

ORIGINAL ARTICLE

Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers

Malou Delplancke,¹ Nadir Alvarez,² Anahí Espíndola,² Hélène Joly,³ Laure Benoit,³ Elise Brouck¹ and Nils Arrigo⁴

¹ Université Montpellier 2, Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175, Montpellier cedex, France

² Department of Ecology and Evolution, Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

³ CIRAD, Centre d'Ecologie Fonctionnelle et Evolutive, Montpellier cedex, France

⁴ Department of Ecology & Evolutionary Biology, University of Arizona, Tucson, AZ, USA

Keywords

Amygdalus, gene flow, genetic diversity, introgression, microsatellites, *Prunus dulcis*.

Correspondence

Malou Delplancke, Université Montpellier 2, Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175, 1919 Route de Mende, 34293 Montpellier cedex 5, France.

Tel.: +33/0 4 67 61 22 58;

fax: +33/0 4 67 41 21 38;

e-mail: malou.delplancke@gmail.com

Received: 24 September 2011

Accepted: 2 November 2011

First published online: 16 December 2011

doi:10.1111/j.1752-4571.2011.00223.x

Abstract

Hybridization has played a central role in the evolutionary history of domesticated plants. Notably, several breeding programs relying on gene introgression from the wild compartment have been performed in fruit tree species within the genus *Prunus* but few studies investigated spontaneous gene flow among wild and domesticated *Prunus* species. Consequently, a comprehensive understanding of genetic relationships and levels of gene flow between domesticated and wild *Prunus* species is needed. Combining nuclear and chloroplastic microsatellites, we investigated the gene flow and hybridization among two key almond tree species, the cultivated *Prunus dulcis* and one of the most widespread wild relative *Prunus orientalis* in the Fertile Crescent. We detected high genetic diversity levels in both species along with substantial and symmetric gene flow between the domesticated *P. dulcis* and the wild *P. orientalis*. These results were discussed in light of the cultivated species diversity, by outlining the frequent spontaneous genetic contributions of wild species to the domesticated compartment. In addition, crop-to-wild gene flow suggests that ad hoc transgene containment strategies would be required if genetically modified cultivars were introduced in the northwestern Mediterranean.

Introduction

Gene flow, the movement of genes among lineages, plays an important role in the evolution of organisms by shuffling the genetic diversity within species (Rieseberg 1997; Petit and Excoffier 2009). Gene flow is quantitatively a major source of genetic variation within populations, thus acting as a primary force to balance the detrimental effects of genetic drift and maintain high effective population sizes (Lynch 2010). Because cultivated species have generally suffered strong bottlenecks through domestication (Doebly et al. 2006), gene flow involving wild species and their domesticated counterparts is valuable in the enrichment of their effective population sizes. Such genetic exchanges have long been reported and exploited by humans (Ellstrand et al. 1999). Wild species have been historically used as a source of genetic variation

for crop improvement programs, resulting in important applications for plant breeding (Papa 2005). For example, based on genetic evidence on grapevine, Myles et al. (2011) demonstrated that Western European *Vitis vinifera* cultivars experienced introgression from local Western European *Vitis sylvestris*. In addition, crop-to-wild gene flow has received growing attention in the last decade (Felber et al. 2007; Arrigo et al. 2011). The phenomenon has important evolutionary consequences for local relatives because it may promote the origin of highly competitive genotypes, resulting in the exclusion of vulnerable wild species (Ellstrand et al. 1999) or into the development of aggressive weeds (Trucco et al. 2009).

The Rosaceae family provides an excellent model for exploring gene flow between domesticated and wild species. Indeed, hybridization has played a central role in the evolutionary history of the family (Coart et al. 2006),

resulting in several large reticulated species complexes. Here, we focus on the almond tree, *Prunus dulcis* (Mill.) D.A. Webb (syn *Amygdalus communis* L. and *Prunus communis* Archang.), an economically important Rosaceae cultivated as a nut crop. The annual world production of almonds exceeds 1.83 million tons (FAO 2008), with half of the production located in California and the other half in Mediterranean Europe.

Almond trees belong to the subgenus *Amygdalus* (L.) Focke, an Irano-Turanian complex of *Prunus* including more than 30 species (Browicz and Zohary 1996) that radiated recently (Ladizinsky 1999; Potter et al. 2002; Yazbek, unpublished data). Although several *Amygdalus* species have been sporadically used for human consumption, only *P. dulcis* was domesticated to produce sweet almonds. The spatio-temporal origin of domestication is still controversial although several lines of evidence suggest that *P. dulcis* domestication originated in the Fertile Crescent during the first half of the Holocene (Browicz and Zohary 1996; Ladizinsky 1999; Willcox et al. 2009; Delplancke 2011). Archeobotanic remains of *P. dulcis* show that almond trees were already cultivated about 11 000 years ago (Willcox et al. 2008) and used throughout the Near East, complementing meat and other plant food (Martinoli and Jacomet 2004).

Prunus orientalis (Duhamel) is one of the wild counterparts of the cultivated almond tree. This taxon is one of the most common *Amygdalus* representatives occurring in the Near Eastern Mediterranean. It is widespread from northeast Iraq to south and central Anatolia, and commonly grows in contact with *P. dulcis* orchards. Because it shows substantial genetic differentiation with almond trees, *P. orientalis* is not considered as the sole potential wild ancestor of *P. dulcis* (Zeinalabedini et al. 2010). The ancestry of the latter species remains controversial, with a probable diffuse domestication process featuring several wild species that contributed to its current genetic pool (see Zeder 2006).

Several hybridization events involving the almond tree and wild relatives from the *Amygdalus* group have been reported. For instance, spontaneous wild-to-crop gene flow was detected in several Italian almond orchards, in which self-compatibility (i.e., species are otherwise self-incompatible) and specific morphological characters had presumably been introgressed from *Prunus webbii* (Spach) (Socias i Company 1998; Godini 2000). Moreover, crop-to-wild exchanges have long been suspected because of the wide range of intermediate phenotypes observed throughout western and central Asian species (Grasselly 1977; Grasselly and Crossa-Raynaud 1980; Denisov 1988; Browicz and Zohary 1996; Gradziel 2009). Such pervasive gene flow is consistent with the mating system (i.e., self-incompatibility), insect-mediated pollination (Dicenta and

Garcia 1993; Socias i Company 1998), and the perennial life cycle of almond, which promotes outcrossing and hybridization (Goodwillie et al. 2005; Petit and Hampe 2006). Finally, gene flow among *Amygdalus* taxa might be facilitated because a large proportion of the group shares a similar diploid chromosome number ($2n = 16$ chromosomes), including *P. dulcis* and *P. orientalis* (Grasselly 1977; Corredor et al. 2004), which may lead to viable hybrids (Browicz and Zohary 1996). In the current context of global genetic erosion, there is urgent need for a comprehensive understanding of the genetic relationships and the coexistence between cultivated, feral, and wild *Prunus* species in their centers of origin.

Because the reality of genetic exchanges between the cultivated form *P. dulcis* and one of its most widespread wild relatives *P. orientalis* has never been examined, we investigate gene flow and hybridization between these two key almond species in the Fertile Crescent, their supposed native area, using nuclear and chloroplastic markers. We outline the reciprocal genetic contributions of one species to the other, relying on microsatellite genotyping [hereafter simple sequence repeat (SSRs)], a category of molecular markers widely used for investigating evolutionary relationships between lineages having diverged recently. Moreover, by combining highly polymorphic, bi-parentally inherited nuclear SSRs with nonrecombinant, maternally inherited chloroplastic SSRs, we aim to

- 1 Assess whether nuclear and chloroplastic microsatellites are efficient markers for delineating species in the *Amygdalus* complex.
- 2 Characterize the genetic diversity of *P. dulcis* and *P. orientalis*.
- 3 Investigate the relative genetic contribution of a common wild species to the gene pool of the cultivated almond crops.
- 4 Assess the level of crop-to-wild gene flow.

