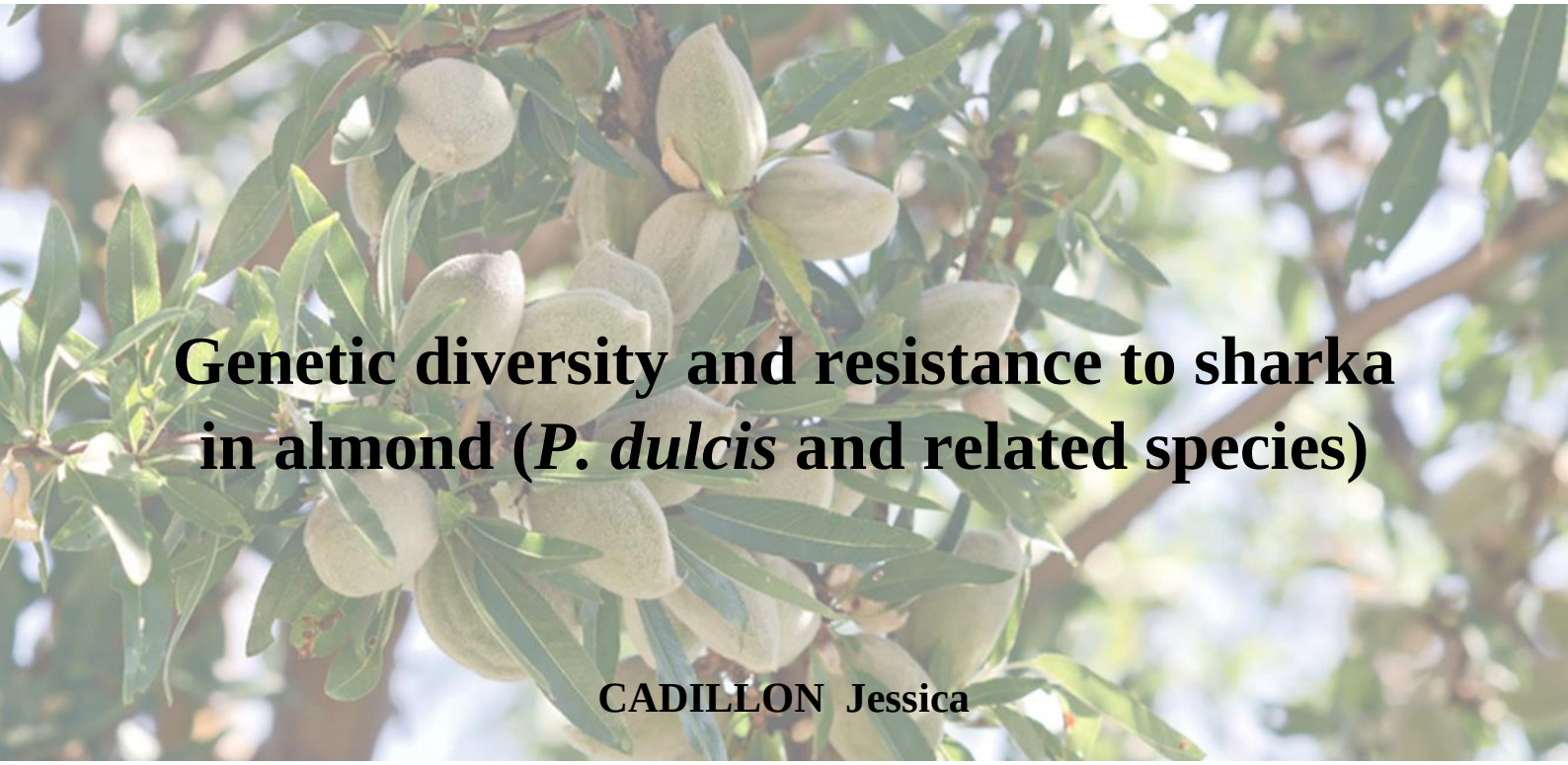
Report for Master's Degree



Genetic diversity and resistance to sharka in almond (*P. dulcis* and related species)

CADILLON Jessica

Received : 4th of June 2021

Internship supervisor : Véronique DECROOCQ

University of Bordeaux, Plant Biology and Agrosciences Master, Plant Breeding, Session 2020-2021

INRAE of Bordeaux, UMR 1332, Plant Virology team

Acknowledgments

I would like to thank Aurélie CHAGUE for preparing all the samples from Bordeaux, but also for her advice in the laboratory and her good mood.

I would also like to thank Stéphane DECROOCQ for guiding me on the use of the different software, for his advice and his listening.

I thank David TRICON for giving me all the information concerning PPV-resistant individuals and for all the conversations leading to hypotheses.

I thank Henri DUVAL and Naïma DLALAH for genotyping the almond trees.

Shuo LIU for the help he gave me for the Factorial Correspondence Analyses.

A big thank you to all the Virology team who once again welcomed me with such kindness and generosity.

Thank you to the girls in my office, for their support and exchanges that we have during these last months.

Finally, I would like to thank Véronique DECROOCQ in particular for giving me the opportunity to contribute to this work, for being present throughout my internship and for giving me guidance. I would like to thank her for her support and for all the work she did in proofreading and correcting this report.

This internship was supported by the GIS Fruits and the PRIMA FREECLIMB project. It benefited from data kindly provided by INRAE GAFL (Naima Dlalah for the SSR data, Christophe Tuero for the PPV resistance tests) and the participation of Centre d'Expérimentation de Pépinières and Montpellier SupAgro (Dr Véronique Marie-Jeanne). The peach x almond genotyping data was prepared by Stéphane Decroocq (INRAE UMR BFP*).*

Presentation of INRAe

The National Research Institute for Agriculture, Food and the Environment (INRAE) resulting from the merger between INRA, Institut national de la recherche agronomique, and Irstea, Institut national de recherches en sciences et technologies pour l'environnement et l'agriculture, is a public research establishment headed by the Ministry of Research and the Ministry of Agriculture. Numerous innovative research projects are conducted and finalized in the fields of plant sciences, ecology and environment in response to various societal challenges (crop production, food security, plant and animal health etc...). There are 18 research centers affiliated to 14 scientific divisions in different thematic fields located in France and many European projects are conducted there.

I performed my internship at the Nouvelle-Aquitaine INRAE research center within the Fruit Biology and Pathology UMR located on the Grande Ferrade site in Villenave d'Ornon. The UMR 1332 is the result of a partnership between INRAE and the University of Bordeaux. Various studies in the field of plant sciences are conducted there. More precisely, I did my internship in the Virology team where the main topic is the deciphering of viral diseases and the interaction between plants and RNA viruses. Their different skills in terms of plant virology, genetics and genomics, cell biology, biochemistry and bioinformatics allow them to better understand the diversity and functioning of these viruses and to develop strategies to control them.

One of the main axes of this team is the study of Plum Pox Virus (PPV) that infects stone fruit species. The aim is to acquire new knowledge about resistance and susceptibility to this virus and to develop biotechnological and genetic tools for the sustainable control of a viral disease.

Table of contents

List of abbreviations

AZE = samples collected from Azerbaijan

ESP = samples collected from Spain

FRA = samples collected from France

GRC = samples collected from Greece

HFR = French hybrids

ISR = samples collected from Israel

ITA = samples collected from Italy

KGZ = samples collected from Kazakhstan

PRT = samples collected from Portugal

SUN = samples collected from Russia

TKM = samples collected from Turkmenistan

TUN = samples collected from Tunisia

TUR = samples collected from Turkey

USA = samples collected from United States

UZB = samples collected from Uzbekistan

ELISA = Enzyme-linked immunosorbent assay

A_r = allelic richness

A_p = private allelic richness

N = number of samples

N_a = number of different alleles

N_e = number of effective alleles

H_o = observed heterozygosity

uH_e= unbiased expected heterozygosity

D_{st} = Nei's standard genetic distance

Jost's D = estimate of differentiation

MCMC = Monte Carlo Markov Chain

NJ = Neighbour-joining

I = Shannon's index

F = Fixation index

cM = Centimorgan

MQM = Multiple QTL Mapping

rMQM = Restricted Multiple QTL Mapping

QTL = Quantitative trait loci

Introduction

Almond (*P. dulcis* (Mill.) D.A.Webb) belongs to the genus Prunus of the Rosaceae family and Prunoideae subfamily (Figure 1). More specifically, among the Prunus species, the subgenus Amygdalus is a monophyletic group of approximately 24 species (Yazbek and Oh, 2013) which occur as trees or shrubs in Asia, Caucasia and around the Mediterranean basin. They grow naturally in temperate forests, dry habitats, but also in mountainous areas between 1,000 and 2,500 m altitude. The subgenus Amygdalus includes two sections: Persica (with the peach and peach-related species) and Amygdalus (with the almond and almond-related species) (Figure 1).

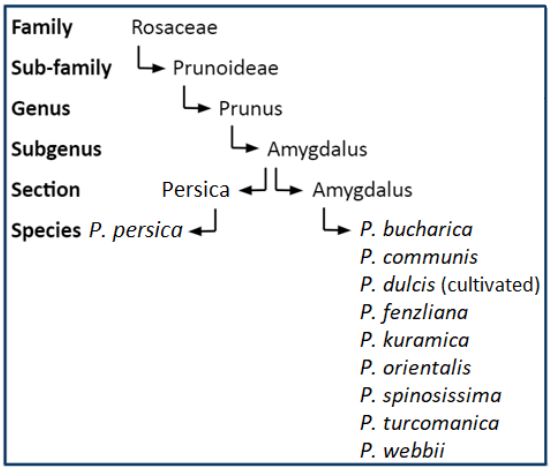


Figure 1 : Classification of the subgenus Amygdalus (For more details about the species see the article Yazbek and Oh, 2013)

Among the species of the Amygdalus section, *P. dulcis* is the domesticated form which is mainly cultivated for its nuts and commercialized throughout the world. Native from the arid mountainous regions of Central Asia, it currently has an important economic share (Browicz and Zohary, 1996). Various studies concerning the domestication of almond have been carried out, however the origin of almonds is still controversial. Different hypotheses have been put forward concerning its domestication and its closest relative(s). One hypothesis is that almonds derived from wild forms of *P. dulcis* which were abundant in the levantine countries (Browicz and Zohary, 1996). There would thus have been a Middle-East domestication event (Browicz and Zohary, 1996). However, other authors argue that the forms of *P. dulcis* that grow spontaneously in the Middle East are wild types that have escaped cultivation. This would suggest that domestication occurred much earlier in Central Asia, where the greatest number of wild relatives of almonds and the greatest number of natural hybrids are still found (Ladizinsky, 1999). In contrast, another theory suggests that almond domestication occurred in a diffuse, protracted manner rather than a localized, single event, involving recurrent genetic exchanges of domesticated forms with local wild relatives (Delplancke et al., 2013).

Previous publications claim the existence of almond wild relatives around the Mediterranean basin, especially in Eastern Anatolia (Zohary, 1996). Most particularly, populations of *P. orientalis* and *P. turcomanica* were described in the Southeast Anatolia region of Turkey (Ak et al, 1999), as well as *P. arabica* forms (Kose et al, 2014). Meanwhile, natural populations of *P. fenzliana* were identified at the Eastern border of Turkey (Naxçivan, autonomous province of Azerbaijan, V. Decroocq pers. comm.). Adaptation to severe environments of Amygdalus species was possible because of their

genetic and associated developmental/physiological diversity promoted by their typically self-sterile, yet inter-specific fertile compatibility. Interestingly, a number of resistance sources have been discovered in almond and almond-related related species, regarding sharka, powdery mildew, green peach aphid, brown rot, root-knot nematodes (Esmenjaud et al., 2009; Lino et al., 2016; Pascal et al., 2017; Pascal et al., 2002(a); Pascal et al., 2002(b); Rubio et al., 2003). In comparison with almond, peach (from the Persica section) is characterized by a low level of genetic variability caused by its self-pollinating behavior and by the small number of founders used in early modern breeding projects which limits the improvement of this crop (Aranzana et al. 2010). Nevertheless, peach and almond are inter-fertile thus providing a mean of genetic improvement, in particular for the transfer of pest and pathogen resistances from almond to peach. In the frame of my internship, we used resistance to sharka as a proof-of-concept for the introgression of valuable agronomic trait in peach cultivars.

Sharka is a viral disease affecting all species of the genus Prunus, from ornementals to fruit trees (Salava J. et al., 2013; James and Thompson, 2006). The causing agent, Plum pox virus (PPV), was first observed on plum trees in Bulgaria in 1917 (Atanassof, 1932). It affects fruit quality as well as productivity. It is one of the most detrimental stone fruit diseases in Europe (Roy and Smith, 1994). It causes strong agronomic losses, but also economic losses (Cambra et al., 2006). Following its discovery, the virus has progressively spread in most of European countries (Németh, 1986), around the Mediterranean basin (Roy and Smith, 1994), in the Near and Middle East, but also worldwide. More recently, it has been discovered in North America (Gildow et al., 2004). At least 10 monophyletic strains of PPV (D, M, EA, C, Rec, W, T, CR, An, and CV) have been identified to date (García et al., 2014). Eradication measures are implemented in some countries but they are costly (Cambra et al. 2006) and have not been able to eliminate the virus. Indeed, despite the establishment of quarantine regulations, sharka spread in most of the stone fruit producing countries thus impairing the international trade of plant material (Garcia et al., 2014). The virus is transmitted by grafting and by aphids, in nurseries and orchards (Lowery et al., 2015). Strategies that aim to reduce the number of aphids by the heavy use of pesticides are counter-productive. As a result, the use of certified plant material and the implementation of quarantine measures have been shown as the best solutions, so far, to limit PPV spread (James, 2017). Sharka is characterized by different symptoms (Figure 2) that can be observed on leaves, fruits, flowers, pits but also bark (Garcia et al. 2014), with more or less latency (Schneider

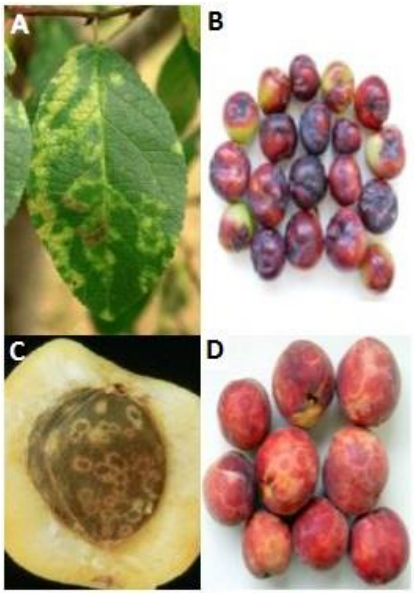


Figure 2 : Symptoms induced by Plum Pox Virus on (A) a domestic plum leaf, (B) domestic plum fruits, (C) apricot stone and (D) peach fruits according to Garcia et al. (2014)

et al. 2011). On leaves, ring-shaped chlorotic discolorations are visible from spring to early summer (Celetti and al., 2002). Infected fruits can show discoloration but also deformation, depending on the species. All these symptoms can vary greatly between species, varieties but also depending on the viral strain (Levy et al., 2000).

Natural resistance to sharka has been shown to occur in apricot (Karayiannis and Mainou, 1994) and in almond (Pascal et al, 2002a). While initially almond was thought to be a non-host for PPV (Németh, 1994), almond susceptibility depends on the strain tested (Dallot et al., 1997) and on the genotype (Pascal et al., 2002a). Indeed, PPV-D has a low pathogenicity compared to PPV-M (Dicenta et al., 2003) whereas testing with PPV-M allowed identifying resistance in ten almond cultivars (cv. Ferragnès, Ardéchoise, Lauranne, Del Cid, Genco, Rumbeta, Mono, Ferraduel, Us179, Az19-4) (Pascal et al., 2002; David Tricon, pers. comm.) and four *P. fenzliana* accessions (cv. Az220-12, Az203-5, Az210-6, Az211-10) (David Tricon, pers. comm.). This raises the question of the origin of resistance to sharka within the Amygdalus germplasm and the true number of different sources of resistance.

