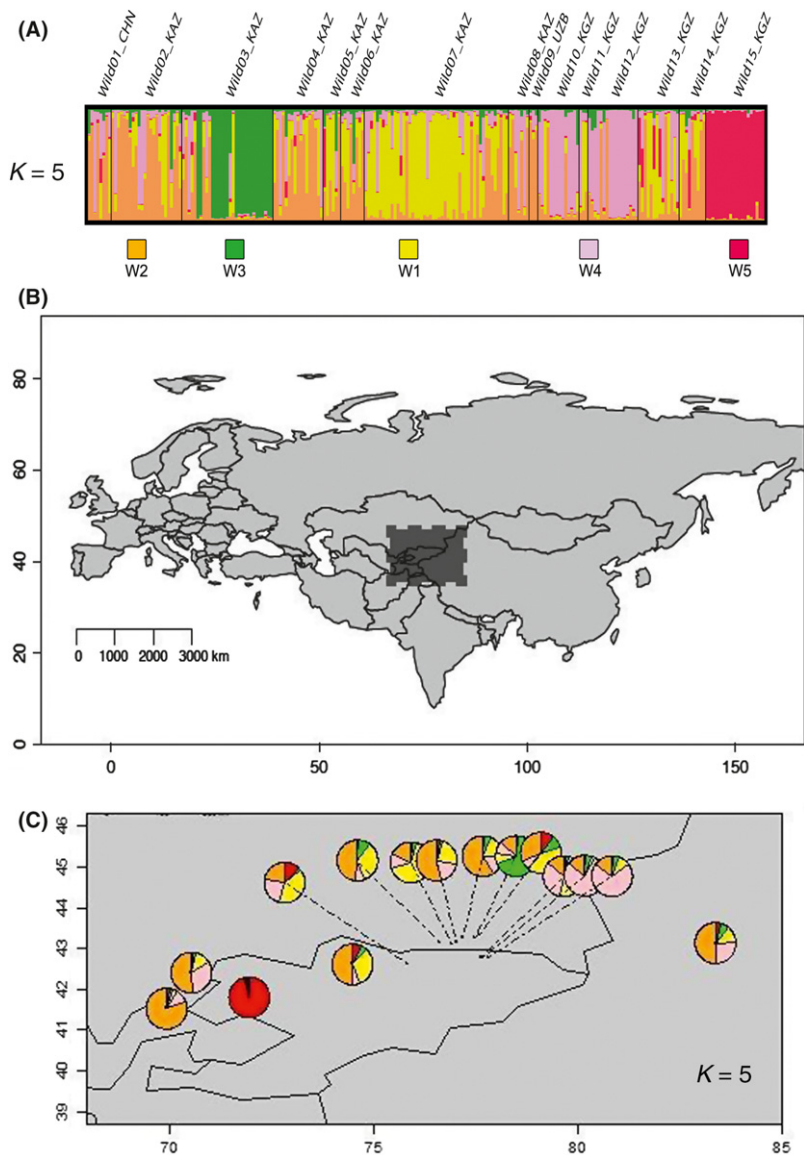
Fig. 2 Spatial population genetic structure of *Prunus armeniaca* natural populations (*n* = 230) inferred by STRUCTURE. (A) Bayesian clustering results obtained with STRUCTURE at *K* = 5. Above the figure are depicted the names of the wild apricot populations as described in Table 1. Each individual is represented by a vertical bar, divided into *K* segments representing the amount of ancestry in its genotype corresponding to *K* clusters. (B) Graphical representation of the sampling area in Central Asia. (C) Visualization of the geographic distribution of the clusters for *K* = 5.

apricot natural populations and W_4 to wild germplasm present in North Kyrgyzstan, around the Issyk-Kol Lake. Altogether, these five wild clusters cover the range of distribution of the Dzhungar-Zailig ecogeographic group (Kostina 1946). Figure 2A shows the distribution of the wild populations based on more detailed analyses (Supporting information). While W_1 (yellow) and W_2 (orange) are spread across the Tien Shan Mountains, the W_4 (pink) formed a narrow spatial cluster restricted to the eastern part of the distribution, and the W_3 (green) cluster corresponds to a single collection site, Wild03_KAZ. The Wild15_KGZ natural population (W_5, red) appears geographically isolated and genetically highly differentiated from the other genetic clusters. No isolation-by-distance pattern was found in wild apricots ($r^2 = 0.06$, $P = 0.35$, Table S4, Supporting information).

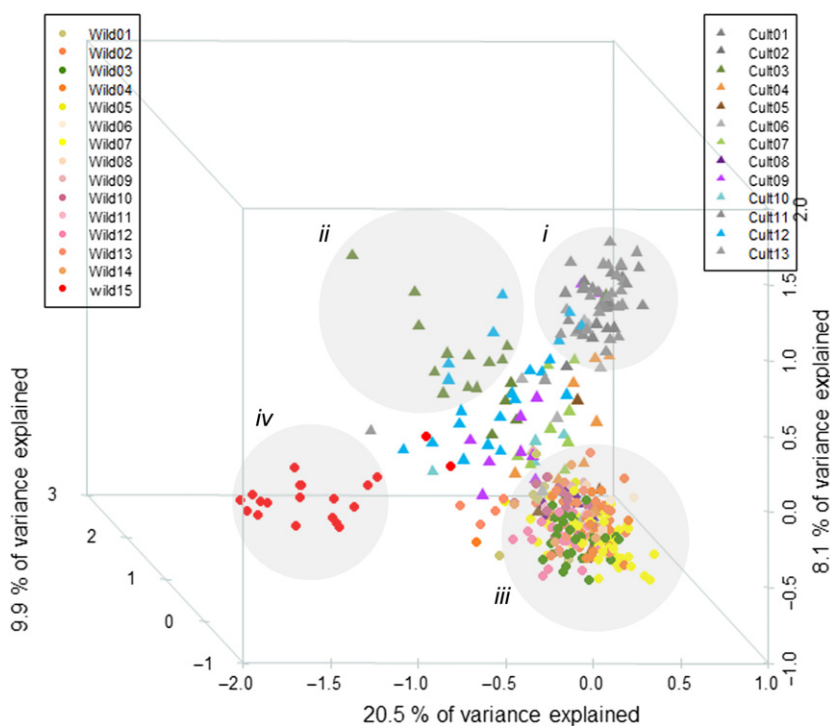*Genetic variation and differentiation among the nine apricot populations*

We built a Neighbor-Joining tree of the nonadmixed individuals (Table S3, Supporting information) based on dissimilarity scores (Fig. 1B). The dendrogram showed three major clades (labelled 1–3 in Fig. 1B). The Q_1 and Q_2 populations formed the first clade, including Irano-Caucasian, European and part of the Chinese cultivated apricot trees. The second clade included the wild populations (number 2 in Fig. 1B, W_1 to W_4), plus a fraction of the Q_4 population from the Almaty Kazakh region. The wild cluster W_5 (Kyrgyzstan) was part of the remaining Chinese and Central Asian cultivars (Q_2–Q_4) in the third clade. Further structure of the dendrogram was also in agreement with the clustering inferred with STRUCTURE, with the exception of Q_2 and Q_4, that each

split into two distinct clades. Dendrograms of populations (Fig. 1C) provided similar patterns of relationships among groups, with Q_1 and Q_2 on one side and W_1–W_4 on the other side.