Material and methods

Sampling and plant material

A total of 428 and 134 individuals were collected for *P. dulcis* and *P. orientalis*, respectively (Table 1). Sampling included 24 traditional orchards of *P. dulcis* and seven spontaneous wild populations of *P. orientalis*, located in the western part of the Fertile Crescent (i.e., Lebanon, Turkey, and Syria). Species were identified using morphological characters described by Browicz and Zohary (1996). The domesticated species *P. dulcis* is a nonspiny tree, containing numerous brachyblasts (i.e., short shoots) bearing relatively large leaves. Its wild counterpart, *P. orientalis*, is a smaller and subspinescent shrub, characterized by white tomentose shoots, leaves,

Table 1. Sampling effort. The taxonomic identification of specimens, the population type (allopatric or sympatric) and name, its geographic coordinates, and the number of genotyped specimens are provided for each surveyed populations and for each marker type.

Group	Population	Country	Latitude (° N)	Longitude (° E)	Nuclear SSRs	Chloroplast SSRs
Allopatric <i>Prunus dulcis</i> (D)	Al Shahar	Syria	32.66	36.64	7	3
	Banian	Syria	35.16	35.99	5	2
	Bire Akaar	Lebanon	34.59	36.24	5	1
	Chenkoy	Turkey	36.04	36.14	25	5
	Dahr al Djabal	Syria	32.67	36.66	22	8
	El Nebi Habil	Syria	33.60	36.06	5	5
	Ergani	Turkey	38.29	39.74	27	3
	Esenpinar	Turkey	36.57	34.11	28	5
	Ferzol	Lebanon	33.89	35.94	14	4
	Fhela	Syria	34.60	36.90	15	5
	Hior Al Louz	Syria	32.73	36.66	16	6
	Jourt Hatar	Syria	32.65	36.60	2	1
	Kirikhan-Hassa	Turkey	36.65	36.45	25	5
	Kosaya	Lebanon	33.82	36.02	9	5
	Lucy	Lebanon	33.67	35.86	32	6
	Nevsehir	Turkey	38.60	34.72	29	4
	Shaat	Lebanon	34.13	36.24	17	9
	Sirgaya	Syria	33.81	36.15	22	7
	Zegrine	Lebanon	34.41	36.32	19	4
	Subtotal				324	88
Sympatric <i>P. dulcis</i> (Ds)	Ciftehan	Turkey	37.52	34.74	29	5
	Irsal	Lebanon	34.19	36.39	16	12
	Labweh	Lebanon	34.17	36.30	28	16
	Seyitusagi	Turkey	38.28	38.20	24	4
	Waadi Kafar Soun	Syria	35.81	36.51	7	4
	Subtotal				104	41
Sympatric <i>Prunus orientalis</i> (Os)	Ciftehan	Turkey	37.52	34.70	24	5
	Irsal	Lebanon	34.19	36.39	24	24
	Labweh	Lebanon	34.17	36.30	8	7
	Seyitusagi	Turkey	38.28	38.20	25	5
	Waadi Kafar Soun	Syria	35.74	36.48	4	4
	Subtotal				85	45
Allopatric <i>P. orientalis</i> (O)	Gazentep	Turkey	37.06	37.53	24	3
	Göreme	Turkey	38.64	34.85	25	0
	Subtotal				49	3
	Total				562	177

SSRs, simple sequence repeats.

and fruits (Fig. 1). Allopatric and sympatric populations of *P. dulcis* and *P. orientalis* were sampled by collecting fresh young leaves (dried and stored in silica gel) from each individual, and by registering the global positioning system coordinates of each sampled tree. Allopatric populations included 324 individuals for *P. dulcis* (19 populations, referred to as the '*P. dulcis* group', hereafter 'D') and 49 for *P. orientalis* (two populations referred to as the '*P. orientalis* group', hereafter 'O'), respectively. Both species co-occurred in five sampling sites, representing 104 individuals for *P. dulcis* (referred to as the '*P. dulcis* sympatric' group, hereafter 'Ds') and 85 for *P. orientalis* (referred to as the '*P. orientalis* sympatric' group, hereafter 'Os').

SSR genotyping

DNA was extracted from silica-dried leaves using the DNeasy® 96 plant kit (Qiagen, Hilden, Germany) with two modifications: samples were lysed 2 h at 65°C, and DNA was eluted in 200 µL of buffer AE. SSR amplification was performed in 96-well plates in a Mastercycler (Eppendorf, Hamburg, Germany) with the following parameters: 10 min at 95°C, 45 cycles of 30 s at 94°C, 90 s at 58°C, 90 s at 72°C, and 30 min at 72°C. Individuals were randomly distributed on PCR plates, with three wells per plate used as negative and migration controls.

All individuals were genotyped with twelve nuclear microsatellite loci (UDP96-001, UDP96-018, UDP96-003,



Figure 1 Specimens of *Prunus dulcis*, *Prunus orientalis* and putative hybrid, with focus on diagnostic morphological characters. The domesticated *P. dulcis* shows a tree habitus (A) and large green leaves (B). Putative hybrid shows an intermediate phenotype with large green and tomentose leaves (C). In contrast, the wild *P. orientalis* is a shrub (D, foreground) with tomentose leaves (E) and thorny shoots (F).

UDP97-401, UDP98-408, UDP98-409, pchgms1, pchgms3, BPPCT017, BPPCT001, BPPCT007, and BPPCT025; Cipriani et al. 1999; Testolin et al. 2000; Dirlewanger et al. 2002), using four multiplexed PCRs. As chloroplasts have smaller effective population sizes than nuclear genomes (Powell et al. 1995) and may thus harbor lower variation at the intrapopulation level because of quicker allele fixation, a subsample of 177 of the 562 individuals was selected by optimizing the geographic coverage of the samples and was further genotyped with ten chloroplastic SSRs (TPScp1, TPScp2, TPScp3, TPScp4, TPScp5, TPScp7, TPScp8, TPScp9, TPScp10, and TPScp11; Ohta et al. 2005), using two multiplexed PCRs. Five loci (TPScp1, TPScp 2, TPScp4, TPScp7, and TPScp8) were later excluded from the dataset, because of a high missing value percentage that demonstrated amplification difficulties and/or presence of null alleles. For nuclear and chloroplastic markers, multiplexed PCRs were carried out with the Type It Microsatellite PCR Kit® (Qiagen) in a final volume of 10 μ L, containing 1 \times of Qiagen Master Mix, 0.2 μ M of a primers' mix (0.4 μ M for the chloroplast primers) and 2 ng/ μ L of template DNA. The reproducibility of reactions was checked using samples replicated on the different plates and reached 95.1% (i.e., the error rate was <0.05). GeneScan was performed on an ABI 3130 XL 16 capillary-sequencer (ABI Prism Applied Biosystems, Foster City, CA, USA), and allele calling was performed by two independent investigators using Genemapper (Applied Biosystems).