In this context, the aim of my internship was three-fold: (1) to characterize the genetic diversity and the relationship between cultivated almonds and its wild relatives, (2) to question a multiple or a single origin of resistance to sharka in almonds and (3) to initiate the identification of the genetic determinants controlling resistance to sharka in one of the PPV resistant almond cultivars, 'Del Cid'.

Material and methods

Plant material

The collection of cultivated almonds analysed in this study is a combination of cultivars maintained at INRAE Bordeaux, INRAE Avignon, IRTA Spain, HAO Greece and CU Turkey. This final collection of *P. dulcis* consists of 156 samples from fourteen countries (see annex Table S1). Part of the cultivated samples and wild relatives were collected between 2008 and 2012 in Azerbaijan, Kyrgyzstan and Turkey. The Turkish samples were collected in Eastern Anatolia along the Euphrates river (Karasu river in Turkey) and on the Akdamar Island of the Van Lake. In total, we have 209 samples which were genotyped by Naima Dlalah at INRAE Avignon. The samples included representatives of different species of the Amygdalus section which correspond to *P. dulcis* (*N*=158) (cultivated almonds) and for the wild relatives, *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). *P. dulcis*, cultivated almonds, are native to different regions: Central Asia (Kyrgyzstan, Turkmenistan and Uzbekistan, *N*=8), Caucasia (Azerbaijan and Turkey,

N=43), Russia (*N*=2), Middle East (Israel, *N*=1), Europe (France, Spain, Greece, Italy and Portugal, *N*=78), North Africa (Tunisia, *N*=1) and North America (United States, *N*=16). They also include nine breeding accessions coming from INRAE breeding programs. Concerning the wild species, *P. communis*, *P. orientalis* and *P. turcomanica* are native from Turkey. P. *fenzliana* from Azerbaijan. *P. kuramica* from Pakistan. *P. pendunculata* from Mongolia and *P. spinosissima* from Kyrgyzstan. *P. bucharica* and *P. webbii* were provided by the ARS-USDA repository (Davis, USA) but their origin remains unknown.

DNA extraction

Genomic DNA was extracted from leaves stored at -80°C. Each sample was ground with 3 mL of buffer 1 (1M Tris-HCl pH 8; 0.5M EDTA; 5M NaCl; completed with water up to 3mL and 0.0117 g of sodium metabisulfite). The amount of buffer was adjusted according to the amount of plant material. As many samples had to be extracted, the extraction was performed in 96-well plates. After grinding the samples, 400 µL of each sample was taken and placed in the plate, followed by 1 min centrifugation at 2,000 RPM. A Falcon® tube containing buffer 2 (0.82g of Hexadecyltrimethylammonium bromide; 5M NaCl; 0.5M EDTA; 1M Tris HCl; completed with water up to 42mL) was placed in the oven at 65°C. In the same Falcon® tube, 1 mL of RNAse was added, followed by shaking. 400 µL of buffer 2 was then placed in the plates containing the grounded material. Followed by 1 min centrifugation at 2,000 RPM and after removing the caps from the tubes, the samples were placed in the oven at 65°C for 20 min. The caps were then replaced and the samples were again placed at 65°C for 40 min. It was important to mix regularly. After 25 min of centrifugation at 3,000 RPM, 400 µL of supernatant from each sample was removed and placed in a new plate. 400 µL of isoamyl alcohol:chloroform was then added followed by a centrifugation of 25 min at 3,000 RPM. 200 µL of supernatant was removed and placed in a new plate. To precipitate the DNA, 200 µL of isopropanol was added and the plate was placed at -20°C for 25 min. The DNA pellet was retrieved by centrifugation 25 minutes at 4°C and 5,600 RPM. Then washed with 300 µL of 70% ethanol, followed by centrifugation with the closed plate for 1 min at 5,600 RPM and a further 20 min centrifugation at 5,600 RPM was then performed this time with the open plate. The pellets were then dried for 2 h in a laminar flow hoods and each pellet was resuspended in 100 µL of 0.1 X TE. The presence and the quality of genomic DNA were verified on 0.8% agarose gel.

Microsatellites markers

Twenty-five microsatellite markers distributed along the eight chromosomes of *P. dulcis*, *P. persica* and *P. salicina* were used to genotype the 209 accessions (see Table S2 in annex). Because some markers did not amplify well, the markers with more than 15% missing data were removed.

SSR genotyping

The INRAE GAFL team genotyped the almond and almond-related samples on an ABI sequencing equipment (Montpellier SupAgro platform). 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 for each marker were checked and adjusted. In addition, each sample and each marker were checked one by one.

Genetic variation and differentiation

The number of different alleles (*Na*), the number of effective alleles (*Ne*), the observed heterozygosity (*Ho*), the expected heterozygosity (*He*) and the fixation index (*F*) were calculated thanks to GenAlEx v6.503 (Peakall and Smouse, 2012), an add-in of Excel. The significance of pairwise genetic differentiation was estimated using *Jost' D* also implemented in GenAlEx. The allelic richness (A_r) and the private allelic richness (A_p) were calculated after adjustment for sample size (set to 4 for the *P. dulcis* analysis and to 5 for the *P. dulcis* and wild related species) differences among groups through the rarefaction procedure implemented in ADZE Allelic Diversity Analyzer v1.0 (Szpiech et al., 2008).

Analysis of population subdivision

From the genotyping data, clonemates present in the dataset were determined using GenoDive v.3.04. (Meirmans, 2020) with a threshold of 5 to avoid biases in allele frequency. One individual per pair identified as clone or siblings was retained.

In order to describe population subdivision, the software STRUCTURE v.2.3.4. (Porras-Hurtado et al., 2013) with no a priori information on the population was used. The clustering method implemented in STRUCTURE is based on Monte Carlo Markov Chain (MCMC) simulations and is used to deduce the proportion of ancestry of genotypes in *K* distinct clusters. Firstly, an analysis with only cultivated almond trees (*P. dulcis*, *N*=138) was performed, followed by a second analysis that included cultivated almond trees and related species (*P. bucharica*, *P. dulcis*, *P. kuramica*, *P. fenzliana*, *P. orientalis*, *P. spinosissima*, *P. turcomanica*, *P. webbii*, *N*=186). For each *K*, STRUCTURE runs consisted of 10 replicates of 10,000 burnin steps followed by 100,000 MCMC iterations. The different analyses were conducted with an admixture model due to the fact that individuals may have admixed ancestry. The analyses were performed with correlated allele frequencies of 1 to 10. Then, the resulting matrices of cluster membership coefficients were permuted with CLUMPP v.1.1.2 (Jakobsson et al., 2007) and the barplots were obtained with DISTRUCT v.1.1 (Rosenberg et al., 2003). The ΔK were 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 is not reflecting the complete subdivision observed in STRUCTURE barplots. It is therefore necessary to compare the obtained barplots in order to choose the *K* value for which all clusters have well assigned individuals.

We considered an individual assigned to a cluster when its assignment probability was ≥90% to that cluster. All individuals with a value below 0.90 (90%) are considered as admixed (hybrids between at least two genetic clusters). In order to better visualize the population structure from a spatial point of view, a factorial correspondence analyses (FCAs) was done with individuals having assignment probability ≥90% with the software GENETIX v.4.05 (Belkhir et al., 2004) and then visualized with "scatterplot3d" R package (Ligges and Mächler, 2002). With the software DARWIN v.6 (Perrier and Jacquemoud-Collet, 2006), genetic differentiation and relationships were estimated using a weighted neighbour-joining tree and 30,000 bootstraps.

Sharka resistance phenotyping

Each 'Del Cid' x 'Honey Blaze' progeny was grafted in 2-3 replicates on four PPV-M infected rootstocks. Eighty-nine F2 individuals were tested, together with the F1 parent and the 'Del Cid' and 'Honey Blaze' grand-parents. Response to PPV infection was scored over two vegetative cycles, with the first one in Spring 2020 and the second one, in Automn 2020. However, because of COVID pandemic issue, we finally gathered reliable data only for the second cycle. Observations of symptoms and ELISA tests were performed twice at three-week intervals, with the first scoring starting 3 weeks after the exit from the cold room (Decroocq et al., 2016). Serological tests (ELISA) were used to measure viral accumulation in both the scion and the rootstock. These tests were performed using the double antibody sandwich (DAS) ELISA technique. The optical density (OD) was measured and normalized using PPV-M-infected, 'GF305' indicator plants used as a positive control on each ELISA plate. The first ELISA assay was performed after bud burst. Two tests were performed at three week intervals per growing season and scores were given according to the degree of infection observed on the plants. Phenotypic observations and serological tests were used to determine which plants were resistant, partially resistant or susceptible to Plum Pox Virus.

Genotyping of the offspring of the interspecific crossing 'Del Cid' x 'Honey Blaze'

A cross between the PPV resistant almond tree 'Del Cid' and the susceptible peach tree 'Honey Blaze' resulted in two F2 populations issued from the self-pollination of two distinct, PPV-resistant F1 individuals (F1(0) and F1(00), respectively). 95 individuals coming from the self-pollination of F1(00) were sequenced by ILLUMINA technology ("SWAG" France Génomique project) and used in the current study for linkage mapping of the resistance factors.

Processing of sequenced reads

The quality of the reads obtained after sequencing was checked with the FastP program (Chen et al., 2018). The sequences obtained were then aligned, in parallel, on the peach reference genome 'Lovell', and the almond 'Texas' genome using the software package BWA (Li & Durbin, 2009), the suite of programs Samtools (Daneceket et al., 2021) and the command line tools Picard (https://broadinstitute.github.io/picard/). Therefore, two sets of Single Nucleotide Polymorphism (SNPs) were obtained, one coming from alignment with the peach genome with a total of 1184 SNPs and one from almond with 895 SNPs, using GATK (DePristo et al., 2011; McKenna et al., 2010)

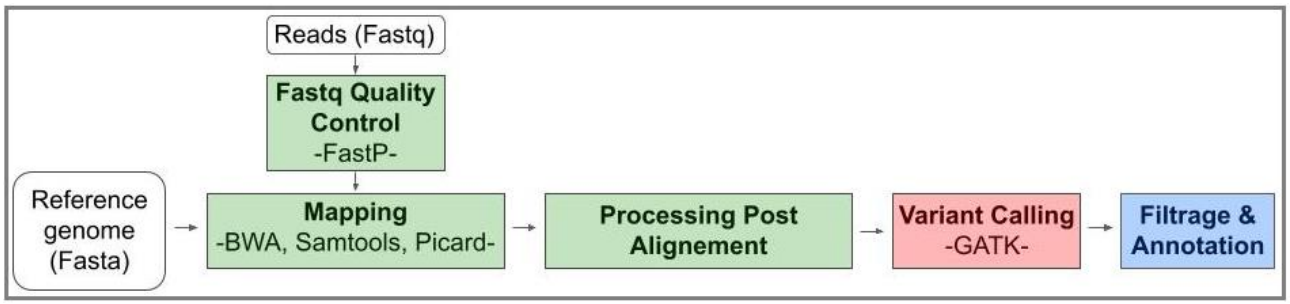


Figure 3 : SNP calling workflow according to the presentation of Robert-Siegwald G., 2018

Linkage analysis and map construction

Linkage between markers and genetic map construction were calculated using Joinmap v. 5 (Stam, 1993). The similarity thresholds for individuals and loci were set at 0.950 to highlight similar individuals and loci. Linkage groups were formed and sorted using a minimum independence LOD (logarithm of odds : probability of linkage between two chromosomal loci) score of 2. Linkage groups were determined using Haldane function. All markers were checked one by one to remove those with too many errors. For each linkage group, approximatively one marker per megabase was positioned. Markers within the resulting linkage groups were ordered relative to each other by automatic multipoint analyses using the default values of JoinMap 5 (Stam, 1993).

QTL detection

Linkage map information obtained with the software Joinmap v.5 (Stam, 1993) have been used to carry out QTL correlated with the Plum Pox Virus resistance. An association analysis between the traits measured and the markers retained was conducted using the interval mapping (IM) analysis with the software MapQTL v.6 (Van Ooijen, 2009). The cofactors on the linkage group have been determined using the Automatic Cofactor Selection (ACS) analysis. The linkage groups with LOD score more than 2.5 were used to find the exact QTL location. Restricted Multiple QTL mapping (rMQM) and Multiple QTL mapping (MQM) analysis have been used to make the graphs for the linkage groups QTLs.

Results

Miscellaneous results

From the genotyping data, four individuals (Turc_114, Vairo (Greek collection), Vairo (INRAE Avignon) and Occhiorosso-2) were removed as they showed more than 50% missing data. SSR fingerprinting also showed that the US143 accession was polyploid while it is commonly acknowledged that all species from the Amygdalus section are diploid. However, Us143 belongs to the *P. pedunculata* species and was sampled in Mongolia where Amygdalus species displaying a chromosomal number over the expected 2n=16 were described (up to 2n=88 according to Ekimova et al., 2012). The authors suppose that this high level of ploidy could be linked to a diversification and adaptation process to very constrained environmental conditions. The markers CPDCT035 and UDP98-408 were also removed as they displayed 17.70% and 22.97% missing data, respectively (see Table S2). We thus retained in our study 23 markers.

Thirty-two pairs of clones have been identified, mostly among the *P. dulcis* cultivated samples, using Genodive V.3.0 software (Meirmans, 2020). One of the two (or three) individuals showing similarities was retained for analysis. It is important to remove clones as replicates would bias allelic frequencies. We also identified through Genodive V.3.0 software (Meirmans, 2020) mis-assigned accessions, i.e. accessions bearing a wrong name of variety (for example, one individual was called 'Texas' when in fact he was a 'Ferragnes' individual). This was the case in particular among the Greek collection of cultivated almonds. After removing clonemates, in total 186 individuals were retained for the subsequent analyses.

Genetic variability and Bayesian clustering of *P. dulcis*

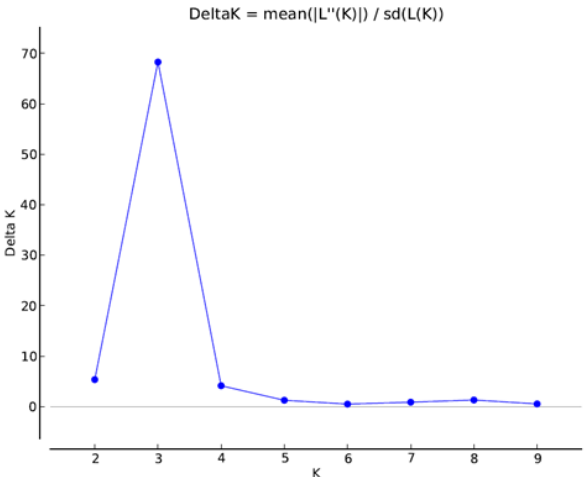


Figure 4 *: ΔK* **plotted against** *K* **values for the** ***P. dulcis*** **set of STRUCTURE analysis**. The *ΔK* was estimated by Structure harvester for the entire *P. dulcis* dataset (*N*=138). STRUCTURE results are based on 23 microsatellite markers.