The FCA (Fig. 3) revealed a similar pattern as inferred with STRUCTURE, with a clear differentiation between eastern and western cultivated apricots. Indeed, part of the Chinese cultivars (Q_2 cluster) was plotted on the top left (green triangles in Fig. 3), apart from the Irano-Caucasian and European/North American apricots, which clustered together on the top right (grey triangles, equivalent to the Q1 cluster). In the FCA, all Dzhungar-Zailig wild apricots (W_1–W_4) clustered together, except for the Sary-Chelek population (W_5 population, red points), that here again clearly separated from all other apricot populations.

We computed population genetic statistics for the nine genetic populations (Table 2). The inbreeding coefficients (F_{IS}) were all low, indicating outcrossing and lack of further strong population subdivision. High genetic diversity was found in the cultivated ($H_E = 0.81 \pm$ standard error SE = 0.09) and wild ($H_E = 0.80 \pm$ SE = 0.09) populations. Genetic diversity was similar across populations, except for Q_2 that showed higher values of allelic richness ($A_R = 8.22 \pm$ SE = 0.46) and private allelic richness ($A_{PR} = 1.39 \pm$ SE = 0.27) than the other populations. Among the wild populations, A_{PR} was higher in the W_5 population ($A_{PR} = 0.39 \pm$ SE = 0.16) than in the other populations.

Pairwise F_{ST} and Jost's D among the nine apricot populations were all significant (Table 3, Fig. S3, Supporting information) and in agreement with the structure found in previous analyses. They indicated a strong differentiation between the Q_1 population (European and Irano-Caucasian cultivated apricots) and all other populations. The W_5 population showed strong differentiation from the other wild or cultivated populations. Differentiation among the other wild populations (W_1–W_4, corresponding to the Dzhungar-Zailig ecogeographic group) was low. Jost's D values further confirmed the strong differentiation between cultivated and wild populations (Table 3), with significantly lower mean Jost's D among W populations (mean $D = 0.35$) than between Q and W populations (mean $D = 0.45$; Student's t-test, $t = 15.07$; $P = 0.045$). The mean D between Q populations was intermediate (mean $D = 0.040$) and not significantly different from either W-W or Q-W means (Student's t-tests, $P > 0.05$).

Testing domestication scenarios

We used an ABC approach for testing different domestication scenarios. The populations simulated in our models were defined based on STRUCTURE inferences at $K = 9$, excluding admixed accessions (CULT04 to CULT07, Fig. 1). We also pooled some populations to limit the number and complexity of scenarios. For the cultivated apricots, we pooled the Q_2–Q_4 populations

Fig. 3 Principal component analysis of the whole apricot data set based on 15 microsatellite markers. Triangles indicate cultivated apricot accessions such as depicted in Table 1, and circles wild apricots collected in natural populations. Axes 1, 2 and 3 explain 20.5%, 8.1% and 9.9% of the total variance, respectively. The right and left legends are depicting the names of the cultivated and wild apricot sample locations as described in Table 1. (i) refers to the Q_1 Irano-Caucasian and European cultivated apricots, (ii) to the Chinese (Q_2) cultivars and landraces; (iii) to the wild Central Asian apricot populations (corresponding roughly to W_1–W_4) and (iv) to the Sary-Chelek (W_5) natural population.

Table 2 Summary statistics of genetic variation among the nine *Prunus armeniaca* populations detected with STRUCTURE

Group	Sample size	H_0	uH_E	F_{IS}	A_R	A_{PR}
Q_1	48	0.652 (0.033)	0.716 (0.035)	0.077	5.15 (0.32)	0.38 (0.17)
Q_2	10	0.732 (0.043)	0.827 (0.030)	0.067	8.22 (0.46)	1.39 (0.27)
Q_3	16	0.683 (0.027)	0.771 (0.022)	0.083	5.86 (0.31)	0.33 (0.10)
Q_4	13	0.754 (0.034)	0.788 (0.023)	−0.001	5.95 (0.42)	0.41 (0.13)
W_1	29	0.703 (0.032)	0.722 (0.035)	0.003	5.38 (0.45)	0.10 (0.05)
W_2	24	0.753 (0.046)	0.763 (0.030)	−0.007	5.92 (0.43)	0.15 (0.07)
W_3	21	0.719 (0.044)	0.713 (0.031)	−0.031	5.03 (0.32)	0.15 (0.10)
W_4	18	0.731 (0.032)	0.738 (0.033)	−0.033	5.46 (0.39)	0.09 (0.04)
W_5	17	0.711 (0.042)	0.741 (0.037)	0.000	5.32 (0.38)	0.39 (0.16)

A_R and A_{PR}, allelic richness and private allele richness averaged across loci, respectively, estimated by rarefaction using a sample size of four; H_0, observed heterozygosity; uH_E, unbiased expected heterozygosity which corresponds to gene diversity corrected for sample size; F_{IS}, inbreeding coefficient; SE, standard error. All values were significant at $P < 0.001$.