Insights on the genetic diversity using nuclear and chloroplastic data

For nuclear SSRs, allelic richness (Ar, the number of distinct alleles), allele size range (Range), observed heterozygosity (Ho), and unbiased heterozygosity (He) were estimated for each population including more than 15 individuals (Table S1). For chloroplastic SSRs, the number of distinct alleles (Ar), effective number of alleles (Ne, computed as $1/\sum p_i^2$ with p_i as the frequency of the i th allele), and the unbiased expected heterozygosity (H) were estimated for each species. The diversity estimates relied on a rarefaction procedure to obtain comparable sampling efforts among groups (Petit et al. 1998), (15 individuals per populations for the nuclear dataset and 48 individuals per species for the chloroplastic dataset, respectively). The estimations were averaged from 1000 resampled datasets and computations were performed using custom R scripts (R Development Team, 2011) (available from the first author, on request). The results were averaged at the species level for the nuclear datasets. Statistical differences among species diversities were assessed by random permutation tests (1000 permutations, using custom R CRAN scripts available from the first author). Finally, the partitioning of genetic variation among species (i.e., *P. dulcis* versus *P. orientalis*) and sampling sites was quantified with a hierarchical analysis of molecular variance (AMOVA, significance levels were tested with 1000 permutations), using Arlequin (Excoffier

and Lischer 2010). Here as well, only sampled sites with more than 15 individuals were included in the analysis.

Species boundaries and Bayesian estimation of admixture levels

Species boundaries were investigated using three distinct approaches. First, the chloroplastic SSRs were analyzed with a median joining network of haplotypes, using *NETWORK* 4.5.6.1 (Bandelt et al. 1999). Second, the nuclear SSRs were investigated using a principal component analysis performed among specimens, using the 'Ade4' *R* CRAN package (R, Core Development Team 2011). Third, for nuclear and chloroplastic SSRs, admixture proportions of *P. dulcis* and *P. orientalis* samples were estimated using a model-based Bayesian clustering of individuals, as implemented in *STRUCTURE* 2.3.1 (Pritchard et al. 2000). This software uses a MCMC framework, in which the algorithm explores a parameter space considering individual admixture proportions, locus-specific ancestries, population allele frequencies, and the expected admixture of the dataset, assuming an user-defined K number of groups. The likelihood of each iteration was then evaluated by computing the probability of the model predictions given the empirical data (the computation assumes Hardy–Weinberg equilibrium within the K groups). The MCMC algorithm was set up for 200 000 burn-in steps (i.e., an initiation phase without results recording), followed by 1 000 000 steps for data acquisition (the remaining parameters were left as default values), assuming admixture. Each analysis was replicated ten times, and only runs with the highest maximum-likelihood values were kept for further investigations. The computations considered K values ranging between two and ten groups, and the optimal number of groups was assessed using the deltaK criterion (Evanno et al. 2005; Fig. S1). The same procedure was applied for the chloroplastic (129 and 48 individuals for *P. dulcis* and *P. orientalis*, respectively) and the nuclear (428 *P. dulcis* and 134 *P. orientalis* individuals) datasets, with haploid and diploid parameterization of the model, respectively. The *STRUCTURE* outputs were handled using the *SIMIL* R script collection (Alvarez et al. 2008).

Coalescent models to estimate population sizes and gene flow

The effective population size of *P. dulcis* and *P. orientalis* and the magnitude of gene flow among species were inferred using coalescent-based methods implemented in *Migrate-n* (Beerli and Felsenstein 2001). Analyses were performed using a subset of six nuclear SSRs that followed a stepwise mutation model (i.e., UDP-408, pchms3, BPPCT007, UDP96-018, UDP96-003, and BPPCT017; the

stepwise mutation model assumption was checked by performing the frequency distribution of allele size – in repeat length – for each loci) and assumed two populations (i.e., *P. dulcis* versus *P. orientalis* specimens, as defined from morphological identifications). Using a maximum-likelihood approach and applying the MCMC search implemented in *Migrate-n*, the algorithm estimates the approximation of the effective population size of each species (as $\Theta = 4N_e\mu$, the effective population size scaled by the mutation rate) and their reciprocal gene flow (as $M = m/\mu$ the migration rate scaled by the mutation rate). The heuristic searches relied on a preliminary run to refine the parameter search space (the initial values of Θ and M were estimated from FSTs, the searches included ten short chains of 1×10^6 generations with 5000 recorded genealogies followed by four long chains of 25×10^6 generations with 50 000 recorded genealogies, and a burnin of 10 000 generations was applied). Demographic parameters and their statistical significance were estimated from five additional independent runs (hereafter the 'final runs') that were initiated using estimations obtained from the preliminary run. Accordingly, Θ and M were initiated using normal distributions (mean = 7 and 18, standard deviation = 1 and 2, for Θ and M , respectively) and searches included five short chains of 12×10^5 generations with 1000 recorded genealogies, followed by two long chains of 2×10^6 generations with 10 000 recorded genealogies (a burnin of 1×10^6 generations was applied). All chains used the Brownian motion approximation as the mutation model and relied on adaptive heating to maximize the visited space (default parameters). The convergence of chains within runs was assessed with the Gelman–Rubin criterion (default parameter). The estimates were assumed as accurate when the 99% confidence intervals of demographic parameters were overlapping in at least two final runs. Finally, the statistical significance of gene flow was assessed with likelihood ratio tests that relied on alternative models considering M either as absent (i.e., $M_{\text{dulcis-to-orientalis}} = M_{\text{orientalis-to-dulcis}} = 0$) or asymmetric (i.e., $M_{\text{dulcis-to-orientalis}}$ and $M_{\text{orientalis-to-dulcis}}$ were alternatively set to null). The complete test procedure was performed in *Migrate-n*.

Results

Genetic diversity of almond trees, as revealed by nuclear and chloroplast markers

The nuclear SSRs (Table 2) revealed high levels of genetic diversity ($Ar = 8.74 \pm 1.24$ alleles per loci, Range = 31.70 ± 2.74) and high heterozygosities ($Ho = 0.73 \pm 0.06$ and $He = 0.85 \pm 0.05$). In contrast, chloroplast markers showed limited polymorphism ($Ar = 2.65 \pm 1.67$ alleles per loci) and low levels of effective number of alleles ($Ne = 1.61 \pm 0.66$).

Table 2. Genetic diversity of *Prunus dulcis* and *Prunus orientalis*, as revealed by nuclear and chloroplast SSRs. The averaged and standard deviation (between parenthesis) across all loci are provided.

	Ar _{nuc} *	Range _{nuc} *	Ho _{nuc} *	He _{nuc} *	N _{cp} †	Ar _{cp} †	Ne _{cp} †	H _{cp} †
<i>P. dulcis</i>	8.61 (1.00)	32.93 (2.45)	0.73 (0.06)	0.85 (0.04)	129	2.89 (1.82)	1.40 (0.37)	0.24 (0.19)
<i>P. orientalis</i>	8.87 (1.47)	30.46 (3.02)	0.73 (0.05)	0.85 (0.05)	48	2.40 (1.67)	1.82 (0.86)	0.34 (0.32)
Global	8.74 (1.24)	31.70 (2.74)	0.73 (0.06)	0.85 (0.05)	177	2.65 (1.67)	1.61 (0.66)	0.29 (0.25)

Ar, allelic richness; range, allele size range; Ho, observed heterozygosity; He, expected heterozygosity; Ne, effective number of alleles; H, unbiased expected heterozygosity; SSRs, simple sequence repeats.

The sampling effort (N) and several diversity estimates are provided.

*Diversity estimates from nuclear SSR markers.

†Diversity estimates from chloroplastic SSR markers, $N_e = 1/\Sigma(p_i^2)$ and $H = 1 - \Sigma(p_i^2)$, where p_i is the frequency of the i th allele.

Both *Prunus* species appeared similarly diversified (Table 2 and Table S1), as attested by nuclear (Ar_{nuc} = 8.61 ± 1.00 and 8.87 ± 1.47, Ho_{nuc} = 0.73 ± 0.06 and 0.73 ± 0.05, and He_{nuc} = 0.85 ± 0.04 and 0.85 ± 0.05 for *P. dulcis* and *P. orientalis*, respectively) and chloroplastic markers (Ar_{cp} = 2.89 ± 1.82 and 2.40 ± 1.67, Ne_{cp} = 1.40 ± 0.37 and 1.82 ± 0.86, H_{cp} = 0.24 ± 0.19 and 0.34 ± 0.32 for *P. dulcis* and *P. orientalis*, respectively). No significant differences among species were detected for any diversity estimate.