The first analysis carried out concerned *P. dulcis* cultivated accessions only (*N*=138 after removal of clonemates and individuals displaying too many missing data). The objective was here to describe the genetic subdivision of this species at the intra-specific level. STRUCTURE v.2.3.4 (Porras-Hurtado et al., 2013) allows to identify subsets of the whole sample by detecting allele frequency differences within the data and can assign individuals to those sub-populations based on analysis of likelihoods. The highest *ΔK* of sample clustering is reached at *K*=3 (Figure 4). However from the STRUCTURE barplots (Figure S1), we

can see that the most relevant *K* is at *K*=4 instead. Indeed, at *K*=4, *P. dulcis* split in 4 clusters in which samples from Central Asia and Caucasus differentiate themselves from European and North American samples (Figure S1). This subdivision appears to be linked to the geographical origin of the cultivated accessions, with a clear distribution from East to West. The majority of individuals from Central Asia and Caucasus belong to the pink cluster (Figure 5). The second, most East European cluster (in purple, Figure 5) corresponds to individuals sampled in the far Eastern part of Eastern Anatolia, on the Van lake, more specifically from the Akdamar Island. Two other clusters are encountered in Europe and North America (light and dark blue, Figure 5). For North-American individuals, it was not surprising to find them assigned to European genetic clusters, since all American cultivars originate either directly or after hybridization from more ancient European varieties.

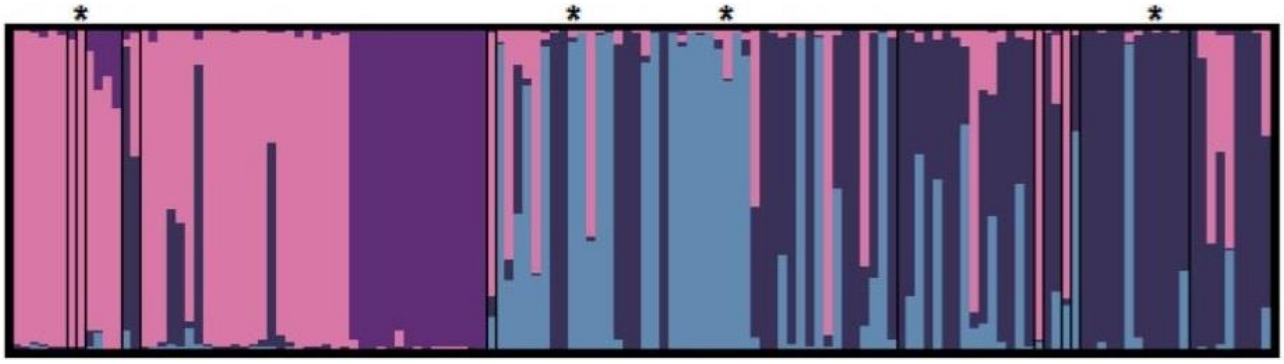


Figure 5 : Genetic subdivision of *P. dulcis* inferred with STRUCTURE for *K*=4 and analysed with 23 microsatellite markers. The one hundred and thirty eight *P. dulcis* accessions include samples from Central Asia (Kyrgyzstan, Turkmenistan and Uzbekistan, *N*=8), Caucasia (Azerbaijan, Russia and Turkey, *N*=44), Middle East (Israel, *N*=1), Europe (Spain, France, Greece, Italy and Portugal, *N*=63), North Africa (Tunisia, *N*=1), North America (United States, *N*=12) and breeding accessions (*N*=9). (*) refers to Plum Pox virus resistant samples

Genetic differentiation and relationships among the cultivated almond genetic cluster

Genetic clusters of *P. dulcis*	*N*	N_a	N_e	A_r	A_p	*I*	H_o	uH_e	*F*
Cluster_1	15	4.565 (0.392)	2.765 (0.287)	2.518 (0.166)	1.123 (0.172)	1.062 (0.109)	0.527 (0.055)	0.554 (0.052)	0.004 (0.052)
Cluster_2	28	14.696 (1.134)	8.281 (0.858)	3.791 (0.181)	1.852 (0.173)	2.181 (0.143)	0.743 (0.042)	0.818 (0.040)	0.077 (0.025)
Cluster_3	20	7.826 (0.585)	4.543 (0.381)	3.250 (0.164)	1.378 (0.148)	1.607 (0.116)	0.717 (0.052)	0.728 (0.047)	-0.006 (0.023)
Cluster_4	35	6.739 (0.508)	4.391 (0.328)	3.201 (0.115)	1.279 (0.116)	1.562 (0.082)	0.661 (0.044)	0.745 (0.027)	0.099 (0.049)
Mean	24.500	8.457 (0.530)	4.995 (0.331)	3.190	1.408	1.603 (0.070)	0.662 (0.025)	0.711 (0.023)	0.044 (0.020)

Table 1 : Genetic diversity estimators for the four *P. dulcis* populations determined with STRUCTURE at *K*=4. Only individuals which were assigned to a genetic cluster with a membership proportion greater than or equal to 90% were retained, i.e. 98 accessions. The number of different alleles (N_a), the effective number of allele (N_e), the Shannon's index (*I*), the observed heterozygoty (H_o), the unbiased expected heterozygoty (*uHe*) and the fixation index (*F*) 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. Standard deviations in brackets.

For analyses hereafter, genotypes were assigned to a given cluster if their membership coefficient for that population was ≥ 0.90, thus removing admixed individuals.

The fixation index *F* is one of the parameters the most widely used to describe population structure (Nagylaki, 1997). It permits to measure the population differentiation based on genetic polymorphism. In the case of *P. dulcis*, we can see that the fixation index is higher in the cluster represented in dark blue (*F*=0.099, +/-0.049, Table 1) comprising European and North American

individuals. However, in case of low sampling size, this estimator may be biased. The same applies to the number of different alleles per population, which is highly dependent on the sample size (N_a).

Jost's D	Cluster_1	Cluster_2	Cluster_3	Cluster_4
Cluster_1	0.000	0.001	0.001	0.001
Cluster_2	0.480	0.000	0.001	0.001
Cluster_3	0.489	0.258	0.000	0.001
Cluster_4	0.552	0.260	0.267	0.000

Table 2 : Pairwise population matrix of Jost's estimate of differentiation (*Jost's D*) among the *P. dulcis* clusters inferred with GenAlEx V.6.503. 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)

Therefore, Jost's *D* index has been calculated. This parameter will give information about the degree of differentiation between genetic clusters. It has been proposed in replacement to F_{ST} as a more accurate measure of genetic differentiation, in case of extreme F_{ST} values, low and high. One of its particularities is that its maximum values are independent of the heterozygosity of the average observed heterozygote rate, a property that the F_{ST} does not share (Alcala & Rosenberg, 2019). Jost's *D* values reveal highly significant genetic differentiation among the *P. dulcis* clusters (p< 0.001). We can observe that the highest degrees of differentiation are between cluster 1 comprising Akdamar island individuals and the three other clusters (Table 2), while the other three clusters are equally differentiated (Jost's *D* values around 0.250, Table 2).

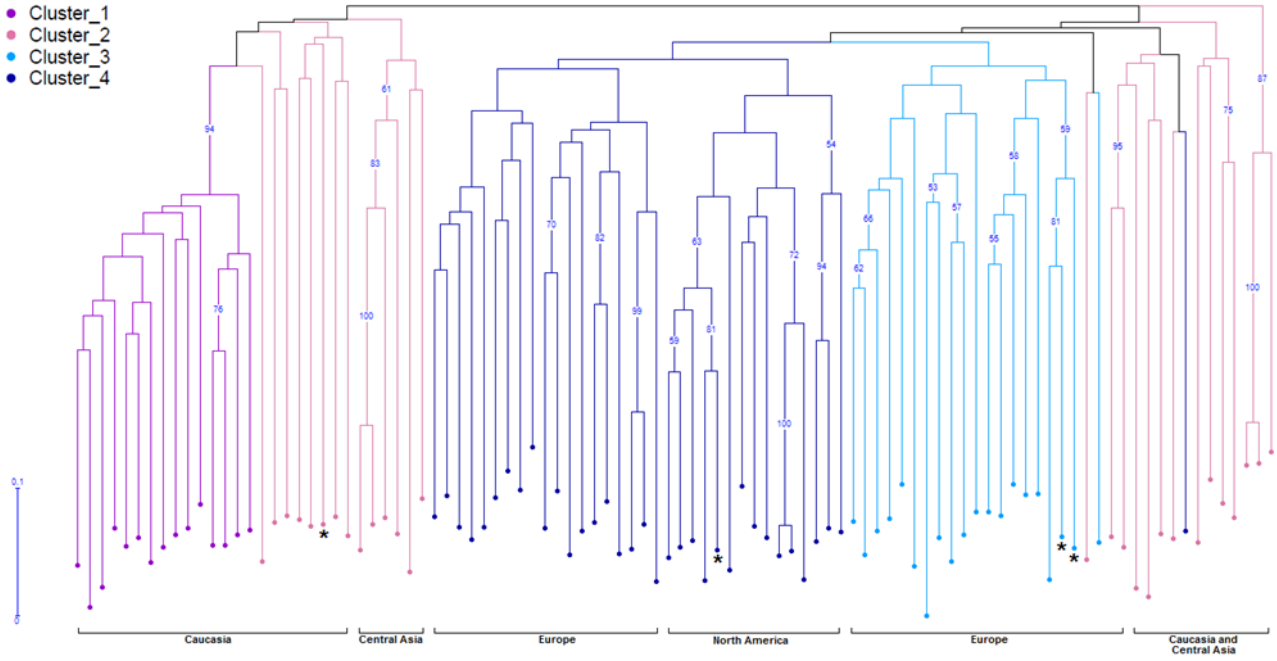


Figure 6 : Weighted neighbour-joining (NJ) tree of *P. dulcis* performed with DARwin. The sampling for this analysis accessions include *P. dulcis* samples from Central Asia (Kyrgyzstan, Turkmenistan and Uzbekistan, *N*=8), Caucasia (Turkey, *N*=33), Europe (Spain, France and Italy, *N*=42), America (United States, *N*=11) and breeding accessions (*N*=4). (*) refers to Plum Pox resistant samples. Cluster 1 is represented in purple, cluster 2 in pink, cluster 3 in light blue and the cluster 4 in dark blue following the colors used in the Structure barplots (Figure 2). 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. (*) refers to Plum Pox virus resistant samples

Thanks to the ADZE software, we calculated the allelic richness. In this estimation, the cluster represented in pink composed of individuals from Central Asia and Caucasia showed the highest

allelic richness (mean ± standard deviation: A_r = 3.791 ± 0.181) and heterozygosity level (mean ± standard deviation: H_e = 0.743 ± 0.042, Table 1). In comparison, cluster of the Akdamar island individuals displayed the lowest heterozygosity (mean ± standard deviation : H_e = 0.527 ± 0.055, Table 1) and allelic richness (mean ± standard deviation : A_r = 2.518 ± 0.166, Table 1). The purple cluster corresponds to the Akdamar population that appears geographically isolated and genetically highly differentiated from the other genetic clusters, most likely contributing to its relative high level of homozygosity.

In order to visualise how the different *P. dulcis* clusters differ, a phylogenetic tree using the Nei's standard genetic distance (D_{st}) was built (Figure 6). The dendrogram showed four major clades with the Akdamar population being related to part of the Caucasian and Central Asian almonds (Figure 6). The European samples are divided into two clades, with one corresponding to the light blue and the second one, grouping with the North-American almonds (dark blue) (Figure 6). Similarly to Bayesian clustering, cultivated almond populations thus differentiate mainly based on their geographical origin.

Bayesian clustering and differentiation of *P. dulcis* and its related species

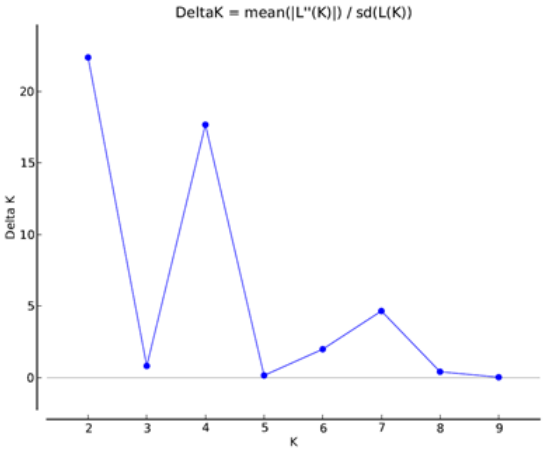


Figure 7 : *ΔK* plotted against *K* values for the *P. dulcis* and related species set of STRUCTURE analysis. The *ΔK* was estimated by Structure harvester for nine different Prunus species including *P. bucharica* (*N*=1), *P. communis* (*N*=5), *P. dulcis* (*N*=138), *P. fenzliana* (*N*=19), *P. kuramica* (*N*=1), *P. orientalis* (*N*=8), *P. spinosissima* (*N*=8), *P. turcomanica* (*N*=5) and *P. webbii* (*N*=1). STRUCTURE results are based on 23 microsatellite markers.

In order to characterize the diversity and genetic subdivision among the cultivated almond and related species, a total of 186 samples 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) were studied using 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 7). The *ΔK* showed an optimal subdivision in four clusters (Figure 7). At *K*=4 we can see that *P. fenzliana* forms a distinct cluster. The same is true for the *P. dulcis* samples from Akdamar Island. A large proportion of the cultivated almonds also form a distinct cluster (Figure S2). But based on the different barplots obtained from *K*=2 to *K*=10 (Figure S2), 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 seven did not reveal well-delimited new cluster, except

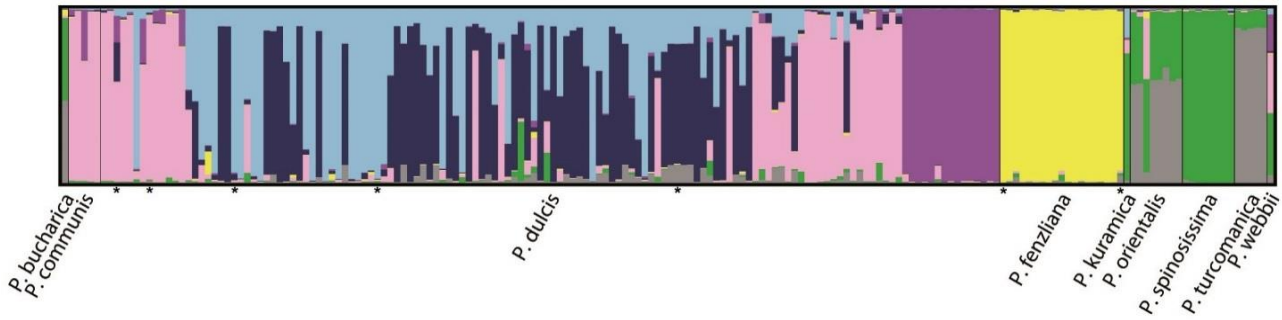


Figure 8 : Genetic subdivision among cultivated almond and related species as inferred with STRUCTURE at *K*=7 and with 23 microsatellite markers. The 186 belongs to nine different Prunus species including *P. bucharica* (*N*=1), *P. communis* (*N*=5), *P. dulcis* (*N*=138), *P. fenzliana* (*N*=19), *P. kuramica* (*N*=1), *P. orientalis* (*N*=8), *P. spinosissima* (*N*=8), *P. turcomanica* (*N*=5) and *P. webbii* (*N*=1). (*) refers to Plum Pox virus resistant samples.

for the ones distinguishing *P. orientalis* and *P. turcomanica* (*K*=8). However, this differentiation is not stable since at *K*>8, the two species are once again undistinguishable (Figure S2). This altogether suggested that *K*=7 corresponded to the most relevant *K* value for our sampling. At *K*=7 (Figure 8), the cultivated almonds split into four distinct clusters, similar to the ones described above (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 Figure 8). The other six wild related species are 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*). Bayesian clustering thus showed that individuals are mainly classified according to their species and/or geographical origin (for the *P. dulcis* clusters).

Figure 9 : Genetic clustering and spatial distribution of *P. dulcis* and related species as performed with Tableau software. The seven genetic clusters and colors correspond to the STRUCTURE barplots (Figures 5). The one hundred and thirty-one to six different Prunus species including *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 cultivated almond are represented in cercle and the related species in square.