Table 3 Pairwise F_{ST} (lower diagonal) and *Jost's D* (upper diagonal) values among the nine apricot populations ($n = 196$) inferred with STRUCTURE

Cluster	Q_1	Q_2	Q_3	Q_4	W_1	W_2	W_3	W_4	W_5
Q_1	—	0.4357	0.3819	0.5338	0.6506	0.5500	0.6935	0.6391	0.6604
Q_2	0.12537	—	0.4036	0.3288	0.3692	0.3158	0.4718	0.3318	0.3408
Q_3	0.12424	0.10050	—	0.3418	0.4313	0.4152	0.4995	0.4943	0.3677
Q_4	0.15797	0.07692	0.08903	—	0.3162	0.2733	0.3951	0.3207	0.3925
W_1	0.20910	0.10226	0.13376	0.09777	—	0.1899	0.2678	0.2058	0.5073
W_2	0.17185	0.07928	0.11873	0.07862	0.06228	—	0.3169	0.2189	0.4424
W_3	0.22235	0.13264	0.15363	0.12396	0.09990	0.10295	—	0.4255	0.4832
W_4	0.20526	0.08747	0.14610	0.09478	0.07016	0.06993	0.14412	—	0.4278

The pairwise F_{ST} values (below the diagonal) were calculated with ARLEQUIN v3.5 as follows: $F_{ST}/(1 − F_{ST})$. All pairwise F_{ST} and *Jost's D* were significant ($P < 0.05$, Number of permutations = 1000).

together because of their weak genetic differentiation (Table 3). Concerning the wild genepool, we simulated the Dzhungar-Zailig ecogeographic group as a single population, pooling all Kazakh and Kirgiz populations together (W_1, W_2, W_3 and W_4), because of their overlapping geographic ranges, weak genetic differentiation and high levels of admixture. For all models of apricot domestication, we therefore assumed four main populations: the cultivated European/Irano-Caucasian population ($Q1_{\_WEST}$), the cultivated Central and eastern Asian group (Q_2, Q_3, Q_4; called $Q_{\_EAST}$), the W_5 wild apricot population and the Dzhungar-Zailig ecogeographic wild group ($W_1 + W_2 + W_3 + W_4$). In our models, we assumed Q_East as the most ancient cultivated group because of its higher level of genetic diversity. We assumed W_5 as the most anciently derived population (i.e. outgroup) because of its strongest genetic differentiation from other populations. In total, we simulated six models of apricot domestication that were designed to distinguish between (i) a single domestication of $Q_{\_EAST}$ from a wild population (either W_5 or $W_{1–4}$, models 1 and 4, respectively, Fig. S5, Supporting information),

followed by a more recent domestication event of $Q_{\_WEST}$ from $Q_{\_EAST}$; (ii) successive domestications of $Q_{\_EAST}$ and $Q_{\_WEST}$ from the same wild population (either W_5 or $W_{1–4}$, models 2 and 5, respectively), (iii) independent domestications of $Q_{\_WEST}$ and $Q_{\_EAST}$ from the two different wild populations ($W_5/W_{1–4}$ or $W_{1–4}/W_5$, respectively, models 3 and 6). In all scenarios, the $Q_{\_EAST}$ and $Q_{\_WEST}$ cultivated groups underwent bottlenecks, whose strength was estimated.

The comparisons of the six scenarios using an ABC framework revealed that the model with the highest posterior probability (P) was the scenario assuming two successive domestications, of $Q_{\_EAST}$ and $Q_{\_WEST}$, respectively, from the same Dzhungar-Zailig group in Asia ($W_{1–4}$), at different times, with rather weak bottlenecks, although stronger western cultivated apricots (Q_1 population) ($P = 0.91$, Fig. 4, Table 4). Overall, the Bayes factor (BF) for scenario 5 against the four others was 10.75 (Table 4), which indicates strong support for this scenario. We checked that the power of the analyses was sufficient to discriminate between the competing models. For scenario 5, against the five other
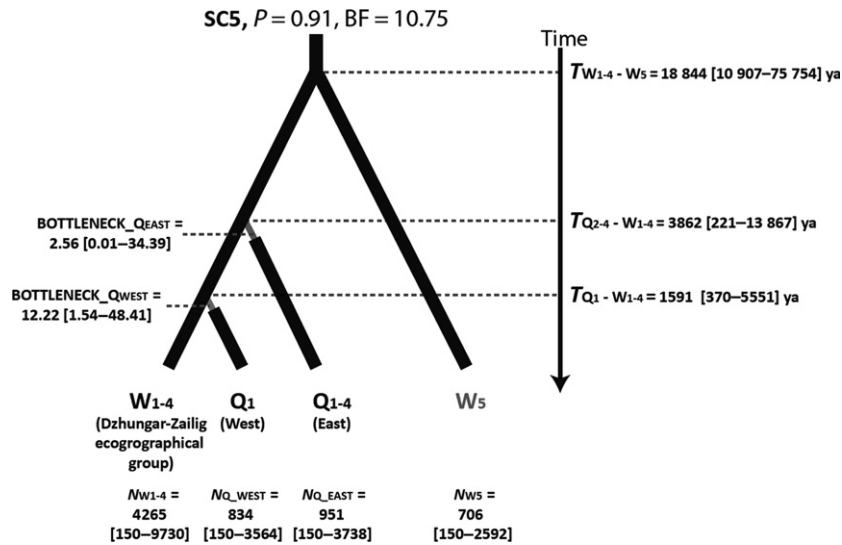
Fig. 4 Most probable scenario of domestication of *Prunus armeniaca* in Eurasia compared by approximate Bayesian computation (i.e. scenario 5: two consecutive independent domestications from the Dzhungar-Zailig ecogeographical group) provided with the model parameter estimates. Each parameter estimate is presented with the 95% confidence interval in brackets. *P*: posterior probability of scenario 5; BF: Bayes factor of scenario 5 vs. the other five scenarios. Details of Bayes factors are summarized in Table 4. T_{X-y}: divergence time between the population x and y, N_x: effective population size of population x; Q_1 (or Q_{WEST}), Q_{2-4} (or Q_{EAST}), W_{1-4}, W_5 referred to populations detected with STRUCTURE (see details in Supporting information online), *Bottleneck_x*: the effective size of the population x one generation earlier in time relative to the effective size of the population x one generation later in time such as $N_T = Bottleneck * N_{T+1}$.