The AMOVA outlined similar patterns in the genetic partition of *P. dulcis* and *P. orientalis* (Table 3). Both species showed similar percentages of genetic diversity throughout the investigated variation levels (FSC = 0.09 and 0.11, FIS = 0.11 and 0.10, FIT = 0.18 and 0.20, for *P. dulcis* and *P. orientalis*, respectively). In addition, most of the variation occurred within individuals, representing 81.55% and 80.09% of the genetic diversity within *P. dulcis* and *P. orientalis*, with only 3.07%, 8.78%, and 9.32% of the diversity occurring among species, among populations, and among individuals, respectively. Finally, species differentiation was significant but low, with genetic variance among species (FCT = 0.03) being lower than genetic variance among populations within species (FSC = 0.09).

Species limits and gene flow among almond trees

The chloroplast and nuclear datasets differed in their ability to discriminate *P. dulcis* from *P. orientalis* specimens. The chloroplast SSRs revealed 18 haplotypes (i.e., Fig. 2A H1 to H18), among which five were shared by both species (H5, H8, H9, H12, and H17) and were found in 76% of the global dataset (including respectively 85% of *P. dulcis* and 52% of *P. orientalis* specimens). The remaining haplotypes were species-specific and showed lower frequencies (except H7 that occurred in 33% of the specimens of *P. orientalis*). Notably, both species differed slightly in terms of private haplotypes (H1, H2, H3, H4, H6, H11, H14, H15, and H16 versus H7, H10, and H18, for *P. dulcis* and *P. orientalis*, respectively). In addition, *P. orientalis* included several haplotypes that were frequent but geographically restricted (H5, H7, H8, and H9), while *P. dulcis* included a frequent widespread haplotype (H5) along with several rare and geographically restricted variants (H1, H2, H3, H6, H11, H14, H15, H16, and H17). Finally, the median joining network (Fig. 2A), outlined unclear genetic limits among both *Prunus* species. These results were corroborated by the Bayesian clustering of specimens based on chloroplast SSRs (optimal K value determined following Evanno's

Table 3. Analysis of molecular variance of nuclear simple sequence repeats, considering the partitioning of four levels of genetic variation (a–d) for *Prunus dulcis* and *Prunus orientalis*.

Variation level	Species	Sum of squares	Variance component	Variation percentage	Fixation indices
a. Among species	Global	91.09	0.17	3.07	FCT = 0.03***
b. Among population within species	<i>P. dulcis</i>	389.51	0.46	8.58	FSC = 0.09***
	<i>P. orientalis</i>	146.59	0.58	10.66	FSC = 0.11***
	Global	536.11	0.49	8.78	FSC = 0.09***
c. Among individuals within pop.	<i>P. dulcis</i>	1857.99	0.53	9.87	FIS = 0.11***
	<i>P. orientalis</i>	645.85	0.51	9.24	FIS = 0.10***
	Global	2503.85	0.52	9.32	FIS = 0.11***
d. Within individuals	<i>P. dulcis</i>	1573.00	4.35	81.55	FIT = 0.18***
	<i>P. orientalis</i>	554.50	4.39	80.09	FIT = 0.20***
	Global	2127.50	4.37	78.83	FIT = 0.21***

Significance levels were tested with 1000 per mutations, ***significant at 0.001.

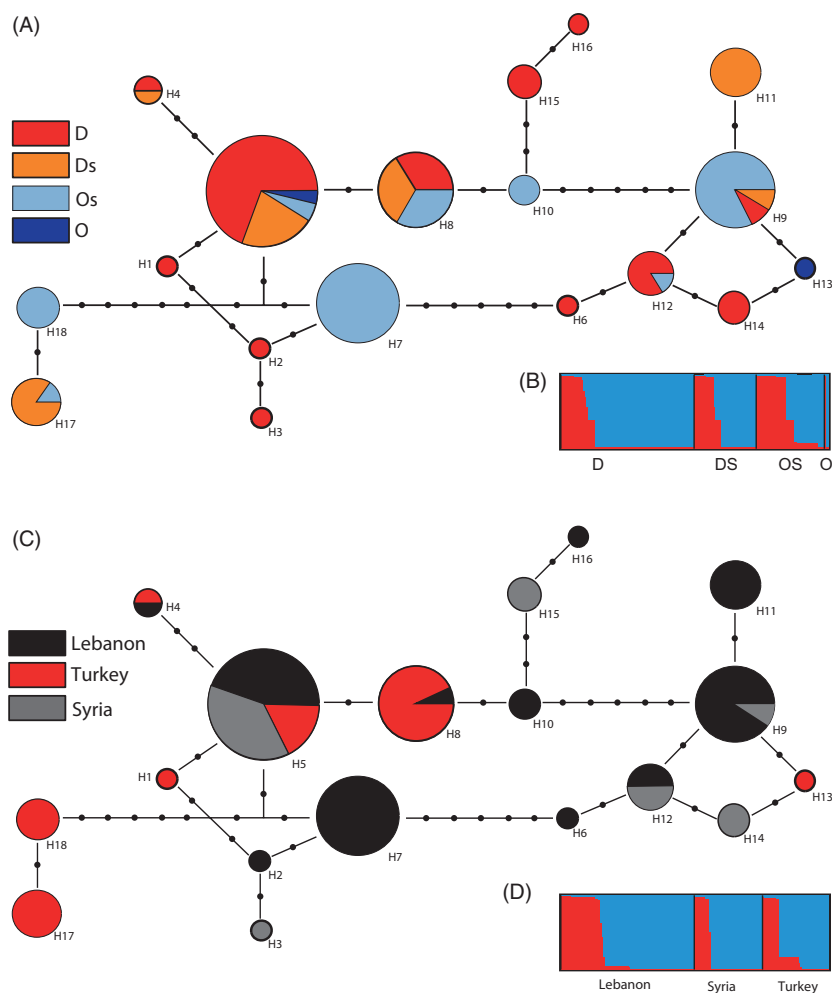


Figure 2 Insights from the chloroplast simple sequence repeats as revealed by median joining networks (A and C) and model-based Bayesian clustering of specimens (B and D). On median joining networks, 18 distinct haplotypes are displayed as pie-charts reflecting the proportions of specimens occurring in (A) allopatric or sympatric populations (i.e., D/Ds and O/Os for *Prunus dulcis* and *Prunus orientalis*, respectively) and in (C) the three surveyed countries (i.e., Lebanon, Syria and Turkey). The radius of pies reflects the frequency of haplotypes in the global dataset, dots along edges correspond to mutational steps. Results from the model-based Bayesian clustering are displayed using barplots. Specimens were assigned to $K = 2$ genetic groups, defined using STRUCTURE. Each specimen is represented as a vertical bar where blue or red sectors reflect assignment probabilities to each of the two groups. Specimens are sorted according to their taxonomical status (B) or their geographical origin (D).

method, see further details in Fig. S1). Still, although chloroplast haplotypes did not delineate clear species limits, they revealed consistent phylogeographic patterns (Fig. 2C). Haplotypes from Turkey (H1, H8, H17, and H18) were genetically distant from those mostly occurring in Syria and Lebanon (H2, H3, H6, H7, H9, H10, H11, H12, H14, H15, and H16). The remaining haplotype (H5) was mostly observed in *P. dulcis* specimens and showed the widest geographic distribution.