Subsequently, in order to estimate 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). Figure 9 shows the distribution of the almond and almond-related clusters inferred by Structure. It confirms a clear spatial differentiation between the four *P. dulcis* clusters between *P. dulcis* and its wild related species following an East to West axis.

Genetic differentiation and relationships among *P. dulcis* and its related species

Almond species	*N*	*Na*	*Ne*	A_r	A_p	*I*	*Ho*	*uHe*	*F*
P. dulcis	92	16.783 (1.399)	7.590 (0.745)	3.948 (0.162)	1.529 (0.119)	2.150 (0.138)	0.666 (0.041)	0.806 (0.038)	0.164 (0.032)
P. fenzliana	19	8.696 (0.710)	4.684 (0.496)	4.118 (0.138)	2.260 (0.137)	1.629 (0.132)	0.639 (0.059)	0.710 (0.051)	0.092 (0.047)
P. orientalis	7	5.957 (0.524)	4.359 (0.437)	3.114 (0.167)	0.905 (0.144)	1.456 (0.133)	0.645 (0.064)	0.722 (0.058)	0.025 (0.059)
P. spinosissima	8	7.870 (0.531)	5.935 (0.558)	3.287 (0.120)	1.068 (0.098)	1.828 (0.088)	0.782 (0.035)	0.847 (0.024)	0.011 (0.033)
P. turcomanica	5	4.565 (0.300)	3.377 (0.251)	3.046 (0.169)	0.878 (0.121)	1.280 (0.089)	0.774 (0.057)	0.721 (0.044)	-0.191 (0.045)
Mean	26.200	7.790 (0.450)	4.786 (0.232)	3.502	1.328	1.562 (0.055)	0.691 (0.023)	0.748 (0.019)	-0.006 (0.023)

Table 3 : Genetic diversity estimators for the almond species. Only individuals which were assigned to a genetic cluster with a membership proportion greater than or equal to 90% were retained, i.e. 131 accessions. The number of different alleles (N_a), the effective number of allele (N_e), the Shannon's index (*I*), the observed heterozygoty (H_o), the unbiased expected heterozygoty (uH_e) and the fixation index (*F*) 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.

As we showed that *P. communis* individuals are in fact feral *P. dulcis*, we thus grouped the two species, *P. communis* and *P. dulcis*, for the subsequent analyses. We then computed population genetic statistics for the five remaining species, *P. webbii*, *P. kuramica* and *P. bucharica* being eliminated because admixed (Table 3). The highest fixation index is that of *P. dulcis* (*F* =0.164, +/- 0.032, Table 3). However, this value is difficult to compare with the values of the other species, because of a significantly unbalanced sample size (i.e. *N*=92 for *P. dulcis* but *N*=5 for *P. turcomanica*). Using ADZE software we were able to determine the allelic richness of the different species balanced with the sample size. As a result, *P. fenzliana* showed significantly higher allelic richness (*Ar*=4.118 ± 0.138, Table 3) than the other four species and it turns out that this same species showed the lowest value of unbiased expected heterozygoty (uH_e=0.710 ± 0.051, Table 3).

The genetic relationships have been explored with a Factorial correspondence analyses (FCAs) (Figure 10). Individuals with an assignment probability <90% to one of the seven clusters at *K*=7 were removed. The FCA (Figure 10) revealed a similar pattern as inferred by STRUCTURE, with a clear differentiation between *P. dulcis* and its wild related species (Fig. 10A) and then between the *P. dulcis* populations (Fig. 10B after removing the non-*P. dulcis* samples).

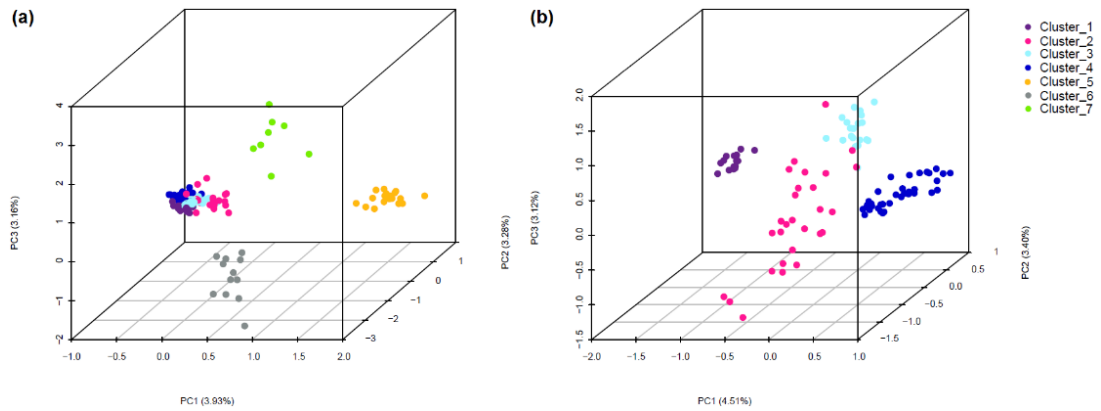


Figure 10 : Factorial correspondence analyses (FCAs). (a) Including P. *dulcis* and related species individuals (*N*=131) with membership coefficient ≥90% to a cluster in the STRUCTURE analysis at *K*=7, colored as in Figure 6. (b) Including *P. dulcis* individuals (*N*=98), colored as in Figure 3. The cluster number correspond to (1-4) *P. dulcis*, (5) *P. fenzliana*, (6) *P. orientalis* and *P. turcomanica* and (7) *P. spinosissima*

In order to visualise the relationships among the different species, a neighbour-joining tree was made with the non-admixed *P. dulcis* and related species (Figure 11). We can observe that the wild accessions are all placed on the left side of the dendogram and the *P. dulcis* clusters, on the right side. Concerning the wild relatives, *P. orientalis* and *P. turcomanica* are monophyletic, indicating a common ancestor while being still well separated into two different branches (Figure 11). The different wild species group together, with however a specific clade for *P. fenzliana*. Further structure of the dendrogram was also in agreement with the clustering inferred by STRUCTURE, with the exception of the Caucasian/Central Asian *P. dulcis* cluster (in pink, Figure 11) that split in 4 clades.

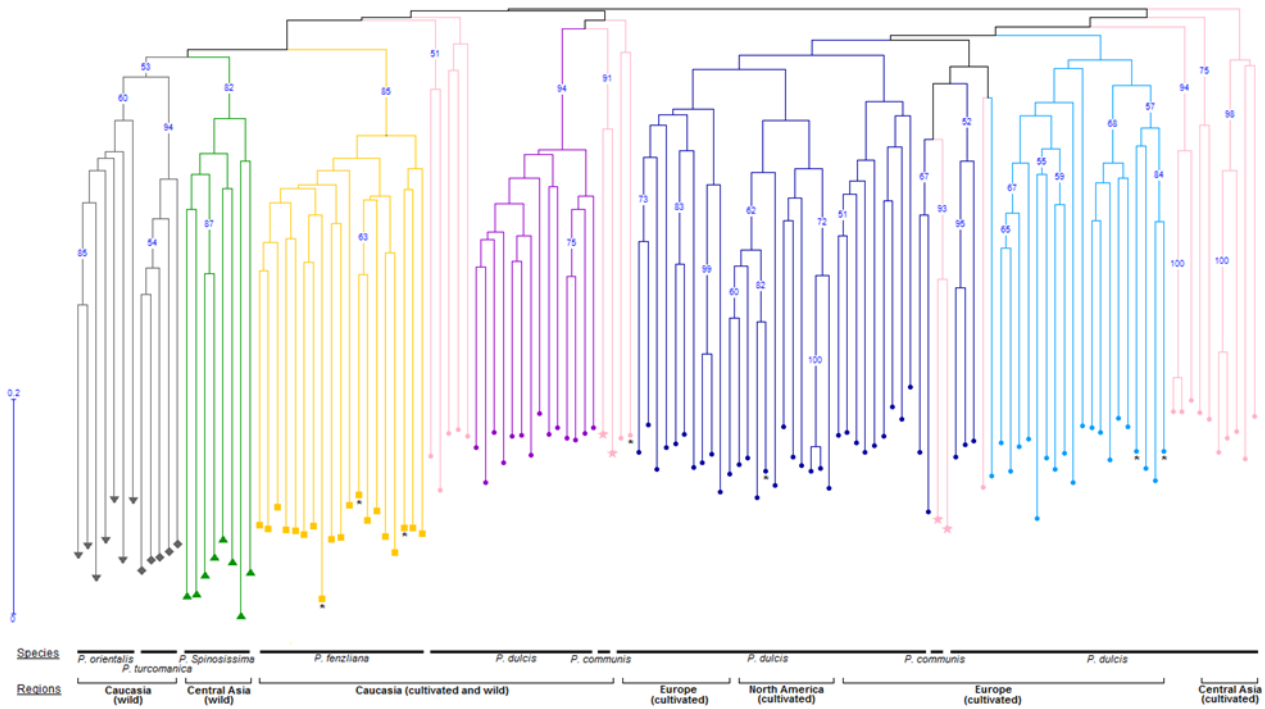


Figure 11 : Weighted neighbour-joining (NJ) tree among cultivated and wild almond species performed with DARwin. The 131 belongs to 6 different *Prunus* species including *P. communis* (*N=4*) in pink ★, *P. dulcis* (*N=89*) in pink, light blue and dark blue ●, *P. fenzliana* (*N=19*) in yellow ■, *P. orientalis* (*N=6*) in grey ▼, *P. spinosissima* (*N=8*) in green ▲, and *P. turcomanica* (*N=5*) in grey ◆. 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. (*) refers to Plum Pox virus resistant samples

Origin of the sharka resistance in the cultivated and wild Amygdalus populations

In the dendogram (Figure 11) and barplots (Figure 8), we indicated (by an asterisk) the Amygdalus accessions that displayed resistance to Plum Pox Virus. The four out of ten PPV resistant almond cultivars tested here (Figure 8, the six other accessions were not fingerprinted in this study) are distributed among three different clades of *P. dulcis* (ie clusters 2 to 4, corresponding to the pink, light and dark blue clusters). In parallel, several but not all *P. fenzliana* accessions were also found resistant to PPV infection. This would indicate at least four potential sources of resistance to PPV, three in *P. dulcis* and one in *P. fenzliana*. However, without further information on the genetic determinants controlling resistance to PPV in almond, we cannot rule out the possibility that they share the same mechanism of resistance. In an effort to answer this question, we initiated the linkage mapping of loci involved in resistance to sharka in the almond 'Del Cid' cultivar, which is affiliated to cluster 3 of *P. dulcis* (light blue, Figures 5, 8 and 11).

Linkage map

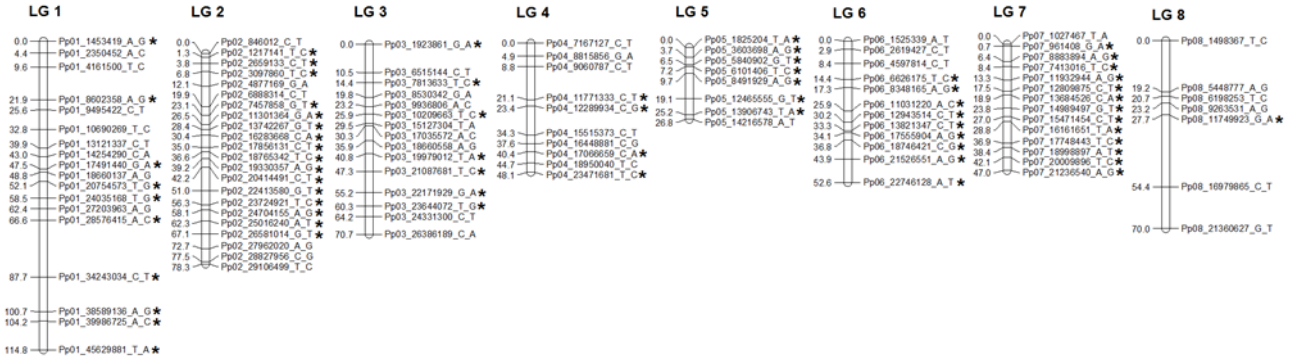


Figure 12 : F1(00) genetic linkage map with 106 peach markers (SNP) distributed over 8 linkage groups. The name of the markers are indicated on the right side of the linkage groups and the genetic distance (cM) on the left side. Marker that showed significant distortions in population are indicated by (*)

Using Joinmap v.5 software, we prepared two genetic maps, one from polymorphism identified after alignment on the peach reference genome (Verde et al., 2013) and the second, after alignment on the 'Texas' almond assembled genome (Alioto et al., 2020). This strategy is expected to reduce the loss of genomic information due to heterologous alignment of the reads, for example alignment of almond fragments onto a peach genome and vice versa. Concerning the one obtained after alignment on the peach genome, the linkage map is composed of 106 SNP markers. It covers 508.3 cM across 8 linkage groups (26.0 to 114.8 cM), with 7 to 22 markers per group (Figure 12). The smallest linkage group is LG5 which contained only 8 markers spanning a length of 26.8 cM. The largest linkage group is LG1 which has 18 markers and a length of 114.8 cM. The distance between adjacent markers on the map does not vary greatly across the different linkage groups and the average marker distance is 5.6, with intervals between loci ranging from 0.7 on LG5 and LG7 to 26.7 on LG8 (Figure 12).

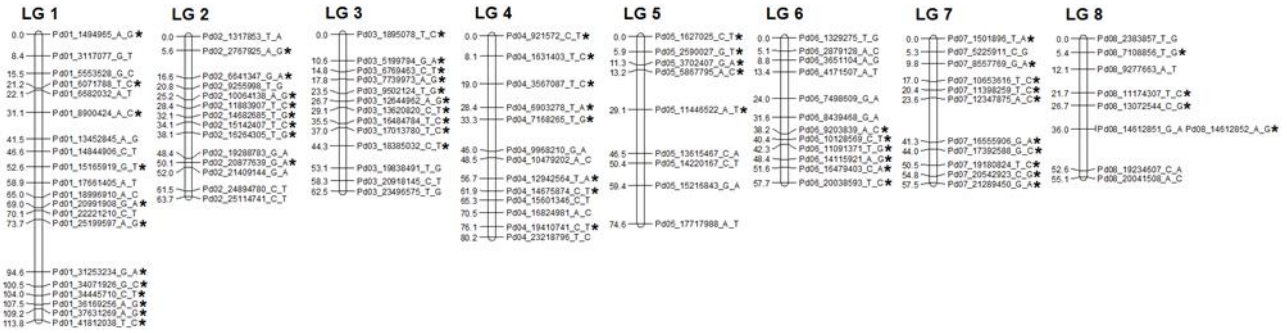


Figure 13 : F1(00) genetic linkage map with 101 almond markers (SNP) distributed over 8 linkage groups. The name of the markers are indicated on the right side of the linkage groups and the genetic distance (cM) on the left side. Marker that showed significant distortions in population are indicated by (*)

Concerning the genetic map obtained from SNP markers selected after alignment with the almond genome, the linkage map is composed of 101 SNP markers. It covers 565.1 cM across 8 linkage groups (55.1 to 113.8), with 9 to 20 markers, depending on the group (Figure 13). The smallest linkage group is LG8 which contained 9 markers spanning a length of 55.1 cM. The largest linkage group is LG1 which has 20 markers and a length of 113.8 cM.

Distribution of distorted markers on the genetic map

With the software Joinmap v. 5 (Stam, 1993), we obtained the frequencies of genotypes for each locus including the chisquare χ^2 test results for segregation according to the Mendelian expectation ratio. Both maps showed significant segregation distortion on all linkage groups (Figure 12 and 13). For the genetic linkage map obtained from peach SNPs (Figure 12), the number of markers showing segregation distortion varies from 1 to 16 per linkage group and the most extreme case of deviation from expected segregation is on LG2 with 16 markers showing distortion. On the genetic linkage map obtained from almond SNPs (Figure 13), there is an equivalent number of markers showing segregation distortion, this time ranging from 4 to 12 depending on the linkage group and the most extreme case of segregation is on LG1 with 12 markers that showed segregation distortion. We decided to add the distorted markers to the final map only when they do not affect the original orders of the markers.