Table 4 Bayes factor (BF) for the six historical models of domestication (i.e. sc1–sc6) assuming different wild groups (i.e. W_{1-4} or/ and W_5) at the origin of the domestication, compared by approximate Bayesian computations

			B					
	Wild group origin	Scenarios	sc1	sc2	sc3	sc4	sc5	sc6
A	W_5	sc1 – Single domestication	—	4.0	0.67	0.22	0.01	0.64
		sc2 – Independent consecutive domestications	0.25	—	0.17	0.05	0.00	0.16
	W_5 and W_{1-4}	sc3 – Independent domestications	1.50	6.0	—	0.32	0.02	0.97
	W_{1-4}	sc4 – Single domestication	4.64	18.6	3.09	—	0.05	2.99
		sc5 – Independent consecutive domestications	**96.17**	**385.7**	**64.05**	**20.74**	—	**62.02**
	W_5 and W_{1-4}	sc6 – Independent domestications	1.55	6.2	1.03	0.33	0.02	—

The models are described in Figs 4 and S5 (Supporting information). Bayes factors are represented for model A vs. model B. Wild group (W_{1-4} and/or W_5) at the origin of domestications and scenario details are presented once in each row; bold values highlight BF of sc5 against other models.

scenarios, the type I error rate was 0.3 and the mean type II error rate was 0.06. Parameter estimates for scenario 5 are provided in Fig. 4. The cross-validation step used to check the accuracy of the marginal posterior distributions estimated using our approach revealed relatively high prediction errors and independent of the tolerance rate (Table S6, Supporting information). These parameter estimates should therefore be regarded with caution. However, the observed values marginally felt within the simulated data ($P = 0.09$), suggesting that the assumed model was capable of reproducing the observed summary statistics. Overall, ABC analyses

thus provided support for two successive domestication events from the wild Dzhungar-Zailig group in Asia, at different times, and with bottlenecks.

Variation of the sharka resistance trait in the wild *Prunus armeniaca* populations

The analyses above revealed high genetic diversity in wild apricots that may therefore harbour valuable genetic resources for breeding, and in particular for disease resistance. For testing this hypothesis, we estimated the frequency of resistance in wild apricots. Fifty

wild apricot trees from seven natural populations were chosen at random for being evaluated for their response to PPV infection following a high-density inoculation (Fig. 5 and Table S8, Supporting information). Susceptibility to PPV was scored by recording symptoms and by semi-quantitative detection of viral accumulation in the grafted scions. The two types of disease scoring estimates were highly correlated (r^2=0.89; $P < 2.2e\text{-}16$) (Fig 5). We observed a high variability in the type of symptoms among the wild apricot individuals when infected with PPV. The Fig. 5-A2 shows the typical

chlorotic symptoms that were recorded both in the wild (UZ 11.5) and the cultivated PPV-susceptible apricots (Manicot and Moniqui cultivars). Atypical symptoms were observed in the four replicates of the genotype CH 128, with chickenpox-like symptoms (Fig. 5-A3).

When considering the accumulation of the virus across two vegetative cycles, we classified as 'resistant', individuals that showed mean optical density scores lower or equal to the negative control (Fig. 5B). Following the results of the PPV-specific serological tests (ELISA), the 50 wild apricot individuals displayed a
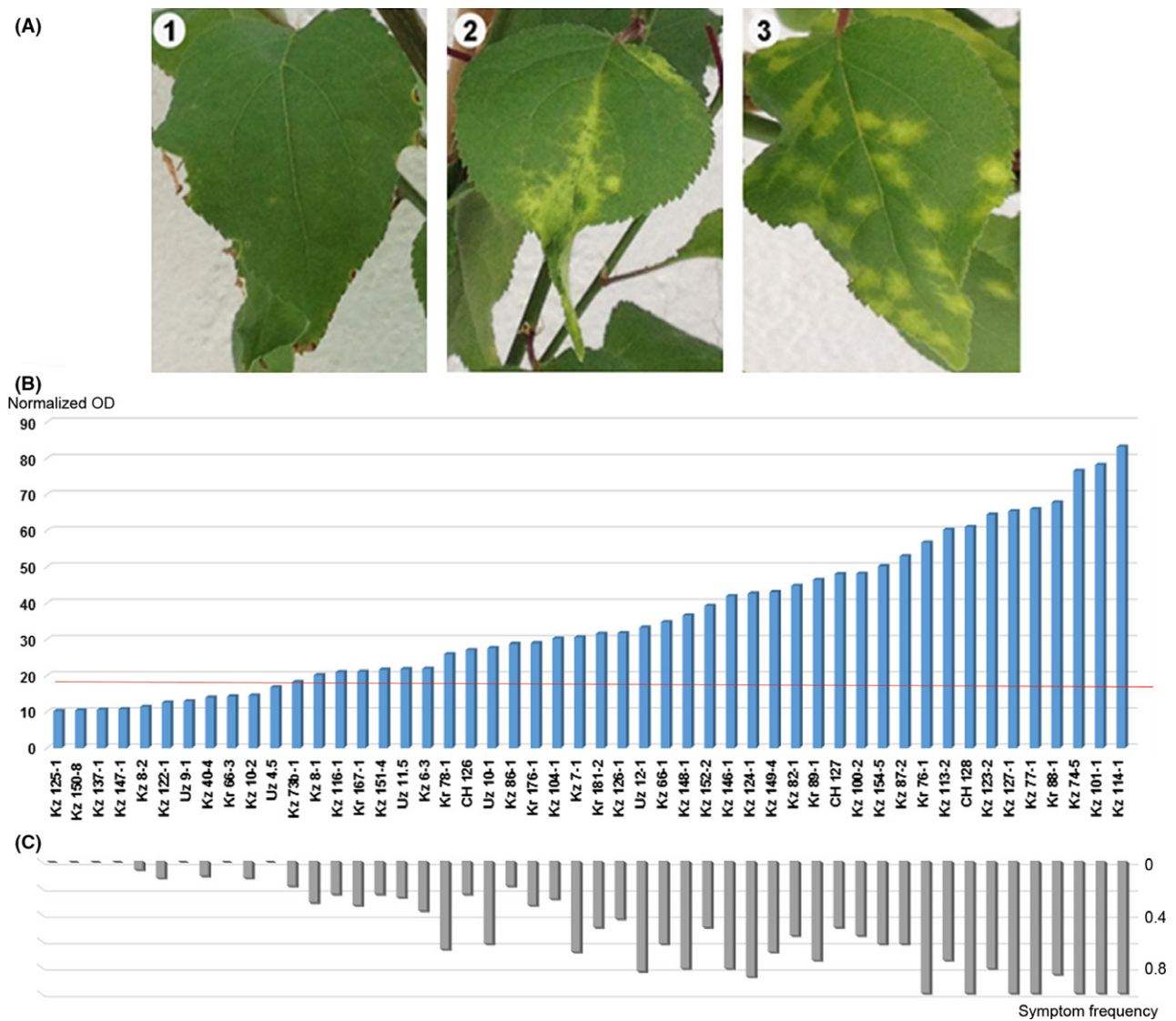
Fig. 5 Plum Pox Virus resistance phenotyping of wild apricot individuals. (A) Symptoms on PPV-resistant (1) or PPV-susceptible (2 and 3) wild apricot individuals ($n = 50$). (B) The graph shows the average of normalized optical density (OD) for four replicates of each wild apricot individual, each being tested four times by ELISA over two vegetative cycles. Accessions are ordered from the lowest to the highest viral accumulation value depicted as the mean OD value. PPV-infected 'GF305' peach indicator that was used as positive control in ELISAs was set at 100. The red line depicts the average optical density of the negative controls (uninfected apricot genotypes) over the two cycles. (C) Frequency of plants showing symptoms over the two vegetative cycles.

continuous phenotypic variation, ranging from high susceptibility to full resistance to PPV. In total, 12 of the 50 tested individuals (24%) were found fully resistant to sharka (Fig. 5B, from KZ 125-1 to KZ 73B-1), showing that wild apricots may in fact harbour valuable genetic resources for controlling diseases.