Nuclear SSRs revealed a clear genetic differentiation between *P. dulcis* and *P. orientalis* (Fig. 3). Indeed, both species were discriminated along the first two eigenaxes of the principal components analysis (Fig. 3A, accounting

for 24% of the observed variance). These results were largely corroborated by the Bayesian clustering of specimens (Fig. 3B,C), where the deltaK spectrum (following Evanno's method) identified $K = 3$ as the most likely number of groups (see further details in Fig. S1). Intra-specific patterns were revealed for *P. dulcis* where specimens were split into two groups; one geographically widespread and another restricted to Syria and Lebanon (Fig. S2). This signal could partly reflect the Lebanon and Taurus Mountains biogeographic splits (respectively parallel to the Mediterranean coast from southern Lebanon into Syria and extending from eastern to south-western Turkey). Finally, *P. orientalis* specimens were

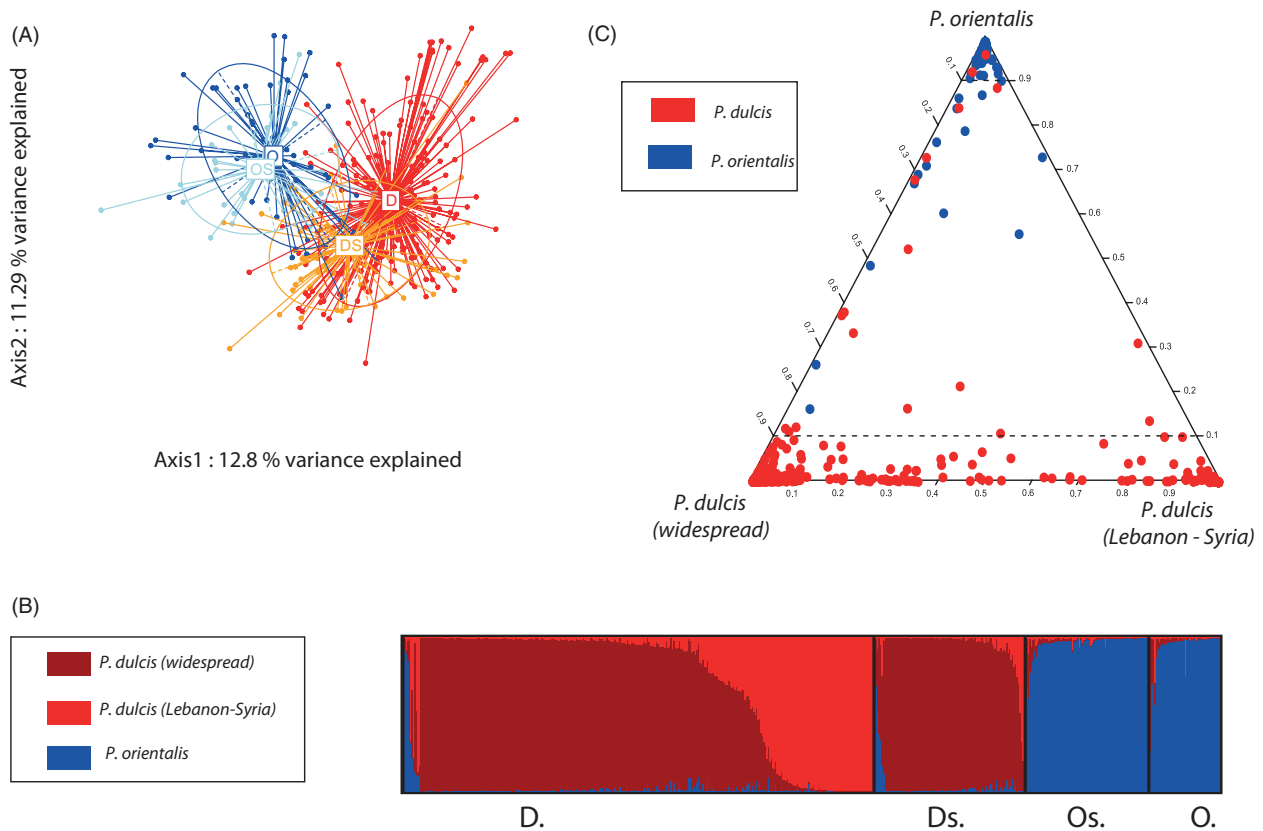


Figure 3 Focus on nuclear simple sequence repeats to highlight gene flow between *Prunus dulcis* and *Prunus orientalis*. (A) Principal components analysis of genotypes. (B) Admixture proportions of specimens –as estimated using STRUCTURE with $K = 3$ groups– displayed in a barplot. Each specimen is represented as a vertical bar, with blue, light red, dark red sectors reflecting assignment probabilities to each of the $K = 3$ groups. (C) Admixture proportions of specimens (figured as red or blue dots for *P. dulcis* and *P. orientalis*) to each of the $K = 3$ groups, are displayed as a ternary plot. Each specimen is represented as a point, with blue or red color reflecting the taxonomic status. For (A), (B) and (C), the specimens are labeled according to their taxonomical status and/or their origin (i.e. D/Ds and O/Os correspond to *P. dulcis* and *P. orientalis* specimens collected in allo- or sympatric populations, respectively).

included in a single group. With respect to interspecific variation, *P. dulcis* and *P. orientalis* were discriminated from each other throughout the Bayesian clusters, but species limits appeared as fuzzy in several cases. For instance, two specimens identified morphologically as *P. dulcis* were assigned to the *P. orientalis* genetic pool by the Bayesian clustering. Moreover, several specimens were assigned to both genetic pools (e.g., 14 and 18 specimens of *P. dulcis* and *P. orientalis*, respectively) with assignment probabilities ranging between 0.1 and 0.9. These specimens were collected either in sympatric or allopatric locations (Fig. 3C).

The magnitude of gene flow, as well as the effective population sizes of *P. dulcis* and *P. orientalis*, estimated using a coalescent-based approach are summarized in Table 4. The estimates obtained from four out of five final runs showed overlapping confidence intervals and only the most likely results were provided. The analysis revealed

large and comparable effective population sizes for both species ($\Theta_{\text{dulcis}} = 7.33$ and $\Theta_{\text{orientalis}} = 7.11$). Furthermore, substantial gene flow among species was detected, with $M_{\text{orientalis-to-dulcis}} = 16.96$ and $M_{\text{dulcis-to-orientalis}} = 15.64$. In addition, likelihood ratio tests outlined gene flow as a significant parameter: scenarios considering absent or asymmetric gene flow produced significantly less explicative models than the full model allowing symmetric gene flow (Table 4).

Discussion

Relative performance of nuclear and chloroplastic molecular markers in discriminating species

Chloroplastic and nuclear markers differed in their ability to discriminate *Prunus* species, a result corroborating other studies that outlined the limited taxonomic resolution of chloroplastic markers, when compared to the

Table 4. Maximum-likelihood estimates of effective sizes (Θ) and gene flow (M) of *Prunus dulcis* and *Prunus orientalis* populations, as revealed by nuclear SSRs. The most likely estimates and 95% confidence intervals (between parenthesis) over all loci are provided for (a) the full or (b) partial models. The partial models consider that gene flow is either totally absent or asymmetric.

Model*	$\Theta_{dulcis}^{\dagger}$	$\Theta_{orientalis}^{\dagger}$	$M_{dulcis-to-orientalis}^{\ddagger}$	$M_{orientalis-to-dulcis}^{\ddagger}$	AIC§	Log-likelihood§	P-value§
a. Full model							
With gene flow	7.33 (0.48)	7.11 (0.63)	15.64 (3.79)	16.96 (2.13)	273.78	−132.89	–
b. Partial models							
Asymmetric gene flow	7.33	7.11	0	16.96	195047.56	−97520.78	1×10^{-6}
Asymmetric gene flow	7.33	7.11	15.64	0	225026.45	−112510.23	1×10^{-6}
Without gene flow	7.33	7.11	0	0	420889.73	−210442.87	1×10^{-6}

SSRs, simple sequence repeats.