QTL mapping

Trait	Linkage map	LG	Interval (cM)	Closest marker	LOD [1,2]	R^2
Viral accumulation	F1(00)^{Peach}	3	48-61	Pp03_22171929_G_A	6.11[1]	7.9
	F1(00)^{Peach}	8	23-50	Pp08_11749923_G_A	5.93[1]	7.6
	F1(00)^{Almond}	3	37-53	Pd03_18385032_C_T	6.67[1]	7.9
	F1(00)^{Almond}	8	12-26	Pd08_11174307_T_C	5.94[1]	7.2
Symptom intensity	F1(00)^{Peach}	5	0-27	Pp05_1825204_T_A	4.22[2]	5.4
	F1(00)^{Peach}	7	10-21	Pp07_11932944_A_G	4.60[2]	6.5
	F1(00)^{Almond}	4	29-39	Pd04_7168265_T_G	5.81[1]	6.6
	F1(00)^{Almond}	5	50-75	Pd05_15216843_G_A	6.81[1]	7.8
	F1(00)^{Almond}	7	10-20	Pd07_10653616_T_C	8.44[1]	10.1

Table 4 : Quantitative trait loci analysis of Plum Pox Virus resistance and associated traits. Linkage map; LG,linkage group; Interval, position on LG; Closest marker; R^2, phenotypic variance explained

F1(00)^{Peach}: linkage map of 'Honey Blaze' x 'Del Cid' with peach SNPs
F1(00)^{Almond}: linkage map of 'Honey Blaze' x 'Del Cid' with almond SNPs
1: MQM
2: rMQM

After obtaining the above linkage maps constructed with Joinmap v. 5 (Stam, 1993), a QTL mapping analysis was run first from normalized optical density values and second from visual notations of PPV symptoms intensity. The QTL approach was performed with MapQTL v.6 (Van Ooijen, 2009) on a total of 89 progenies.

From optical density values that are reflecting viral accumulation, MQM mapping detected two loci, one on LG3 (-log10(*p*-value) of 6.11) and the second on LG8 (-log10(*p*-value) of 5.93) (Figure 14 and Table 4). Concerning LG3, we identified a significant QTL between 48 and 61 cM with the peak SNP marker being Pp03_22171929_G_A. On LG8, the region is comprised between 23 and 50 cM with the peak marker Pp08_11749923_G_A. On LG3, higher PPV accumulation is correlated with the homozygous peach alleles at the locus thus indicating that the resistance is coming from almond and is dominant (Figure S3). The QTL on LG8 is more complex since both peach and almond homozygous alleles are linked to resistance to PPV. Similar loci contributing to PPV accumulation were identified when the QTL analysis was performed with the linkage map obtained from almond SNPs (Table 4).

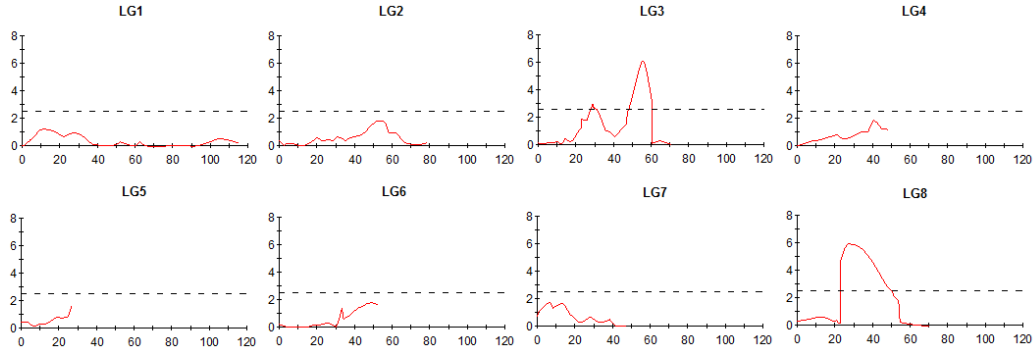


Figure 14 : Genetic mapping of loci linked to PPV accumulation in the 'Honey Blaze' x 'Del Cid' progenies. LOD (-log10(p-value)) plot (in red) for PPV accumulation (normalized optical density) was calculated with the MapQTL v.6 software by MQM mapping. The LOD threshold is represented by the dot line

Concerning the loci involved in the intensity of symptoms, we identified three QTLs (LG4, LG5 and LG7, Table 4, Figure 15 and Figure S5) over the linkage map obtained from almond markers and only two, with the peach markers (LG5 and LG7, Table 4 and Figure S5).

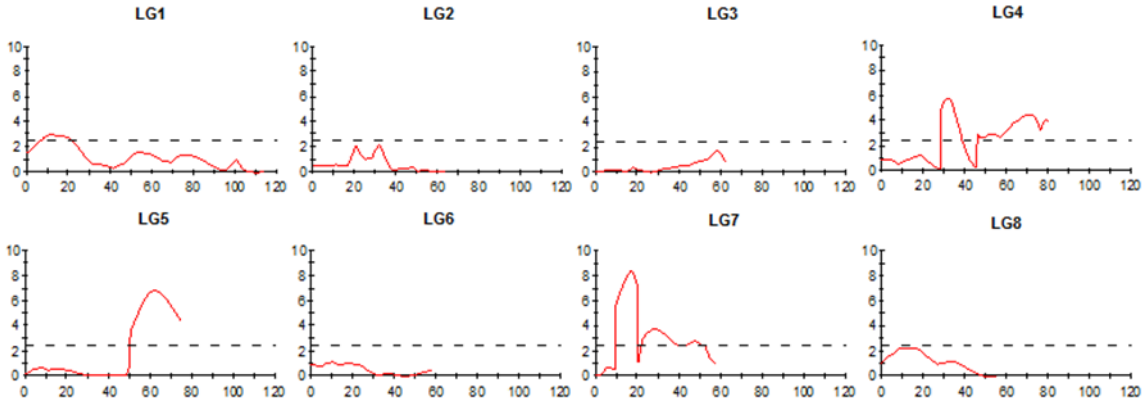


Figure 15 : Genetic regions involved in the intensity of PPV symptoms. LOD (-log10(p-value)) plot (in red) for symptom intensity was calculated with the MAPQTL6 software by MQM. The LOD threshold is represented by the dot line

Discussion

After the divergence between peach and almond that happened millions of years ago in Central Asia, Amygdalus species dispersed westwards and established in particular in severe environments, from Central Asia to the Mediterranean area, including Caucasus (Velasco et al, 2016). Previous publications showed the existence of several almond wild relatives around the Mediterranean basin but none of them questioned the relationship between cultivated almonds and its wild relatives because of 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).

In the current study, we significantly extended sampling of Amygdalus species, including to worldwide almond cultivars natural populations of wild relatives from Turkey, Caucasia and Central Asia. We found four genetically differentiated clusters of cultivated almonds (*P. dulcis*) with contrasted geographical distributions from Central Asia to Europe and North America. Turkey appears as a diversification center of cultivated almonds, with two distinct genetic clusters that may result from population isolation and island speciation in the case of the Akdamar cluster. This late population would have thus diverged due to local adaptation or reproductive isolation. 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. Besides those two genetic clusters present in Turkey, we observed two other ones in Europe. One of the two European clusters occurs specifically in Spanish almond cultivars, it might therefore correspond to the Middle East and North African population described by Delplancke *et al.* (2013). Indeed, after Muslim conquest of the Iberian Peninsula (711-726 AC), Southern Spain has been for a long time a hub of trade between Arabic and European countries (Delplancke *et al.*, 2013). A close relationship between Southern Spanish and North-African gene pool has already been documented for another stone fruit species, apricot (Bourguiba et al, 2012).

We performed Bayesian analysis highlighting the differentiation between cultivated almond trees and their wild related species (Figure 8). Cluster analysis using SSRs as valuable markers for discriminating among accessions allowed to estimate relationships among accessions. When wild related species are included in the analysis, *P. dulcis* is still differentiated in four clusters while *P. communis* samples group with the Turkish, Caucasian and Central Asian *P. dulcis* cultivars and landraces. Moreover, in the phylogenetic tree, *P. communis* individuals are placed in the same clade

as *P. dulcis*. We can therefore deduce that the species *P. communis* is a synonym of *P. dulcis*, as already proposed by Rahemi et al. (2011). Concerning *P. fenzliana*, Ladizinsky (1999) proposed this species as the wild ancestor of almond. According to our phylogenetic study, *P. fenzliana* appears as a true species distinct from the other Amygdalus species, with limited footprints of admixture. Our phylogenetic study did not support an exclusive relationship between *P. dulcis* and *P. fenzliana*. In fact, it showed a kinship between a set of *P. dulcis* cultivars (from Turkey and Caucasia) and the four wild relatives: *P. turcomanica*, *P. orientalis*, *P. spinosissima* and *P. fenzliana*, nothing more specific than that (Figure 11). Concerning *P. orientalis* and *P. turcomanica*, it is still not clear at this stage if they are two different species or not. Our Bayesian clustering and phylogenetic results suggested that they are sister species and that they share a common ancestor.

Thanks to preliminary data provided by colleagues from GAFL and BFP, we also questioned the diversity of sources of resistance to sharka in almond and its wild relatives. From ten *P. dulcis* accessions identified as resistant to PPV (Pascal et al, 2002(a); David Tricon, pers. comm.), four of them were included in our study. Those four PPV resistant almond cultivars distributed among three out of four *P. dulcis* genetic clusters, suggesting multiple origins of the resistance. In addition, we also identified variability in response to sharka among the *P. fenzliana* natural population with four out of 20 individuals being resistant. If the hypothesis of Ladizinsky is true and *P. fenzliana* is the most likely ancestor of *P. dulcis* (Ladizinsky, 1999), it is possible that *P. dulcis* has inherited the same resistance mechanism from *P. fenzliana*. This preliminary result would benefit, in the future, from a more extended testing for PPV resistance among the Amygdalus species as well as from the identification of the genetic factors controlling resistance to sharka in almond.

In a first attempt to map the genomic regions linked to PPV resistance, we carried out a QTL analysis among an interspecific peach x almond F2 population. The two grand-parents of the progenies are ‘Honey Blaze’ for the peach cultivar and ‘Del Cid’ for the almond PPV resistant genitor. A total of two genomic regions were identified as being involved in viral accumulation and three other loci were linked to symptom severity. While the likelihood (LOD score) was rather significant, the explained variance (i.e. the individual contribution to phenotypic variance) remained low. This could be due to the limited number of individuals tested (89) and/or the limited number of vegetative cycles taken into consideration here (one cycle). Moreover, symptom scoring appeared to be affected by a certain degree of subjectivity tightly dependent on the person doing the notations. In consequence, the quality of the phenotypic data has a significant impact on the reliability of our results. Nevertheless, this preliminary approach allowed setting up pipelines of data analysis for the upcoming cycles (planned for Automn 2021 and two more in 2022).

In a previous genome-wide association study performed in peach, Cirilli et al (2017) identified three associated loci on chromosome 2 and 3, accounting for most of the reduction in PPV-M susceptibility (Cirilli et al., 2017). In addition, weak associations in collinear region have also been demonstrated on chromosome 3 in apricot (Mariette et al., 2016; Marandel et al, 2009). Here, we have also identified a region at the bottom of chromosome 3 potentially involved in sharka resistance. While the resistance trait on chromosome 3 is dominant and the resistant allele is coming from almond, the resistance mechanism linked to the locus on chromosome 8 appears more complex at this stage. It requires further analysis such as the estimation of QTL interactions and of the additive effect. As for the analysis of symptoms intensity in response to PPV infection, we identified two regions from the peach SNP-based genetic map (LG5 and LG7) and three (LG4, LG5 and LG7) from the almond SNP-based map. Such result would indicate that the LG4 QTL is specific to almond and the corresponding region is missing in peach. Interestingly, none of the regions linked to symptom severity colocalised with QTLs involved in viral accumulation. This is either questioning the relevance of the symptom notation for resistance mapping in peach x almond progenies or it indicates that the genetic determinants for symptom severity are not linked to the resistance trait. It is also possible that symptom severity in the 'Honey Blaze' x 'Del Cid' population is not correlated to the level of virus accumulation, as we showed previously in *Arabidopsis thaliana* (Sicard et al, 2008). The results we have presented here are essentially preliminary and will need to be completed by further analyses of data collected during the next two vegetative cycles. Following this, it will be possible to determine reliable molecular markers linked to PPV resistance in almond and possibly candidate genes. These markers will be then tested on a >1,000 individuals large F2 population of 'Honey Blaze' x 'Del Cid' with the aim to implement genome-wide prediction and selection of new peach cultivars resistant to sharka.

Bibliography

Ak, B.E., Acar I. and Sakar. E. (2001). An investigation on the determination of pomological and morphological traits of wild almond grown at Sanliurfa province. CIHEAM. Options Mediterraneennes 56:139-144. http://om.ciheam.org/article.php?IDPDF=1600167

Alcala, N. and Rosenberg, N.A. (2019). Jost's D., and FST are similarly constrained by allele frequencies: A mathematical, simulation, and empirical study. Mol Ecol 28: 1624–1636. DOI: 10.1111/mec.15000

Alioto, T. et al. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: results from the almond genome sequence. Plant J 101: 455–472. DOI: 10.1111/tpj.14538

Aranzana, M., Abbassi, E.-K., Howad, W., and Arús, P. (2010). Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genet 11: 69. DOI: 10.1186/1471-2156-11-69

Atanasoff, D. (1932) Plum pox.A new virus disease. Annals of the University of Sofia, Faculty of Agriculture and Silviculture 11, 49–69.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Retrieved from http://www.genetix.univ-montp2.fr/genet ix/genet ix.htm

Bortiri, E., Oh, S.-H., Jiang, J., Baggett, S., and Granger, A. (2009). Phylogeny and Systematics of Prunus (Rosaceae) as Determined by Sequence Analysis of ITS and the Chloroplast trnL-trnF Spacer DNA. SYSTEMATIC BOTANY 26: 12. DOI: 10.1043/0363-6445-26.4.797

Bourguiba, H., Audergon, J.-M., Krichen, L., Trifi-Farah, N., Mamouni, A., Trabelsi, S., D'Onofrio, C., Asma, B.M., Santoni, S., and Khadari, B. (2012). Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. BMC Plant Biol 12: 49. DOI: 10.1186/1471-2229-12-49

Browicz, K. and Zohary, D. (1996). The genus Amygdalus L. (Rosaceae): Species relationships, distribution and evolution under domestication. Genet Resour Crop Evol 43: 229–247. DOI: 10.1007/bf00123275

Cambra, M., Capote, N., Myrta, A., and Llácer, G. (2006). Plum pox virus and the estimated costs associated with sharka disease. EPPO Bulletin 36: 202–204. DOI: 10.1111/j.1365-2338.2006.01027.x

Celetti, M., Fraser, H., Carter, N., and Llewellyn, J., (2002). Sharka (plum pox virus) of stone fruit and ornamental Prunusspecies. Factsheet No. 634, Ontario Ministry of Agriculture,Food and Rural Affairs. http:// http://www.omafra.gov.on.ca/english/crops/facts/02-001.htm#relatedlinks

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884–i890. DOI: 10.1093/bioinformatics/bty560

Cirilli, M., Rossini, L., Geuna, F., Palmisano, F., Minafra, A., Castrignanò, T., Gattolin, S., Ciacciulli, A., Babini, A.R., Liverani, A., and Bassi, D. (2017). Genetic dissection of Sharka disease tolerance in peach (P. persica L. Batsch). BMC Plant Biol 17: 192. DOI: 10.1186/s12870-017-1117-0