We then checked whether the markers previously shown to be linked to resistance loci in cultivated apricots were also linked to resistance in wild apricots. This would allow detecting resistance without phenotyping. The most susceptible genotype appeared to lack the resistance-linked alleles for all three *PPVres*-associated markers (KZ 114-1, Fig. 5B and Table S8, Supporting information). However, many other susceptible accessions displayed at least one or two resistance-linked alleles at the *PPVres* or *MetaPPV1b* locus (see, e.g., KZ 127-1 in Table S8, Supporting information). Besides, the KZ 8-2 or KZ 147-1 individuals, that displayed resistance-associated allele at only one *PPVres* marker, were scored as resistant in phenotypic tests. In conclusion, no specific *PPVres* or *MetaPPV1b* haplotype could be associated with PPV resistance or susceptibility in wild apricots (Table S8, Supporting information). These markers are therefore not useful for screening resistance in wild populations.

Discussion

Our study aimed at clarifying the identity of the ecogeographic groups previously described (Kostina 1969) and the relationships between the cultivated and wild compartments, to elucidate the domestication history and to provide insights into the genetic diversity of wild apricots in comparison with common cultivars and ancient, local varieties, in particular regarding the resistance to sharka.

Genetic differentiation between cultivated and wild apricots

Overall, we detected four populations of cultivated apricots and five populations of wild apricots. Clustering analyses clearly differentiated cultivated from wild apricot accessions, especially the western (European/ Irano-Caucasian) cultivars (Q_1) and Central Asian wild forms (W_1–W_4). The Central Asian cultivated genepool (Q_2–Q_4) displayed a higher level of admixture, and partly with the wild compartment.

Based on morphological and location criteria, the existence of three ecogeographic groups (Irano-Caucasian, Central Asian and Dzunghar-Zailig) of apricot has been previously suggested (Kostina 1969). Based on their sampling sites, they correspond to our Q_1, Q_3 and Q_4 genetic clusters of cultivated apricots, respectively. Our study thus confirms, using genetic data, the

relevance of these groups and provides information on their genetic structure, diversity and relationships. Our study further reveals the occurrence of a fourth genetic group (Q_2), in China. This fourth genetic group of Chinese cultivars and landraces (Q_2) was well separated from the wild apricot populations growing in North Xinjiang (China), which grouped instead with the Kazakh wild populations.

Within the wild germplasm sampled in Central Asia, we identified five populations. The W_1 and W_2 populations were largely distributed along the Tien Shan Mountains, while W_3–W_5 were each restricted to a single sampling site, and W_4 was mostly found in the southern range of the Tien Shan Mountains (around the Issyk-kol Lake, Kyrgyzstan). The W_1–W_4 wild populations form a rather homogeneous group that corresponds to the Dzhungar-Zailig ecogeographic group, together with the Q_4 cultivated population. The W_5 population, displaying the highest genetic divergence with the other wild apricot clusters, is located into the Sary-Chelek national park, close to the Fergana valley, several hundreds of kilometres away from the Tien Shan Mountains. Our data indicate that W_5 does not belong to the Dzhungar-Zailig ecogeographic group, but instead to a genetically differentiated genepool that may correspond to the second ecogeographic group, that is the Central Asian group, that Kostina previously described (1969).

Apricot domestication history

Approximate Bayesian computation allowed to distinguish between different plausible scenarios of domestication and indicated that the most likely history was two independent and successive domestication events from the same wild genepool, the Dzhungar-Zailig ecogeographic group, at different times. This is in agreement with the origin and dissemination routes previously proposed for cultivated apricots (Faust *et al.* 1998). The domestication history has been studied in only few other fruit trees (Gaut *et al.* 2015). A single domestication event has been reported in apple trees but with subsequent hybridization with another wild species than the original progenitor (Cornille *et al.* 2012). In contrast, multiple domestication events have been suggested in olive trees and grapes (Breton *et al.* 2009; Myles *et al.* 2011; Besnard *et al.* 2013; Diez *et al.* 2015). In annual plants, several cases of multiple independent domestications have been documented, for instance in barley, common bean, sorghum and rice (Morrell & Clegg 2007; Bitocchi *et al.* 2012; Lin *et al.* 2012; Gaut *et al.* 2015; Poets *et al.* 2015).

Our estimates of dates of domestication had large confidence intervals, but they included the dates

provided by historical and archaeological evidence. Our findings further indicated the occurrence of weak bottlenecks during apricot domestication. ABC analyses indeed pointed to only slight decrease in population size in cultivated apricots. The inferred bottleneck was a bit stronger in western cultivated apricots (Q_1 population), as expected given modern improvement. These results are in agreement with a previous study (Bourguiba *et al.* 2012) that suggested a loss of diversity during domestication and/or during diffusion of apricots from the Irano-Caucasian area to the Mediterranean basin. The ABC estimates showed high confidence intervals but pointed to weaker bottlenecks than previously inferred in cereals (Glemin & Bataillon 2009). Furthermore, we found that the current apricot cultivated genepool retained high genetic diversity compared to wild apricots. Altogether, this is in agreement with findings in other fruit trees, where domestication bottlenecks were found limited. This has been explained by the limited number of generations since domestication due to long juvenile phases, clonal propagation (which restricts even more the number of sexual cycles separating domesticated species from their wild progenitors) and ongoing crop–wild gene flow (Arroyo-Garcia *et al.* 2006; Miller & Gross 2011; Cornille *et al.* 2012).