*The analysis includes 428 *P. dulcis* and 134 *P. orientalis* specimens; the estimates are computed from a subset of six nuclear SSRs that follow a stepwise mutation model.

$\dagger\Theta = 4N_e\mu$ is an estimate of the effective population size, scaled by the mutation rate (μ).

$\ddagger M_{source-to-sink} = m/\mu$ is an estimate of the gene flow magnitude among source and sink populations, scaled by the mutation rate (μ). The number of immigrants in a sink population can be computed as $N_e m = \frac{1}{4}\Theta_{sink} \times M_{source-to-sink}$.

§Likelihood ratio test. The Akaike information criterion (AIC), the log-likelihood and the P-value (Ho: Partial = Full model) of models are provided for either the full or partial (i.e., constraining gene flow to null) models.

nuclear ones (Rieseberg et al. 1996). Nuclear SSRs provided robust and highly informative signals, consistent with earlier studies showing the suitability of such markers to discriminate species and detect hybridization among *Prunus* species (Cipriani et al. 1999). In contrast, chloroplastic SSRs had a moderate transferability among species and showed low levels of polymorphism (Table 2). Furthermore, the detected haplotypes were more informative about phylogeographic patterns than for delimiting species boundaries (Fig. 2). These results could reflect limitations in taxonomic (177 individuals investigated with chloroplastic SSRs versus 562 for the nuclear SSRs) or genetic sampling (only four polymorphic markers), possibly resulting in underestimated levels of species differentiation. More likely, pervasive gene flow among species, with chloroplastic lineages evolving largely independently from species boundaries, but constrained by geographic features, might explain our results. Such patterns were indeed already reported from other plant species (Petit et al. 2005) and could result from unbalanced contributions of pollen and seeds to gene flow – a process causing more reticulated signals for chloroplast than nuclear markers (Petit et al. 2005). As a consequence, we mainly focused on nuclear genotypes to detect interspecific gene flow.

Genetic diversity of domesticated and wild almond trees

Our investigation revealed genetic diversity levels in domesticated almond trees that were as high as those reported from wild tree species. Indeed, *P. dulcis* and its wild counterpart, *P. orientalis* appeared as highly heterozygous (average of 0.73 and 0.85 for Ho and He, respectively, Table 2). Furthermore, genetic diversity was

higher within individuals (78.83%) than between populations (8.78%), a pattern similar to that observed for other tree species that might be explained by both high level of pollen flow and life cycle characteristics of trees (juvenile phase and overlapping generations, Austerlitz et al. 2000). These results were also consistent with the self-incompatible mating system of both species. In addition, *P. dulcis* and *P. orientalis* appeared as similarly diversified (Table 2) when considering the number of alleles (Na), allelic richness (Ar), heterozygosities (He and Ho), and coalescent-based estimations of effective population sizes ($\theta = 7.33$ and 7.11 for *P. dulcis* and *P. orientalis*, respectively). Finally, both species differed slightly in terms of phylogeographic structures, with a stronger regional differentiation observed in the domesticated *P. dulcis* (See Fig. S2). Our results thus contrasted with the high level of diversity loss usually observed in many annual seed-propagated crops (maize: Matsuoka et al. 2002; common bean: Papa et al. 2007; wheat: Haudry et al. 2007) but were congruent with insights revealed from perennial crops. For example, several cultivated perennials retained relatively high variation throughout domestication (e.g., tropical fruit trees, Hollingsworth et al. 2005; grapevine, Myles et al. 2011). Our results also suggested that the domestication of almond trees might not have suffered a substantial reduction in the original gene pool, because extant almond trees appeared as diversified as other wild *Prunus* species. This hypothesis would require further investigation (e.g., using sequence data). Indeed, the signature of bottlenecks might have been underestimated by our SSR markers, owing to their high mutation rates that are able to quickly replenish diversity losses typical of domestication processes (Glemin and Bataillon 2009). Alternatively,

cultural practices might also partly explain our results. Almond orchards are propagated through seed reproduction, a strategy that has probably maintained high diversity levels in the domesticated forms (Grasselly and Crossa-Raynaud 1980). Finally, gene flow between crop and wild species could also account for the high observed diversities (e.g., Mariette et al. 2010).

Contributions of wild *Prunus orientalis* to domesticated cultivars

Our results clearly outlined ongoing gene flow between the wild *P. orientalis* and the domesticated *P. dulcis* (Figs 2 and 3, Table 4). Accordingly, several *P. dulcis* specimens were either admixed (with Bayesian assignment probabilities ranging between 0.1 and 0.9) or even genetically clustered within the *P. orientalis* genetic group (Fig. 3B,C). Furthermore, coalescent-based estimations of gene flow (Table 4) revealed that over the complete sampling area, *P. orientalis* contributed about 31 immigrants per generation to the *P. dulcis* pool ($\Theta_{\text{dulcis}} = 7.33$ and $M_{\text{orientalis-to-dulcis}} = 16.96$, $N_e m = \frac{1}{4}\Theta M = 31$). These results were corroborated by insights from other crops that revealed substantial genetic contributions from wild local species to the domesticated pool. For instance, patterns with a similar trend were observed for the maize (Matsuoka et al. 2002), the olive tree (Breton et al. 2008), several European grapevine cultivars (Myles et al. 2011), apples (Coart et al. 2006), or the bread wheat (Caldwell et al. 2004). In addition, wild species have long been used as a source of genetic novelty in breeding programs and almond cultivars are not an exception (Denisov 1988; Socias i Company 1998). Still, the present study outlined gene flow that was most likely an unintended genetic contribution from a wild species to domesticated almonds (i.e., the sampling included only traditional orchards, with no modern cultivars). From an ethnobotanic perspective, these results point out the question whether wild-to-crop exchanges remained exclusively spontaneous or whether traditional practices could have facilitated the fixation of introgressed wild genes in domesticated almond trees. The latter case was reported from many traditional agro-systems and involved for instance fig trees (Achtak et al. 2010) or olive trees (Aumeeruddy-Thomas et al. 2009). For almonds, several authors reported the direct use of wild species as rootstocks (Denisov 1988; Martinoli and Jacomet 2004), and in Italy, Godini (2000) and Socias i Company (1998) reported that self-compatibility (the allele Sf at the S-locus) could have been transferred spontaneously into cultivars from *P. webbii*, another wild relative. Additional ethno-botanical investigations are needed because traditional practices in almonds breeding remain largely unknown.

Genetic transfers from the domesticated *Prunus dulcis* to its wild relative

Nuclear SSRs showed that genetic exchanges between *P. dulcis* and *P. orientalis* were bidirectional and outlined substantial crop-to-wild gene flow (Figs 2 and 3 and Table 4). Indeed, the coalescent-based approach revealed that, over the complete sampled area, 28 domesticated migrants were introgressed into *P. orientalis*, at each generation (Table 4, $\Theta_{\text{orientalis}} = 7.11$ and $M_{\text{dulcis-to-orientalis}} = 15.64$, $N_e m = \frac{1}{4}\Theta M = 28$). These results were consistent with mentions of hybridization within the *Amygdalus* species group (Browicz and Zohary 1996) and confirmed that *P. dulcis* genes could spontaneously be introgressed into their wild relatives. This scenario is highly likely if genetically modified almond cultivars (e.g., *Agrobacterium*-mediated transformation, see Gradziel 2009) were introduced in the Western Mediterranean. Such genetic transfers can have significant evolutionary consequences, especially if the inserted transgene is adaptative under natural conditions (see Felber et al. 2007 for a review). For instance, traits such as enhanced fertility (e.g. Kron and Husband 2009), resistance to pests (Fladung et al. 2006) or viruses (e.g., plum pox virus, Scorza et al. 2007) could favor the emergence of highly competitive phenotypes if transferred into wild species. In addition, genetic exchange among related *Prunus* species (Grasselly 1977) could cause an uncontrolled spread of transgenes across the *Amygdalus* species complex (i.e., the bridge species concept, Felber et al. 2007).