Dallot, S., Bousalem, M., Boeglin, M., Renaud, L.Y., and Quiot, J.B. (1997). Potential role of almond in sharka epidemics: susceptibility under controlled conditions to the main types of plum pox potyvirus and survey for natural infections in France. EPPO Bulletin 27: 539–546. DOI: 10.1111/j.1365-2338.1997.tb00681.x

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience 10: giab008. DOI: 10.1093/gigascience/giab008

Decroocq, S. et al. (2016). New insights into the history of domesticated and wild apricots and its contribution to Plum pox virus resistance. Mol Ecol 25: 4712–4729. DOI: 10.1111/mec.13772

Delplancke, M., Alvarez, N., Benoit, L., Espíndola, A., I Joly, H., Neuenschwander, S., and Arrigo, N. (2013). Evolutionary history of almond tree domestication in the Mediterranean basin. Mol Ecol 22: 1092–1104. DOI: 10.1111/mec.12129

DePristo, M.A. et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491–498. DOI: 10.1038/ng.806

Dicenta, F., Martinez-Gomez, P., Rubio, M., and Audergon, J.M. (2003). Localisation and movement of Plum pox virus in apricot stem tissues. Ann Applied Biology 142: 99–105. DOI: 10.1111/j.1744-7348.2003.tb00234.x

Earl, D.A. and vonHoldt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genet Resour 4: 359–361. DOI: 10.1007/s12686-011-9548-7

Ekimova, N.V., Muratova, E.N., and Silkin, P.P. (2012). The role of polyploidy in adaptation and settling of steppe shrubs in Central Asia. Russ J Genet Appl Res 2: 105–109. DOI: 10.1134/S2079059712020037

Esmenjaud, D., Voisin, R., Van Ghelder, C., Bosselut, N., Lafargue, B., Di Vito, M., Dirlewanger, E., Poëssel, J.L., and Kleinhentz, M. (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. DOI: 10.1007/s11295-008-0173-x

García, J.A., Glasa, M., Cambra, M., and Candresse, T. (2014). Plum pox virus and sharka: a model potyvirus and a major disease: Plum pox virus. Molecular Plant Pathology 15: 226–241. DOI: 10.1111/mpp.12083

Gildow, F., Damsteegt, V., Stone, A., Schneider, W., Luster, D., and Levy, L. (2004). Plum Pox in North America: Identification of Aphid Vectors and a Potential Role for Fruit in Virus Spread. Phytopathology® 94: 868–874. DOI: 10.1094/PHYTO.2004.94.8.868

Jakobsson, M. and Rosenberg, N.A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801–1806. DOI: 10.1093/bioinformatics/btm233.

James, D. and Thompson, D. (2006). Hosts and symptoms of Plum pox virus: ornamental and wild Prunus species. EPPO Bulletin 36: 222–224. DOI: 10.1111/j.1365-2338.2006.00976.x

James, D. (2017). Perspective on strategies for controlling the spread of Plum pox virus, causal agent of sharka/plum pox disease. Acta Hortic.: 129–136. DOI: 10.17660/ActaHortic.2017.1163.19

Karayiannis, I. and Mainou, A. (1994). Resistance to plum pox potyvirus in apricots. EPPO Bulletin 24: 761–765. DOI: 10.1111/j.1365-2338.1994.tb01093.x

Kose, A. (2014) Determination of naturally grown amygdalus species and different types in Adiyaman province. Master thesis. Harran University, Institute of Science, Sanliurfa, 6-7.

Ladizinsky, G. (1999). On the origin of almond. Genetic Ressources and Crop Evolution 46: 143-147. DOI: 10.1023/A:1008690409554

Levy, L., Damsteegt, V., Scorza, R., and Kolber, M. (2000). Plum Pox Potyvirus Disease of Stone Fruits. APSnet Features. DOI: 10.1094/APSnetFeature-2000-0300

Li, H. and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26: 589–595. DOI: 10.1093/bioinformatics/btp698

Ligges, U., & Mächler, M. (2002). Scatterplot3d - An R package for visualizing multivariate data. Retrieved from https://EconP apers.repec.org/RePEc:zbw:sfb47 5:200222

Liu, S., Decroocq, S., Harte, E., Tricon, D., Chague, A., Balakishiyeva, G., Kostritsyna, T., Turdiev, T., Fisher-Le Saux, M., Dallot, S., Giraud, T., and Decroocq, V. (2021). Genetic diversity and population structure analyses in the Alpine plum (Prunus brigantina Vill.) confirm its affiliation to the Armeniaca section. Tree Genetics & Genomes 17. DOI : 10.1007/s11295-020-01484-6

Lowery, D.T., Vickers, P.M., Bittner, L.A., Stobbs, L.W., and Foottit, R.G. (2015). Aphid Transmission of the Ontario Isolate of Plum Pox Virus. JOURNAL OF ECONOMIC ENTOMOLOGY: 6. DOI: 10.1093/jee/tov172

Marandel, G., Salava, J., Abbott, A., Candresse, T., and 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. DOI: 10.1111/j.1364-3703.2009.00535.x

Mariette, S. et al. (2016). Genome-wide association links candidate genes to resistance to Plum Pox Virus in apricot (Prunus armeniaca). New Phytol 209: 773–784. DOI: 10.1111/nph.13627

Martínez-Gómez, P., & Dicenta, F. (2001). Distribution of coat protein and nucleic acid of Plum pox virus (PPV) in seedlings of peach rootstock GF305 and apricot cv. Real Fino. Phytopathologia Mediterranea, 40(2), 157-164. DOI: 10.14601/Phytopathol_Mediterr-1596.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20: 1297–1303. DOI: 10.1101/gr.107524.110

Meirmans, P.G. (2020). Genodive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. Mol Ecol Resour 20: 1126–1131. DOI: 10.1111/1755-0998.13145

Nagylaki, T. (1998). Fixation Indices in Subdivided Populations. Genetics 148: 1325–1332. http:// https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1460034/pdf/9539445.pdf

Németh, M. (1986). Virus, Mycoplasma and Rickettsia Diseases of Fruit Trees. Martinus-Nijhoff Publishers, Dordrecht, the Netherlands.

Oliveira Lino, L., Pacheco, I., Mercier, V., Faoro, F., Bassi, D., Bornard, I., and Quilot-Turion, B. (2016). Brown Rot Strikes Prunus Fruit: An Ancient Fight Almost Always Lost. J. Agric. Food Chem. 64: 4029–4047. DOI: 10.1021/acs.jafc.6b00104

Pascal, T., Pfeiffer, F., and Kervella, J. (2002a). PRELIMINARY OBSERVATIONS ON THE RESISTANCE TO SHARKA IN PEACH AND RELATED SPECIES. Acta Hortic.: 699–704. DOI: 10.17660/ActaHortic.2002.592.98

Pascal, T., Pfeiffer, F., Kervella, J., Lacroze, J.p. and Sauge, M.H. (2002b). Inheritance of green peach aphid resistance in the peach cultivar 'Rubira'. Plant Breeding 121, 459-461. DOI: 10.1046/j.1439-0523.2002.734333.x

Pascal, T., Aberlenc, R., Confolent, C., Hoerter, M., Lecerf, E., Tuéro, C., and Lambert, P. (2017). Mapping of new resistance (Vr2, Rm1) and ornamental (Di2, pl) Mendelian trait loci in peach. Euphytica 213: 132. DOI: 10.1007/s10681-017-1921-5

Peakall, R. and Smouse, P.E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics 28: 2537–2539. DOI: 10.1093/bioinformatics/bts460

Perrier, X., Jacquemoud-Collet, J.P. (2006). Retrieved from http://darwin.cirad.fr/

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, Á., and Lareu, M.V. (2013). An overview of STRUCTURE: applications, parameter settings, and supporting software. Front. Genet. 4. DOI: 10.3389/fgene.2013.00098

Rosenberg, N.A. (2003). Distruct: a program for the graphical display of population structure: PROGRAM NOTE. Molecular Ecology Notes 4: 137–138. DOI : 10.1046/j.1471-8286.2003.00566.x

Roy, A.S. and Smith, I.M. (1994). Plum pox situation in Europe. EPPO Bulletin 24: 515–523. DOI: 10.1111/j.1365-2338.1994.tb01064.x

Salava, J., Polak, J., and Oukropec, I. (2013). Evaluation of the Prunus Interspecific Progenies for Resistance to Plum Pox Virus. Czech Journal of Genetics and Plant Breeding 49(2):65-69. DOI: 10.17221/252/2012-CJGPB

Schneider, W.L., Damsteegt, V.D., Gildow, F.E., Stone, A.L., Sherman, D.J., Levy, L.E., Mavrodieva, V., Richwine, N., Welliver, R., and Luster, D.G. (2011). Molecular, Ultrastructural, and Biological Characterization of Pennsylvania Isolates of Plum pox virus. Phytopathology® 101: 627–636. DOI: 10.1094/PHYTO-09-10-0256

Sicard, O., Loudet, O., Keurentjes, J.J.B., Candresse, T., Le Gall, O., Revers, F., and Decroocq, V. (2008). Identification of Quantitative Trait Loci Controlling Symptom Development During Viral Infection in Arabidopsis thaliana. MPMI 21: 198–207. DOI: 10.1094/MPMI-21-2-0198

Soundararajan, P., Won, S.Y., and Kim, J.S. (2019). Insight on Rosaceae Family with Genome Sequencing and Functional Genomics Perspective. BioMed Research International 2019: 1–12. DOI: 10.1155/2019/7519687

Souquet, M., Borron, S., Marchat, S., Rimbaut, L., Thébaud, G., et al. (2015). 15è Rencontres de Virologie Végétale (RVV). hal-01602149

Stam, P. (1993). Construction of integrated genetic linkage maps by means of a new computer package: Join Map.: 6. Retrieved from https://www.kyazma.nl/index.php/JoinMap/

Szpiech, Z.A., Jakobsson, M., and Rosenberg, N.A. (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. Bioinformatics 24: 2498–2504. DOI: 10.1093/bioinformatics/btn478

Rubio, M., Martínez-Gómez, P., Dicenta, F., and Weber, W.E. (2003). Resistance of almond cultivars to Plum pox virus (sharka). Plant Breeding 122: 462–464. DOI: 10.1046/j.1439-0523.2003.00872.x

Van Ooijen, J.W. (2009). MapQTL® 6: software for the mapping of quantitative trait loci in experimental populations of diploid species Kyazma BV. Wageningen, The Netherlands, p 59. Retrieved from https://www.kyazma.nl/index.php/MapQTL/

Velasco, D., Hough, J., Aradhya, M., and Ross-Ibarra, J. (2016). Evolutionary Genomics of Peach and Almond Domestication.: 9. DOI:10.1534/g3.116.032672

The International Peach Genome Initiative et al. (2013). The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45: 487–494. DOI: 10.1038/ng.2586

Yazbek, M. and Oh, S.-H. (2013). Peaches and almonds: phylogeny of Prunus subg. Amygdalus (Rosaceae) based on DNA sequences and morphology. Plant Syst Evol 299: 1403–1418. DOI: 10.1007/s00606-013-0802-1

Zohary, D. (1996) The mode of domestication of the founder crops of Southwest Asian agriculture. In: The Origins and Spread of Agriculture and Pastoralism in Eurasia (ed. Harris DR), pp. 142–158. University College London Press, London.

Annexes

Table S1 : List of individuals used : The name of the indivdiduals, species name, the presumed origin of the sample and the sample supplier correspond to (1) INRAE Bordeaux, (2) INRAE Avignon, (3) HAO Greece, (4) IRTA Spain and (5) CU Turkey