The Chinese population (Q_2) in particular displays a very high level of genetic variation, characterized by a strikingly higher private allelic richness compared to the eight other populations, both cultivated and wild. Private alleles were found at 10 different loci in the Chinese germplasm, in agreement with previous studies (Zhebentyayeva *et al.* 2003; Pedryc *et al.* 2009). The occurrence of specific alleles within the Chinese germplasm in comparison with the Central Asian germplasm may be due to either the existence of a second centre of domestication of apricot, in China, or to postdomestication hybridization in China with closely related wild species sharing the same ecological niche as *P. armeniaca* L. (i.e. *P. sibirica*, *P. manshurica*, *P. ansu*, *P. mume*). Gene flow between common Chinese cultivated germplasm and its wild relatives is considered to be still occurring, while wild Siberian apricot resources are declining due to land management and deterioration of their natural environment (Li *et al.* 2014; Wang *et al.* 2014). Further research on the Chinese cultivated and wild apricot(s), together with representatives of closely related species (*P. siberica*, *P. mandshurica*, *P. mume* and *P. ansu*), is required to infer the precise diversity and population structure of apricot in China. Our study was based on a dozen of microsatellite markers, but these can give reliable information for deciphering crop centres of origin and dispersal history. Several previous studies indeed showed congruent results when inferring domestication demographic histories with a limited number of microsatellites and then by genome resequencing, for example in maize, peach and grape (Aradhya *et al.* 2003; Grassi *et al.* 2003; Barnaud *et al.* 2006, 2010; Aranzana *et al.* 2010; Myles *et al.* 2011; Hufford *et al.* 2013; Verde *et al.* 2013).

Are the wild apricots from the Dzhungar-Zailig natural populations at the origin of the resistance to sharka?

Our findings using microsatellite markers thus allowed inference on the domestication history of apricots and on its centre of origin in central Asia. However, the origin and diversity of adaptive traits may follow a different evolutionary trajectory from neutral markers. Phenotyping of important agronomic traits, such as disease resistance, is therefore required for identifying valuable genetic resources. This may also provide fundamental knowledge on the co-evolution history between trees and their pathogens. We therefore also investigated the frequency of resistance in wild populations.

Based on markers linked to *PPVres*, the sharka resistance occurring in the North American germplasm has previously been suggested to have a Chinese origin (Zhebentyayeva *et al.* 2008). However, the discrepancy between our genotypic and phenotypic data invalidates the strict association between allelic variation at the *PPVres* locus and resistance to sharka. Indeed, no *PPVres* haplotype was systematically linked to resistance to PPV infection of the wild apricots. Such association breakage between phenotypic resistance and the *PPVres* resistance-associated alleles had already been observed in several breeding progenies and cultivated germplasm (Decroocq *et al.* 2014; Rubio *et al.* 2014). This can be due to recombination events between markers and resistance genes in natural apricot populations and/or to the presence of other factors or genes involved in the mechanism of resistance to sharka. A genomewide association study conducted on the apricot cultivated germplasm recently showed the implication of at least another locus, *MetaPPV1b* (Mariette *et al.* 2016). However, no clear linkage between phenotypic resistance and the *MetaPPV1b* genotype was found here in wild apricots either. These linked markers therefore do not appear useful for investigating resistance frequencies in wild populations or the origin of sharka resistance, and only phenotyping resistance after experimental inoculations appeared reliable.

Our results on phenotypes suggest that resistance to sharka occurs at relatively high frequency in natural populations of the Dzhungar-Zailig ecogeographic group, with an estimated proportion of fully resistant genotypes of 24%. This may sound surprising, as the wild populations in Central Asia were considered so far

to have evolved in the absence of PPV, thought to have originated in the Balkans (Atanasoff 1935). PPV would have been more recently introduced in the Almaty region of the Tien Shan mountain ranges, less than two decades ago (Spiegel *et al.* 2004). According to this hypothesis, apricot would have been domesticated in the absence of PPV, before its transfer to its western Irano-Caucasian and, later, European area of cultivation. This would explain the lack of resistance to PPV in the Irano-Caucasian and European cultivated germplasm (Zhebentyayeva *et al.* 2008). However, this raises questions about the origin of PPV resistance in natural populations and the forces driving apricot diversity and variation in resistance to sharka, in the absence of its pathogen. We can formulate hypotheses, for example that the genetic factors controlling resistance to sharka were linked to other adaptive trait(s), including resistance to other pathogens or/and pests. In fact, the sources of PPV resistance currently used in the European breeding programmes were linked to resistance to high winter chilling (Badenes *et al.* 1996; Zhebentyayeva *et al.* 2008). Another hypothesis is that PPV, or an ancestral form, emerged much earlier than its first description in 1917, not in the Balkans, but within the native area of stone fruit species in Asia. Under such a scenario, apricot and PPV could have co-evolved for long, in natural populations, thereby explaining the existence of PPV resistance traits in the wild compartment. This discovery, in addition to being important for breeding-resistant cultivars, thus suggests that resistance to PPV has a more complex evolutionary history than once thought when information was based only on breeding progenies, raising questions on the role of the wild apricot genepool in the plant–pathogen interactions and its association with other adaptive traits.

Conclusion

In conclusion, our study showed that apricot natural populations in Central Asia, including China, have great and original genetic diversity that can provide valuable germplasm for further breeding. In particular, studying the genetic structure of this germplasm is expected to provide valuable sources of resistance to sharka, a pool for adaptation to changing environments and a theoretical basis for understanding the biodiversity and potential of apricot resources in the area. The high genetic diversity at microsatellite markers in wild populations further suggests that variability for other adaptive traits may be also present in Central Asia. Furthermore, our findings reveal that domestication occurred in apricots at least twice independently, at different times, but from the same wild genetic pool, and with limited bottlenecks. This contributes to our general understanding of domestication processes in fruit trees, that are still little explored for these aspects (Gaut *et al.* 2015).

References

Aradhya MK, Dangl GS, Prins BH *et al.* (2003) Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genetical Research*, **81**, 179–192.

Aranzana M-J, Abbassi E-K, Howad W, Arus P (2010) Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. *BMC Genetics*, **11**, 69.

Arroyo-Garcia R, Ruiz-Garcia L, Bolling L *et al.* (2006) Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp *sativa*) based on chloroplast DNA polymorphisms. *Molecular Ecology*, **15**, 3707–3714.

Atanasoff D (1933) Plum pox. A new virus disease. Faculty of Agriculture and Forestry, University of Sofia Yearbook (1932–1933), Vol **11**, 49–69.

Atanasoff D (1935) Mosaic of stone fruits. *Phytopathology*, **8**, 259–284.

Badenes ML, Asins MJ, Carbonell EA, Glacer G (1996) Genetic diversity in apricot, *Prunus armeniaca*, aimed at improving resistance to plum pox virus. *Plant Breeding*, **115**, 133–139.

Barnaud A, Lacombe T, Doligez A (2006) Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. *Theoretical and Applied Genetics*, **112**, 708–716.