As might be expected for a crop and its wild relative (Zohary 1984), the present study revealed substantial and ongoing gene flow among *P. dulcis* and *P. orientalis*. The magnitude of the detected gene flow consistently reflected the self-incompatible mating system of almond trees species and supported earlier mentions of admixture among cultivated and wild germplasms (Ortega and Dicenta 2003; Gradziel 2009).

Furthermore, we provided genetic evidence that wild lineages could have spontaneously contributed to the current cultivated gene pool. These results could confirm that the domestication of almonds might have been diffuse and characterized by recurrent genetic exchanges among the domesticated forms and the local wild relatives (Zeder 2006). Our results also highlighted the importance of including wild relatives when documenting the origins of almond domestication using genetic data, because gene flow from wild relatives can bias distance-based ancestry inferences (e.g., see van Heerwaarden et al. 2011).

Finally, our study revealed that crop-to-wild gene flow occurred commonly among the domesticated almond and at least one of its wild relatives. These results suggested

that transgenes could potentially introgress into *P. orientalis* populations and further outlined the need for detailed characterization of crop-to-wild gene flow within the *Amygdalus* species complex. Therefore, ad hoc containment strategies of transgenes might be necessary if genetically modified almond cultivars are grown in sympatry with their wild relatives.

Acknowledgements

The present study was funded by the 'Fruit Med Project' and distributed by the French Agropolis Foundation. N. Arrigo and N. Alvarez were funded by the Swiss National Science Foundation (grant No. 132747 and an Ambizione fellowship PZ00P3_126624, respectively). Genotyping was done in the 'Service Commun de Marqueurs Génétiques en Ecologie' of the CEF. The authors are very grateful to Rebecca Rundell for constructive comments on the manuscript and to fieldwork collaborators for their help during sample collections: A. Chehade, L. Chalak, B. Hamadeh, A. El Bitar (Lebanese Agricultural Research Institute); S. Beyazit, S. Cerce; K. Gündünz (Mustapha Kemal University, Turkey); and A. Al Ibrahim (General Commission for Scientific Agricultural Research, Syria), A. El Ibrahim (Centre National de Recherche Agronomique, Syria).

Data archiving statement

Data for this study will be available at Dryad: doi:10.5061/dryad.5f41fq18.

Literature cited

- Achtak, H., M. Ater, A. Oukabli, S. Santoni, F. Kjellberg, and B. Khadari. 2010. Traditional agroecosystems as conservatories and incubators of cultivated plant varietal diversity: the case of fig (*Ficus carica* L.) in Morocco. *BMC Plant Biology* **10**:28.
- Alvarez, N., N. Arrigo, and Intrabiodiv Consortium. 2008. SIMIL: an R (CRAN) scripts collection for computing genetic structure similarities based on STRUCTURE 2 outputs. *Molecular Ecology Resources* **8**:757–762.
- Arrigo, N., R. Guadagnuolo, S. Lappe, S. Pasche, C. Parisod, and F. Felber. 2011. Gene flow between wheat and wild relatives: empirical evidence from *Aegilops geniculata*, *Ae. neglecta* and *Ae. triuncialis*. *Evolutionary Applications* **4**:685–695.
- Aumeeruddy-Thomas, Y., Y. H'mimsa, M. Ater, and B. Khadari. 2009. Beyond the divide between wild and domesticated: spatiality, domesticity and practices pertaining to fig (*Ficus carica* L.) and olive (*Olea europaea* L.) agroecosystems in Morocco. Brussels: Crops and people: choices and diversity through time. Earth EU.
- Austerlitz, F., S. Mariette, N. Machon, P. H. Gouyon, and B. Godelle. 2000. Effects of colonization processes on genetic diversity: differences between annual plants and tree species. *Genetics* **154**:1309–1321.
- Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**:37–48.
- Berli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of The National Academy of Sciences of the United States Of America* **98**:4563–4568.
- Breton, C., C. Pinatel, F. Medail, F. Bonhomme, and A. Berville. 2008. Comparison between classical and Bayesian methods to investigate the history of olive cultivars using SSR-polymorphisms. *Plant Science* **175**:524–532.
- Browicz, K., and D. Zohary. 1996. The genus *Amygdalus* L. (Rosaceae): species relationships, distribution and evolution under domestication. *Genetic Resources and Crop Evolution* **43**:229–247.
- Caldwell, K. S., J. Dvorak, E. S. Lagudah, E. Akhunov, M. C. Luo, P. Wolters, and W. Powell. 2004. Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. *Genetics* **167**:941–947.
- Cipriani, G., G. Lot, W. G. Huang, M. T. Marrazzo, E. Peterlunger, and R. Testolin. 1999. AC/GT and AG/CT microsatellite repeats in peach [*Prunus persica* (L.) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. *Theoretical and Applied Genetics* **99**:65–72.
- Coart, E., S. Van Glabeke, M. De Loose, A. S. Larsen, and I. Roldan-Ruiz. 2006. Chloroplast diversity in the genus *Malus*: new insights into the relationship between the European wild apple [*Malus sylvestris* (L.) Mill.] and the domesticated apple (*Malus domestica* Borkh.). *Molecular Ecology* **15**:2171–2182.
- Corredor, E., M. Roman, E. Garcia, E. Perera, P. Arus, and T. Naranjo. 2004. Physical mapping of rDNA genes establishes the karyotype of almond. *Annals of Applied Biology* **144**:219–222.
- Denisov, V. P. 1988. Almond genetic resources in the USSR and their use in production and breeding. *Acta Horticulturae* **224**:299–306.
- Delplancke, M. 2011. Histoire évolutive de l'amandier cultivé (*Prunus dulcis*) en Méditerranée: regards croisés sur la domestication, dialogue entre la biologie et l'ethnobiologie, Biologie des populations, Université Montpellier 2, Montpellier.
- Dicenta, F., and J. E. Garcia. 1993. Inheritance of self-compatibility in almond. *Heredity* **70**:313–317.
- Dirlwanger, E., P. Cosson, M. Tavaud, M. J. Aranzana, C. Poizat, A. Zanetto, P. Arus et al. 2002. Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). *Theoretical and Applied Genetics* **105**:127–138.
- Doebley, J. F., B. S. Gaut, and B. D. Smith. 2006. The molecular genetics of crop domestication. *Cell* **127**:1309–1321.
- Ellstrand, N. C., H. C. Prentice, and J. F. Hancock. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* **30**:539–563.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**:2611–2620.
- Excoffier, L., and E. L. L. Heidi. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**:564–567.
- FAO. 2008. FAO Statistical Databases: Agriculture. FAO, Rome.
- Felber, F., G. Kozłowski, N. Arrigo, and R. Guadagnuolo. 2007. Genetic and ecological consequences of transgene flow to the wild flora. *Advances in Biochemical Engineering/Biotechnology* **107**:173–205.