Individual name	Species	Origin	Supplier	Individual name	Species	Origin	Supplier	Individual name	Species	Origin	Supplier
5/1	*P. dulcis*	TUR	5	az270_2	*P. fenzliana*	AZE	1	STELLIETTE	*P. dulcis*	FRA	4
06/12	*P. dulcis*	TUR	5	az271_3	*P. fenzliana*	AZE	1	Tardive de la Verdière	*P. dulcis*	FRA	2
7/27	*P. dulcis*	TUR	5	Bartre-JR	*P. dulcis*	FRA	2	Tardy Non Pareil	*P. dulcis*	USA	2
(770 x FL)190	*P. dulcis*	HFR	2	Belona	*P. dulcis*	ESP	2	TARDY NONPAREIL	*P. dulcis*	USA	4
(884 x (335 x 189)80)19	*P. dulcis*	HFR	2	Bertina	*P. dulcis*	ESP	2	TARRACO	*P. dulcis*	ESP	4
(884 x (335 x 486)44)56	*P. dulcis*	HFR	2	BLANQUERNA	*P. dulcis*	ESP	4	TARRAGONÉS	*P. dulcis*	ESP	4
(R269 x P. webbii)22	*P. dulcis*	HFR	2	BONITA SAINT BRAS	*P. dulcis*	PRT	2	TEXAS	*P. dulcis*	ESP	4
(R335 x R219)28	*P. dulcis*	HFR	2	CARRERO	*P. dulcis*	ESP	4	Texas	*P. dulcis*	USA	2
(R486 x (335 x 486)77)22	*P. dulcis*	HFR	2	CAVALIERA	*P. dulcis*	ESP	4	Texas	*P. dulcis*	USA	3
(R486 x dehiscens)2	*P. dulcis*	HFR	2	CEBAS-1	*P. dulcis*	ESP	4	TITAN	*P. dulcis*	USA	4
(R769 x dehiscens)8	*P. dulcis*	HFR	2	CONSTANTÍ	*P. dulcis*	ESP	4	TOKYO	*P. dulcis*	USA	2
101/13	*P. dulcis*	TUR	5	CRISTOMORTO	*P. dulcis*	ESP	4	Tuono_inra	*P. dulcis*	ITA	2
17/1	*P. dulcis*	TUR	5	D. LARGUETA ULLDEMOLINS	*P. dulcis*	ESP	4	Turc 106	*P. dulcis*	TUR	1
21/9	*P. dulcis*	TUR	5	DEL CID	*P. dulcis*	ESP	4	Turc 107	*P. dulcis*	TUR	1
42/27	*P. dulcis*	TUR	5	Depranoto	*P. dulcis*	GRC	2	Turc 108	*P. dulcis*	TUR	1
47/2	*P. dulcis*	TUR	5	DOMINGO DABRIO	*P. dulcis*	ESP	4	Turc 109	*P. dulcis*	TUR	1
48/1	*P. dulcis*	TUR	5	ESPERANZA FORTA	*P. dulcis*	ESP	4	Turc 110	*P. dulcis*	TUR	1
48/4	*P. dulcis*	TUR	5	FALSA BARESE	*P. dulcis*	ESP	4	Turc 111	*P. dulcis*	TUR	1
48/6	*P. dulcis*	TUR	5	FELISIA	*P. dulcis*	ESP	4	Turc 112	*P. orientalis*	TUR	1
48/9	*P. dulcis*	TUR	5	Ferraduel	*P. dulcis*	FRA	3	Turc 113	*P. orientalis*	TUR	1
A6	*P. dulcis*	KGZ	1	FERRADUEL MUTANTE	*P. dulcis*	FRA	4	Turc 114	*P. orientalis*	TUR	1
A7	*P. dulcis*	KGZ	1	FERRAGNES	*P. dulcis*	FRA	4	Turc 115	*P. orientalis*	TUR	1
ACHAAK	*P. dulcis*	ESP	4	Ferragnes	*P. dulcis*	FRA	3	Turc 117	*P. dulcis*	TUR	1
AGER 1	*P. dulcis*	ESP	4	Ferragnès	*P. dulcis*	FRA	2	Turc 118	*P. dulcis*	TUR	1
AGER 2	*P. dulcis*	ESP	4	Ferralise	*P. dulcis*	FRA	2	Turc 119	*P. dulcis*	TUR	1
AGER 3	*P. dulcis*	ESP	4	Ferrastar	*P. dulcis*	FRA	2	Turc 120	*P. dulcis*	TUR	1
Aï	*P. dulcis*	FRA	2	FERRASTAR	*P. dulcis*	FRA	4	Turc 121	*P. dulcis*	TUR	1
AÏ	*P. dulcis*	FRA	4	GABAIX	*P. dulcis*	ESP	4	Turc 122	*P. dulcis*	TUR	1
AKBADEM (48/2)	*P. dulcis*	TUR	5	Gulcan 1 (101/23)	*P. dulcis*	TUR	5	Turc 123	*P. dulcis*	TUR	1
ALNEM-1	*P. dulcis*	ISR	2	HACI ALIBEY(48/5)	*P. dulcis*	TUR	5	Turc 124	*P. dulcis*	TUR	1
AM 34	*P. dulcis*	FRA	2	Independence	*P. dulcis*	USA	2	Turc 125	*P. dulcis*	TUR	1
Amandier rose	*P. dulcis*	HFR	2	JEFFRIES	*P. dulcis*	USA	4	Turc 126	*P. dulcis*	TUR	1
Amyg_Spino	*P. spinosissima*	KGZ	1	kr19-8	*P. dulcis*	KGZ	1	Turc 127	*P. dulcis*	TUR	1
ANGONES	*P. dulcis*	ESP	4	kr19-9	*P. dulcis*	KGZ	1	Turc 128	*P. dulcis*	TUR	1
Anxaneta	*P. dulcis*	ESP	2	KR21-1	*P. dulcis*	KGZ	1	Turc 129	*P. dulcis*	TUR	1
ANXANETA	*P. dulcis*	ESP	4	KR22-1	*P. dulcis*	KGZ	1	Turc 130	*P. dulcis*	TUR	1
ARDECHOISE	*P. dulcis*	FRA	4	Lauranne	*P. dulcis*	FRA	3	Turc 131	*P. dulcis*	TUR	1
Ardèchoise	*P. dulcis*	FRA	2	Macaluso	*P. dulcis*	TUN	2	Turc 132	*P. dulcis*	TUR	1
ASC13	*P. spinosissima*	KGZ	1	MAKAKO	*P. dulcis*	ESP	4	Turc 133	*P. dulcis*	TUR	1
ASC2	*P. spinosissima*	KGZ	1	Mandaline	*P. dulcis*	FRA	2	Turc 81	*P. communis*	TUR	1
ASC3	*P. spinosissima*	KGZ	1	Marcona	*P. dulcis*	ESP	2	Turc 82	*P. communis*	TUR	1
ASC4	*P. spinosissima*	KGZ	1	MARCONA	*P. dulcis*	ESP	4	Turc 83	*P. communis*	TUR	1
ASC5	*P. spinosissima*	KGZ	1	Mardia	*P. dulcis*	ESP	2	Turc 84	*P. communis*	TUR	1
ASC6	*P. spinosissima*	KGZ	1	Marinada	*P. dulcis*	ESP	2	Turc 85	*P. communis*	TUR	1
ASC7	*P. spinosissima*	KGZ	1	Marinada	*P. dulcis*	ESP	3	Turc 86	*P. orientalis*	TUR	1
ASPERILLA	*P. dulcis*	ESP	4	MARQUET	*P. dulcis*	ESP	4	Turc 87	*P. orientalis*	TUR	1
Avola	*P. dulcis*	ITA	2	MARTA	*P. dulcis*	ESP	4	Turc 88	*P. orientalis*	TUR	1
Ayles	*P. dulcis*	ESP	2	MENA D'EN MUSTE	*P. dulcis*	ESP	4	Turc 89	*P. orientalis*	TUR	1
az19-1	*P. dulcis*	AZE	1	MILOW	*P. dulcis*	USA	2	Turc 90	*P. orientalis*	TUR	1
az19-3	*P. dulcis*	AZE	1	MOLL CARRISET	*P. dulcis*	ESP	4	Turc 91	*P. turcomanica*	TUR	1
az19-5	*P. dulcis*	AZE	1	MONO	*P. dulcis*	USA	2	Turc 92	*P. turcomanica*	TUR	1
az19-8	*P. dulcis*	AZE	1	NANO	*P. dulcis*	ESP	4	Turc 93	*P. turcomanica*	TUR	1
az203-5	*P. fenzliana*	AZE	1	Nimes_mazet	*P. dulcis*	FRA	2	Turc 94	*P. turcomanica*	TUR	1
az205-1	*P. fenzliana*	AZE	1	Non Pareil	*P. dulcis*	USA	2	Turc 95	*P. turcomanica*	TUR	1
az205-5	*P. fenzliana*	AZE	1	NONPAREIL	*P. dulcis*	USA	4	us012	*P. dulcis*	ESP	1
az206_1	*P. fenzliana*	AZE	1	NORMAN	*P. dulcis*	USA	2	us039	*P. kuramica*	PAK	1
az207-2	*P. fenzliana*	AZE	1	NORTHLAND	*P. dulcis*	USA	2	us127	*P. bucharica*	INC	1
az209_5	*P. fenzliana*	AZE	1	Occhiorosso-2	*P. dulcis*	ITA	2	us143	*pedunculata*	MNG	1
az210-6	*P. fenzliana*	AZE	1	PAUET	*P. dulcis*	ESP	4	us179	*P. dulcis*	UZB	1
az211-10	*P. fenzliana*	AZE	1	Penta	*P. dulcis*	ESP	2	us183	*P. dulcis*	TKM	1
az213_4	*P. fenzliana*	AZE	1	PEP DE JUNEDA	*P. dulcis*	ESP	4	us193	*P. webbii*	INC	1
az214_1	*P. fenzliana*	AZE	1	PESTAÑETA	*P. dulcis*	ESP	4	UZBEK-923	*P. dulcis*	SUN	2
az215-1	*P. fenzliana*	AZE	1	Pointue d'Aureille	*P. dulcis*	FRA	2	Vairo	*P. dulcis*	ESP	2
az216_5	*P. fenzliana*	AZE	1	RAMILLETE	*P. dulcis*	ESP	4	Vairo	*P. dulcis*	ESP	3
az217_3	*P. fenzliana*	AZE	1	ROF	*P. dulcis*	USA	2	VERD	*P. dulcis*	ESP	4
az219-1	*P. fenzliana*	AZE	1	RUMBETA	*P. dulcis*	ESP	4	VIALFAS	*P. dulcis*	ESP	4
az220-12	*P. fenzliana*	AZE	1	Sekerci	*P. dulcis*	TUR	5	WAWONA	*P. dulcis*	ESP	4
az222_1	*P. fenzliana*	AZE	1	Soleta	*P. dulcis*	ESP	2	YALTA	*P. dulcis*	SUN	2
az268_5	*P. fenzliana*	AZE	1	SOLETA	*P. dulcis*	ESP	4	YOSEMITE	*P. dulcis*	USA	2
az269_4	*P. fenzliana*	AZE	1	Soleta	*P. dulcis*	ESP	3				

Table S2 : List of microsatellite markers used : the name of the locus; the repeat motif; the forward and reverse primer sequences; the temperature of annealing; the locus size; the percent of missing data; the number of different alles (N_a) and the number of effective alleles (N_e)

MP	Marker names	Repeat motif	Forward primer sequence	Reverse primer sequence	Temperature of annealing	Size (bp)	Missing data (%)	N_a (*cultivated*)	N_a (cultivated)	N_a (cultivated and related species)	N_e (cultivated and related species)
MP1	UDP96-001	(CA)17	AGTTTGATTTTCTGATGCATCC	TGCCATAAGGACCGGTATGT	57	120	4	13	2.807	21	3.818
MP11	UDP98-409	(AG)19	GCTGATGGGTTTTATGGTTTTC	CGGACTCTTATCCTCTATCAACA	57	129	13	22	8.248	39	12.248
MP12	UDP96-013	(AG)22[TG]8TT[TG]10	ATTCTTCACTACACGTGCACG	CCCCAGACATACTGTGGCTT	57	198	8	38	13.350	57	19.791
MP139	CPPCT033	(CT)16	TCAGCAAACTAGAAACAAACC	TTGCAATCTGTGTTGATGTT	50	151	3	19	9.207	19	9.991
MP14	UDP97-401	(GA)19	TAAGAGGATCATTTTTGCCTTG	CCCTGGAGGACTGAGGGT	57	130	14	22	10.178	27	12.581
MP17	pchgms1	(AC)12(AT)6	GGGTAAATATGCCCATTGTGCAATC	GGATCATTGAACTACGTCAATCCTC	63	194	2	23	8.568	29	12.452
MP19	pchgms3	(CT)14	ACGGTATGTCCGTACACTCTCCATG	CAACCTGTGATTGCTCCTATTAAAC	57	179	5	18	5.899	33	9.301
MP204	CPSCT012	GA	ACGGGAGACTTTCCCAGAAG	CTTCTCGTTTCCTCCCTCCT	62	156	14	26	11.398	32	14.269
MP206	CPSCT018	(CA)5(CT)20	AGGACATGTGGTCCAACCTC	GGGTTCCCCGTTACTTTCAT	52	162	5	26	9.374	36	13.062
MP226	CPDCT005	(CT)14	TTCAAGGAGAAGGCCTGAAA	ATTGTGGGTTCCAACCAATG	62	114	7	23	8.216	28	10.894
MP238	CPDCT025	(CT)10	GACCTCATCAGCATCACCAA	TTCCCTAACGTCCCTGACAC	62	172	2	22	9.630	29	12.990
MP244	CPDCT034	(CT)4-(CT)7-(CT)6	GAGAACCTTTTGTTTGGCCTTA	CGTCGTATTTAGTGCCGTTG	62	165	6	17	8.645	22	11.399
MP253	CPDCT045	(GA)16	TGTGGATCAAGAAAGAGAACCA	AGGTGTGCTTGCACATGTTT	62	142	3	27	13.099	35	17.823
MP307	EPDCU5100	GA	CTCTTCTCGCCTCCCAATTT	TGCTTAGCCCTGGGTACAAG	57	174	6	6	1.587	7	1.651
MP51	BPPCT001	(GA)27	AATTCCCAAAGGGATGTGTATGAG	CAGGTGAATGAGGCCAAAGC	57	159	7	24	7.791	31	11.066
MP59	BPPCT010	(AG)4 GG(AG)10	AAAGCACAGCCCATAATGC	GTACTGTTACTGCTGGGAATGC	57-60	131	5	25	9.910	30	12.836
MP66	BPPCT017	(GA)28	TTAAGAGTTTGTGATGGGAACC	AAGCATAATTTAGCATAACCAAGC	57-60	174	8	17	7.034	27	10.317
MP7	UDP96-018	(AC)21	TTCTAATCTGGGCTATGGCG	GAAGTTCACATTTAGGACAGGG	57	253	3	8	4.266	22	6.370
MP74	BPPCT025	(GA)29	TCCTGCGTAGAAGAAGGTAGC	CGACATAAAGTCCAAATGGC	57-60	197	15	22	12.610	28	15.962
MP76	BPPCT027	(GA)11	CTCTCAAGCATCATGGGC	TGTTGCCCGGTTGTAATATC	51	249	9	4	1.368	10	1.499
MP8	UDP96-003	(CT)11(CA)28	TTGCTCAAAAGTGTCGTTGC	ACACGTAGTGCAACACTGGC	57	143	11	18	6.845	32	10.403
MP85	BPPCT036	(AG)11	AAGCAAAGTCCATAAAAACGC	GGACGAAGACGCTCCATT	57-60	253	10	9	2.578	16	2.926
MP9	UDP96-005	(AC)16TG(CT)2CA(CT)11	GTAACGCTCGCTACCACAAA	CCTGCATATCACCACCCAG	57	155	6	25	12.141	39	16.776
							Mean	19.73913043	8.033	28.217	10.888
							SE	1.625	0.746	0.433	1.024
Markers excluded											
MP16	UDP98-408	(CT)14	ACAGGCTGTTGAGCATGTG	CCCTCGTGGGAAAATTTGA	57	100	23	-	-	-	-
MP245	CPDCT035	(GA)17	TCGAAGGAGGATGAAGTTGC	ATATCACGAGGGGGCAAAATG	62	146	18	-	-	-	-

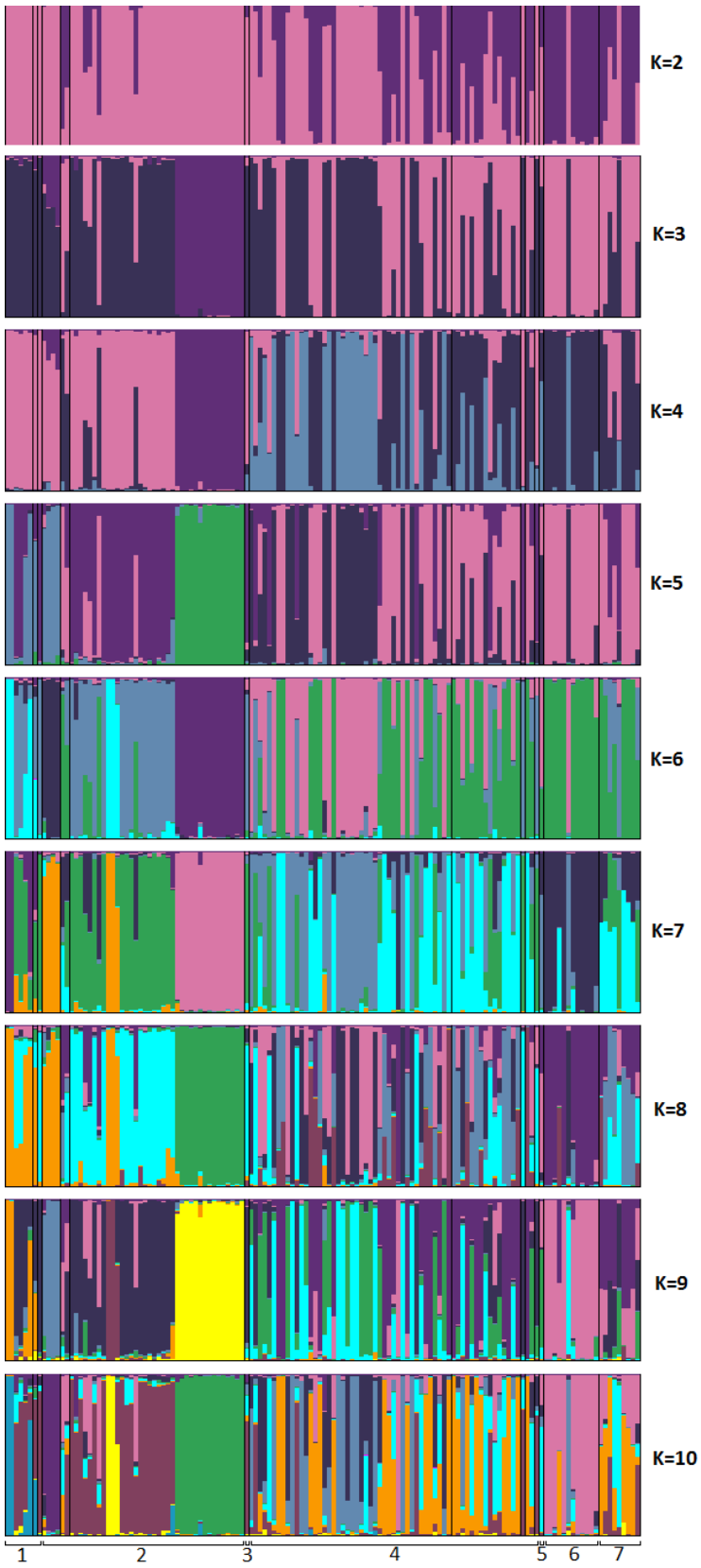


Figure S1 : Genetic subdivision of *P. dulcis* inferred with STRUCTURE for *K*=2 to *K*=10 and analysed with 23 microsatellite markers. The one hundred and thirty eight *P. dulcis* accessions include samples from Central Asia (Kyrgyzstan, Turkmenistan and Uzbekistan, *N*=8), Caucasia (Azerbaijan, Russia and Turkey, *N*=44), Middle East (Israel, *N*=1), Europe (Spain, France, Greece, Italy and Portugal, *N*=63), North Africa (Tunisia, *N*=1), North America (United States, *N*=12) and breeding accessions (*N*=9). Each individual was represented by a vertical bar and each color represented the proportion of membership in each genetic group *K* as implemented in STRUCTURE for each individual. (*) refers to Plum Pox virus resistant samples. The number in the legend correspond to (1) Central Asia, (2) Caucasia, (3) Middle East, (4) Europe, (5) North Africa, (6) North America and (7) Breeding accessions.