Barnaud A, Laucou V, This P, Lacombe T, Doligez A (2010) Linkage disequilibrium in wild French grapevine, *Vitis vinifera* L. subsp *silvestris*. *Heredity*, **104**, 431–437.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Besnard G, Khadari B, Navascues M *et al.* (2013) The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122833.

Bitocchi E, Nanni L, Bellucci E et al. (2012) Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E788–E796.

Bourguiba H, Audergon JM, Krichen L et al. (2012) Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. *BMC Plant Biology*, **12**, 49.

Breton C, Terral JF, Pinatel C et al. (2009) The origins of the domestication of the olive tree. *Comptes Rendus Biologies*, **332**, 1059–1064.

Chakraborty R, Jin L (1993) Determination of relatedness between individuals using DNA fingerprinting. *Human Biology*, **65**, 875–895.

Cornille A, Gladieux P, Smulders MJM 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**, e1002703.

Decroocq S, Chague A, Lambert P et al. (2014) Selecting with markers linked to the PPVres major QTL is not sufficient to predict resistance to Plum pox virus (PPV) in apricot. *Tree Genetics & Genomes*, **10**, 1161–1170.

Diez CM, Trujillo I, Martinez-Urdiroz N et al. (2015) Olive domestication and diversification in the Mediterranean Basin. *New Phytologist*, **206**, 436–447.

Dondini L, Lain O, Vendramin V et al. (2011) Identification of QTL for resistance to plum pox virus strains M and D in Lito and Harcot apricot cultivars. *Molecular Breeding*, **27**, 289–299.

Dosba F, Denise F, Maison P, Massonié G, Audergon J-M (1991) Plum pox virus resistance of apricot. *Acta Horticulturae*, **293**, 569–579.

Doyle J, Doyle J (1987) A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, **19**, 11–15.

Dzhangaliev AD, Salova TN, Turekhanova PM (2003) The wild fruit and nut plants of Kazakhstan. *Horticultural Reviews*, **29**, 305–371.

EPPO (2014) PQR: EPPO database on quarantine pests (available online). https://www.eppo.int/DATABASES/pqr/pqr.htm.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Faust M, Surányi D, Nyujtó F (1998) Origin and dissemination of apricot. *Horticultural Reviews*, **22**, 225–266.

Gaut BS, Diez CM, Morrell PL (2015) Genomics and the contrasting dynamics of annual and perennial domestication. *Trends in Genetics*, **31**, 709–719.

Gerbault P, Allaby RG, Boivin N et al. (2014) Storytelling and story testing in domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 6159–6164.

Glemin S, Bataillon T (2009) A comparative view of the evolution of grasses under domestication. *New Phytologist*, **183**, 273–290.

Grassi F, Labra M, Imazio S et al. (2003) Evidence of a secondary grapevine domestication centre detected by SSR analysis. *Theoretical and Applied Genetics*, **107**, 1315–1320.

Halász J, Pedryc A, Hegedűs A (2007) Origin and dissemination of the pollen-part mutated SC haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). *New Phytologist*, **176**, 792–803.

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.

He T-M, Chen X-S, Xu Z et al. (2007) Using SSR markers to determine the population genetic structure of wild apricot (*Prunus armeniaca* L.) in the Ily valley of West China. *Genetic Resources and Crop Evolution*, **54**, 563–572.

Hufford MB, Lubinksy P, Pyhaejaervi T et al. (2013) The genomic signature of crop-wild introgression in maize. *PLoS Genetics*, **9**, e1003477.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Janik J (2005) The origins of fruits, fruit growing and fruit breeding. *Plant Breeding Reviws*, **25**, 255–320.

Jost L (2008) GST and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015–4026.

Kostina KF (1946) The origin and evolution of cultivated apricot (in Russian). *Proceedings (Trudi) of the Nikita Botanical Garden*, **24**, 25–31.

Kostina KF (1969) The use of varietal resources of apricots for breeding. *Proceedings (Trudi) of the Nikita Botanical Garden*, **40**, 45–63.

Lambert P, Dicenta F, Rubio M, Audergon J-M (2007) QTL analysis of resistance to sharka disease in the apricot (*Prunus armeniaca* L.) 'Polonais' × 'Stark Early Orange' F1 progeny. *Tree Genetics & Genomes*, **6**, 291–304.

Li M, Zhao Z, Miao X, Zhou J (2014) Genetic diversity and population structure of Siberian apricot (*Prunus sibirica* L.) in China. *International Journal of Molecular Sciences*, **15**, 377–400.

Lin ZW, Li XR, Shannon LM et al. (2012) Parallel domestication of the Shattering1 genes in cereals. *Nature Genetics*, **44**, 720–724.

Lingdi L, Bartholomew B (2003) Armeniaca. In: *Flora of China* (eds Wu CY, Raven PH), pp. 396–401. Science Press and Missouri Botanical Garden Press, Beijing.

Maghuly F, Borroto Fernandez E, Ruthner S, Pedryc A, Laimer M (2005) Microsatellite variability in apricots (*Prunus armeniaca* L.) reflects their geographic origin and breeding history. *Tree Genetics & Genomes*, **1**, 151–165.

Marandel G, Salava J, Abbott A, Candresse T, Decroocq V (2009) Quantitative trait loci meta-analysis of *Plum pox virus* resistance in apricot (*Prunus armeniaca* L.): new insights on the organization and the identification of genomic resistance factors. *Molecular Plant Pathology*, **10**, 347–360.

Mariette S, Wong Jun Tai F, Roch G et al. (2016) Genome wide association analysis links specific genetic factors to a critically important agronomic trait (resistance to sharka) in *Prunus armeniaca* (apricot) germplasm. *The New Phytologist*, **209**, 773–784.

Miller AJ, Gross BL (2011) From forest to field: perennial fruit crop domestication. *American Journal of Botany*, **98**, 1389–1414.

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 3289–3294.

Myles S, Boyko AR, Owens CL 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.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. 2. Gene-frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

Pedryc A, Szabolcs R, Hermán R *et al.* (2009) Genetic diversity of apricot revealed by a set of SSR markers from linkage group G1. *Scientia Horticulturae*, **121**, 19–26.

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. http://darwin.cirad.fr/.

Pilařová P, Marandel G, Decroocq V *et al.* (2010) Quantitative trait analysis of resistance to Plum pox virus in the apricot F1 progeny 'Harlayne' × 'Vestar'. *Tree Genetics and Genomes*, **6**, 467–475.