- Fladung, M., D. Ewald, A. Balestrazzi, G. Allegro, and M. Confalonieri. 2006. Genetically modified trees expressing genes for insect pest resistance. In M. Fladung, and D. Edwalds, eds. *Tree Transgenesis*, pp. 253–273. Springer, Berlin, Heidelberg.
- Glemin, S., and T. Bataillon. 2009. A comparative view of the evolution of grasses under domestication. *New Phytologist* **183**:273–290.
- Godini, A. 2000. About the possible relationship between *Amygdalus webbii* Spach and *Amygdalus communis* L. *NUCIS Newsletter* **9**:17–19.
- Goodwillie, C., S. Kalisz, and C. G. Eckert. 2005. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology Evolution and Systematics* **36**:47–79.
- Gradziel, T. M. 2009. Almond (*Prunus dulcis*) breeding. In S. Mohan Jain, and P. M. Priyadarshan, eds. *Breeding Plantation Tree Crops: Temperate Species*, pp. 1–31. Springer, New York.
- Grasselly, C. 1977. Réflexions sur les caractéristiques des espèces sauvages d'amandier et sur leur utilisation éventuelle dans des programmes d'amélioration génétique. In 3e Colloque du groupe de recherche et d'étude méditerranéen pour le pistachier et l'amandier. GREMPA, CIHEAM, Bari, Italy.
- Grasselly, C., and P. Crossa-Raynaud. 1980. L'amandier. Techniques agricoles et productions méditerranéennes. G.P. Maisonneuve et Larose, Paris, XII.
- Haudry, A., A. Cenci, C. Ravel, T. Bataillon, D. Brunel, C. Poncet, I. Hochu *et al.* 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. *Molecular Biology and Evolution* **24**:1506–1517.
- van Heerwaarden, J., J. Doebley, W. H. Briggs, J. C. Glaubitz, M. M. Goodman, J. J. Sanchez Gonzalez, and J. Ross-Ibarra. 2011. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proceedings of the National Academy of Sciences of the United States of America* **108**:1088–1092.
- Hollingsworth, P. M., I. K. Dawson, W. P. Goodall-Copestake, J. E. Richardson, J. C. Weber, C. S. Montes, and R. T. Pennington. 2005. Do farmers reduce genetic diversity when they domesticate tropical trees? A case study from Amazonia. *Molecular Ecology* **14**:497–501.
- Kron, P., and B. C. Husband. 2009. Hybridization and the reproductive pathways mediating gene flow between native *Malus coronaria* and domestic apple, *M. domestica*. *Botany-Botanique* **87**:864–874.
- Ladizinsky, G. 1999. On the origin of almond. *Genetic Resources and Crop Evolution* **46**:143–147.
- Lynch, M. 2010. Scaling expectations for the time to establishment of complex adaptations. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **107**:16577–16582.
- Mariette, S., M. Tavaud, U. Arunyawat, G. Capdeville, M. Millan, and F. Salin. 2010. Population structure and genetic bottleneck in sweet cherry estimated with SSRs and the gametophytic self-incompatibility locus. *BMC Genetics* **11**:77.
- Martinoli, D., and S. Jacomet. 2004. Identifying endocarp remains and exploring their use at Epipalaeolithic Okuzini in southwest Anatolia, Turkey. *Vegetation History And Archaeobotany* **13**:45–54.
- Matsuoka, Y., Y. Vigouroux, M. M. Goodman, G. J. Sanchez, E. Buckler, and J. Doebley. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **99**:6080–6084.
- Myles, S., A. R. Boyko, C. L. Owens, P. J. Brown, F. Grassi, M. K. Aradhya, B. Prins *et al.* 2011. Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America* **108**:3530–3535.
- Ohta, S., T. Katsuki, T. Tanaka, T. Hayashi, Y. Sato, and T. Yamamoto. 2005. Genetic variation in flowering cherries (*Prunus* subgenus *cerasus*) characterized by SSR markers. *Breeding Science* **55**:415–424.
- Ortega, E., and F. Dicenta. 2003. Inheritance of self-compatibility in almond: breeding strategies to assure self-compatibility in the progeny. *Theoretical and Applied Genetics* **106**:904–911.
- Papa, R. 2005. Gene flow and introgression between domesticated crops and their wild relatives. In *Proceedings of the International Workshop on the Role of Biotechnology for the Characterisation and Conservation of Crop, Forestry, Animal and Fishery Genetic Resources*. Turin, Italy.
- Papa, R., E. Bellucci, M. Rossi, S. Leonardi, D. Rau, P. Gepts, L. Nanni *et al.* 2007. Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Annals of Botany* **100**:1039–1051.
- Petit, R. J., and L. Excoffier. 2009. Gene flow and species delimitation. *Trends in Ecology & Evolution* **24**:386–393.
- Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* **37**:187–214.
- Petit, R. J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* **12**:844–855.
- Petit, R. J., J. Duminil, S. Fineschi, A. Hampe, D. Salvini, and G. G. Vendramin. 2005. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology* **14**:689–701.
- Potter, D., F. Gao, P. E. Bortiri, S. H. Oh, and S. Baggett. 2002. Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. *Plant Systematics and Evolution* **231**:77–89.
- Powell, W., M. Morgante, C. Andre, J. W. McNicol, G. C. Machray, J. J. Doyle, S. V. Tingey *et al.* 1995. Hypervariable microsatellites provide a general source of polymorphic DNA markers for the chloroplast genome. *Current Biology* **5**:1023–1029.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
- R Foundation for Statistical Computing. 2011. R: A Language and Environment for Statistical Computing, Reference Index Version 2.13.1. R Foundation for Statistical Computing, Vienna, Austria.
- Rieseberg, L. H. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28**:359–389.
- Rieseberg, L. H., J. Whitton, and C. R. Linder. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. *Acta Botanica Neerlandica* **45**:243–262.
- Scorza, R., J. Hily, A. Callahan, T. Malinowski, M. Cambra, N. Capote, I. Zagari *et al.* 2007. Deregulation of plum pox resistant transgenic plum 'HoneySweet'. In *Proceedings of the International Symposium on Biotechnology of Temperate Fruit Crops and Tropical Species*.
- Socias i Company, R. 1998. Fruit tree genetics at a turning point: the almond example. *Theoretical and Applied Genetics* **96**:588–601.
- Testolin, R., T. Marrazzo, G. Cipriani, R. Quarta, I. Verde, M. T. Dettori, M. Pancaldi *et al.* 2000. Microsatellite DNA in peach (*Prunus persica* L. Batsch) and its use in fingerprinting and testing the genetic origin of cultivars. *Genome* **43**:512–520.
- Trucco, F., T. Tatum, A. L. Rayburn, and P. J. Tranel. 2009. Out of the swamp: unidirectional hybridization with weedy species may explain the prevalence of *Amaranthus tuberculatus* as a weed. *New Phytologist* **184**:819–827.

- Willcox, G., S. Fornite, and L. Herveux. 2008. Early Holocene cultivation before domestication in northern Syria. *Vegetation History and Archaeobotany* **17**:313–325.
- Willcox, G., R. Buxo, and L. Herveux. 2009. Late Pleistocene and early Holocene climate and the beginnings of cultivation in northern Syria. *Holocene* **19**:151–158.
- Zeder, M. A. 2006. *Documenting Domestication: New Genetic and Archaeological Paradigms*. University of California Press, Berkeley.
- Zeinalabedini, M., M. Khayam-Nekoui, V. Grigorian, T. M. Gradziel, and P. Martinez-Gomez. 2010. The origin and dissemination of the cultivated almond as determined by nuclear and chloroplast SSR marker analysis. *Scientia Horticulturae* **125**:593–601.
- Zohary, D. 1984. Modes of evolution of plants under domestication. In W. F. Grant, ed. *Plant Biosystematics*, pp. 579–596. Academic Press, Toronto, ON.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Screening for the optimal *K* number of groups, using the deltaK criterion (Evanno et al. 2005).

Figure S2. Geographical map of specimen assignment probabilities, as revealed by Bayesian clustering of genotypes based on nuclear SSRs.

Table S1. Genetic diversity of *P. dulcis* and *P. orientalis* populations, as revealed by nuclear SSRs.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.