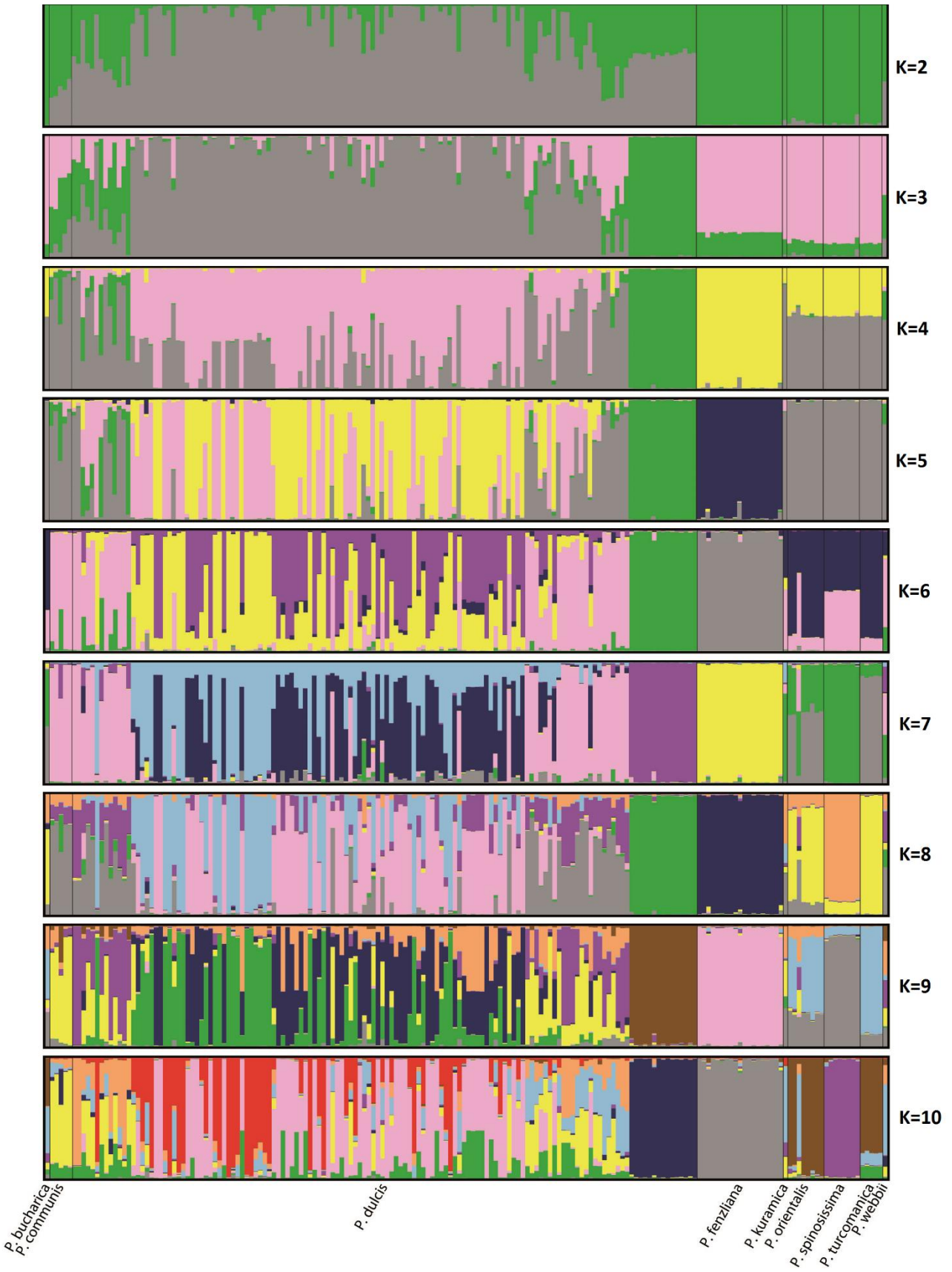


Figure S2 : Genetic subdivision of *P. dulcis* and wild related species inferred with STRUCTURE for *K*=2 to *K*=10 and with 23 microsatellite markers. The 186 belongs to nine different Prunus species including P. bucharica (N=1), *P. communis* (*N*=5), *P. dulcis* (*N*=138), *P. fenzliana* (*N*=19), *P. kuramica* (*N*=1), *P. orientalis* (*N*=8), *P. spinosissima* (*N*=8), *P. turcomanica* (*N*=5) and *P. webbii* (*N*=1). Each individual was represented by a vertical bar and each color represented the proportion of membership in each genetic group *K* as implemented in STRUCTURE for each individual. (*) refers to Plum Pox virus resistant samples.

Figure S3 : Genetic mapping of QTLs involved in viral accumulation : blue line represents peach homozygous alleles, the light pink line almond homozygous alleles and the dark pink the heterozygous state

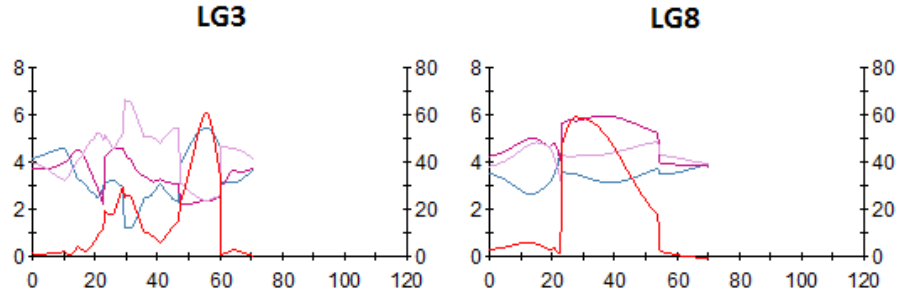


Figure S4 : Genetic mapping of QTLs involved in PPV symptoms intensity : blue line represents almond homozygous alleles, the light pink line peach homozygous alleles and the dark pink the heterozygous state

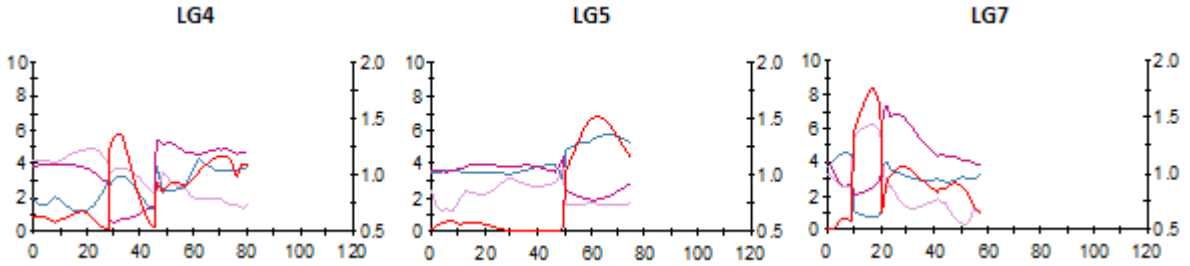
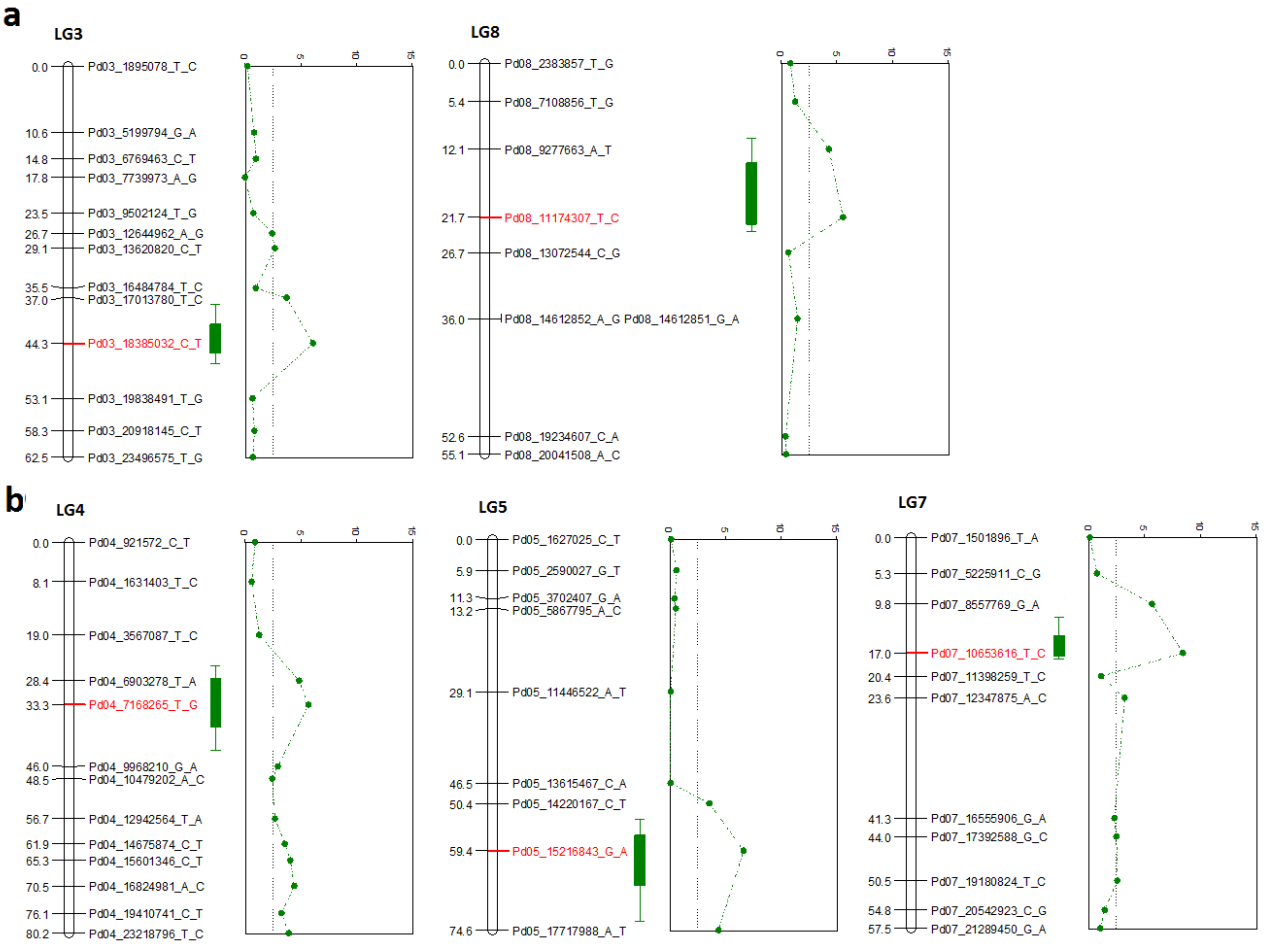


Figure S5 : : Genetic regions involved in (a) PPV accumulation and in (b) intensity of PPV symptoms in the 'Honey Blaze' x 'Del Cid' progenies obtained after alignment on the almond 'Texas' genome. The name of the markers are indicated on the right side of the linkage groups, the genetic distance (cM) on the left side of the linkage groups and LOD scores are on the right of the markers.



Abstract

Native from the mountainous regions of Central Asia, *P. dulcis*, commonly called almond, is the cultivated form of the Amygdalus species. Natural resistance to Plum Pox Virus has been shown to occur in almond, designating it as a potential donor of resistance for peach. In the current study, we determined the genetic diversity and subdivision of *P. dulcis* at the intraspecific level. We highlighted four differentiated clusters with contrasting geographical distributions, from Central Asia to North America, across Europe. One of the clusters, corresponding to samples from Akdamar Island (Turkey), is particularly distinct from the others. The origin of this population is puzzling, it might have diverged a long time ago, first by population isolation followed by local adaptation. We then included to our study representatives of almond wild related species. By Bayesian clustering with one of the four *P. dulcis* populations, we showed that *P. communis* is actually a feral form of *P. dulcis*. Our phylogenetic study illustrated the relationship between the *P. dulcis* and its wild relatives: *P. turcomanica, P. orientalis*, *P. spinosissima* and *P. fenzliana*. Among our Amygdalus wild and cultivated gene pool, PPV resistant accessions are affiliated to at least one of four possible genetic clusters, three *P. dulcis* and one *P. fenzliana*. The almond 'Del Cid' cultivar is one of the *P. dulcis* PPV resistant accessions. Starting from an F2 'Honey Blaze' x 'Del Cid' interspecific population, we identified two genomic regions involved in virus accumulation and three related to symptom severity. While being still preliminary, it is the first genetic mapping of PPV resistance determinants in almond that will require further confirmation in the next two vegetative cycles.
Keywords: *P. dulcis*, related species, diversity, Plum Pox Virus, resistance, QTL.

Résumé

Originaire des régions montagneuses d'Asie Centrale, *P. dulcis* est la forme cultivée de l'amandier, tel que nous le connaissons aujourd'hui. Des sources de résistance naturelle à la sharka existent chez l'amandier et pourront servir à introgresser ce caractère chez le pêcher. L'étude que nous avons réalisée a permis dans un premier temps de déterminer la diversité et la structuration génétique de *P. dulcis* au niveau intraspécifique. Nous avons ainsi mis en évidence quatre clusters génétiquement différenciés présentant une distribution géographique contrastée, allant de l'Asie Centrale à l'Amérique du Nord en passant par l'Europe. L'un de ces clusters, regroupant les individus de l'île d'Akdamar en Turquie, se distingue tout particulièrement des autres. Il semblerait que cette population ait divergé suite à un isolement géographique suivi d'une adaptation locale. Nous avons dans un second temps inclus dans notre analyse les espèces sauvages apparentées à l'amandier. A l'un des 4 groupes génétiques de *P. dulcis* se joint *P. communis* qui semblerait donc être une forme ensauvagée de l'amandier. Par une approche phylogénétique, nous avons pu démontrer la proximité génétique de certains amandiers avec les espèces sauvages : *P. turcomanica, P. orientalis, P. spinosissima* et *P. fenzliana*. Au sein de ce germoplasme cultivé et sauvage, les accessions résistantes à la sharka se répartissent en trois groupes génétiques de *P. dulcis* ainsi que dans l'espèce *P. fenzliana*. A partir de l'amandier européen 'Del Cid', résistant à la sharka, croisé avec le pêcher 'Honey Blaze', nous avons initié la cartographie génétique des facteurs de l'hôte contrôlant ce caractère. Une approche QTL (Quantitative Trait Loci) a permis d'identifier deux régions génomiques liées à l'accumulation virale et trois autres impliquées dans la sévérité des symptômes. Cependant, ces résultats demeurent préliminaires et nécessiteront d'être confirmés lors des deux prochains cycles végétatifs.
Mots-clés : *P. dulcis*, espèces apparentées, diversité, Plum Pox Virus, resistance, QTL.