Poets AM, Fang Z, Clegg MT, Morrell PL (2015) Barley landraces are characterized by geographically heterogeneous genomic origins. *Genome Biology*, **16**, 173.

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

Riaz S, Boursiquot J-M, Dangl G *et al.* (2013) Identification of mildew resistance in wild and cultivated Central Asian grape germplasm. *BMC Plant Biology*, **13**, 149.

Rimbaud L, Dallot S, Gottwald T *et al.* (2015) Sharka epidemiology and worldwide management strategies: learning lessons to optimize disease control in perennial plants. *Annual Review of Phytopathology*, **53**, 357–378.

Romero C, Pedryc A, Muñoz V, Llácer G, Badenes ML (2003) Genetic diversity of different apricot geographical groups determined by SSR markers. *Genome*, **46**, 244–252.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rubio M, Ruiz D, Egea J, Martínez-Gómez P, Dicenta F (2014) Opportunities of marker assisted selection for Plum pox virus resistance in apricot breeding programs. *Tree Genetics and Genomes*, **10**, 513–525.

Soriano J, Domingo M, Zuriaga E *et al.* (2012) Identification of simple sequence repeat markers tightly linked to *plum pox virus* resistance in apricot. *Molecular Breeding*, **30**, 1017–1026.

Spiegel S, Kovalenko EM, Varga A, James D (2004) Detection and partial molecular characterization of two Plum pox virus Isolates from plum and wild apricot in southeast Kazakhstan. *Plant Disease*, **88**, 973–979.

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

Ugurtan Yilmaz K, Gurcan K (2012) Chapter 13: genetic diversity in apricot. In: *Genetic Diversity in Plants* (ed. Caliskan M), pp. 249–270. InTech Europe, Rijeka, Croatia. ISBN: 978-953-51-0185-7. http://www.intechopen.com/books/genetic-diversity-in-plants/genetic-diversity-in-apricot.

Vavilov NI (1951) Phytogeographical basis of plant breeding. In: *The Origin, Variation, Immunity and Breeding of Cultivated Plants*, Vol. **13**, Chester KS (trans), pp. , 13–54. Chronica Botanica, Ronald Press company, New York.

Vera Ruiz E, Soriano J-M, Romero C *et al.* (2011) Narrowing down the apricot Plum pox virus resistance locus and comparative analysis with the peach genome syntenic region. *Molecular Plant Pathology*, **12**, 535–547.

Vercken E, Fontaine MC, Gladieux P *et al.* (2010) Glacial refugia in pathogens: European genetic structure of anther smut pathogens on *Silene latifolia* and *Silene dioica*. *PLos Pathogens*, **6**, e1001229.

Verde I, Abbott AG, Scalabrin S *et al.* (2013) The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics*, **45**, 487–494.

Wang Z, Kang M, Liu H *et al.* (2014) High-level genetic diversity and complex population structure of Siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers. *PLoS One*, **9**, e87381.

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, **11**, 116.

Yuan Z, Chen X, He T-M *et al.* (2007) Population genetic structure in apricot (*Prunus armeniaca* L.) cultivars revealed by fluorescent-AFLP markers in Southern Xinjiang, China. *Journal of Genetics and Genomics*, **34**, 1037–1047.

Zhang Q-P, Liu D-C, Liu S *et al.* (2014) Genetic diversity and relationship of common apricot (*Prunus armeniaca* L.) in China based on simple sequence repeat (SSR) markers. *Genetic Resources and Crop Evolution*, **61**, 357–368.

Zhebentyayeva TN, Reighard GL, Gorina VM, Abbott AG (2003) Simple sequence repeat (SSR) analysis for assessment of genetic variability in apricot germplasm. *Theoretical and Applied Genetics*, **106**, 435–444.

Zhebentyayeva TN, Reighard GL, Lalli DA *et al.* (2008) Origin of resistance to plum pox virus in Apricot: what new AFLP and targeted SSR data analyses tell. *Tree Genetics and Genomes*, **4**, 403–417.

Zuriaga E, Soriano JM, Zhebentyayeva T *et al.* (2013) Genomic analysis reveals MATH gene(s) as candidate(s) for Plum pox virus (PPV) resistance in apricot (*Prunus armeniaca* L.). *Molecular Plant Pathology*, **14**, 663–677.

Data accessibility

Files with microsatellite genotypes and input data for Structure are available on Data Dryad: DOI: 10.5061/dryad.93633.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Geographic origins of *Prunus armeniaca* accessions used in this study (*n* = 372, including 230 wild and 142 cultivated trees).

Fig. S2 Estimated number of populations in *Prunus armeniaca* from STRUCTURE analyses using the Ln(PD) and Δ*K*.

Fig. S3 Matrix of pairwise F_{ST} in *Prunus armeniaca* among A) all 28 cultivated groups and wild sites, B) the nine populations defined with STRUCTURE.

Fig. S4 Bayesian clustering results for the wild *Prunus ameniaca* sample (*n* = 240) in Central Asia, obtained with STRUCTURE from *K* = 2 to *K* = 7.

Fig. S5 The six scenarios of apricot domestication compared using approximate Bayesian computation.

Table S1 Description of the *Prunus armeniaca* cultivated and wild accessions used in this study, with their geographic origin and, when applicable, provider.

Table S2 Descriptive genetic diversity estimates for the 15 microsatellite markers and the markers associated with the sharka resistance *PPVres* locus in the whole apricot data set.

Table S3 Percentage and counting of individuals assigned to each population defined using STRUCTURE for $2 \leq K \leq 9$ and $2 \leq K \leq 7$ for the whole (*n* = 372 including cultivated and wild trees) and the wild (*n* = 230 including only trees from natural populations) data sets, respectively, based on the 15 microsatellite marker data set.

Table S4 Pairwise estimates of genetic ($F_{ST}/(1 - F_{ST})$, lower diagonal) and geographic distances (in kilometres, upper diagonal) between wild populations previously defined with STRUCTURE.

Table S5 Prior distributions used in approximate Bayesian computations.

Table S6 Prediction errors for each model parameter (i.e. scenario 5, two independent domestications from W_{1-4}) for each tolerance rate (0.01%, 0.1%, 1%).

Table S7 Frequency among the nine apricot clusters of *PPVres* alleles putatively linked to PPV resistance.

Table S8 Genotypes at the *PPVres* and *MetaPPV1b* loci of wild apricot accessions tested for resistance to sharka.

Appendix S1 Supplementary text describing the methods used for microsatellite marker analysis, filtering, spatial population structure and Approximate Bayesian Computation (ABC).