Population genetics of Mediterranean and Saharan olives: geographic patterns of differentiation and evidence for early generations of admixture

G. Besnard¹,*, A. El Bakkali², H. Haouane², D. Baali-Cherif³, A. Moukhli³ and B. Khadari²

¹CNRS-UPS-ENFA, EDB, UMR 5174, Bât. 4R1, 31062 Toulouse cedex 9, France, ²INRA/CBNMed, UMR 1334, AGAP, 34060 Montpellier, France, ³INRA, UR APCRPG, BP 578, Meknès, Morocco, ⁴BP44, Laboratoire de Recherche sur les Zones Arides, USTHB/INA, Alger, Algeria and ⁵INRA Marrakech, UR Amélioration des Plantes, Marrakech, Morocco

*For correspondence. E-mail guillaume.besnard@univ-tlse3.fr

Received: 5 June 2013  Returned for revision: 28 June 2013  Accepted: 5 July 2013  Published electronically: 6 September 2013

INTRODUCTION

The origins of the Mediterranean cultivated olive (Olea europaea subsp. europaea) are hotly debated, but it is usually accepted that its primary domestication started in the Levant as attested to by archaeological, historical and molecular evidence (Kanieowski et al., 2012; Zohary et al., 2012; Besnard et al., 2013a). Multiple local selections of cultivars have also been reported by genetic analyses (e.g. Besnard et al., 2001a; Belaj et al., 2002; Baldoni et al., 2006; Breton et al., 2008), and secondary diversification of the crop followed the oleiculture diffusion over the whole Mediterranean basin (Terral. 1997; Besnard et al., 2001a). The contribution of the western wild relatives distributed over three continents, from the Mediterranean basin to South Africa and southwestern Asia. Recent studies suggested that this crop originated in the Levant while a secondary diversification occurred in the Sahara. Although the data available were too limited to draw definite conclusions, here, patterns of genetic differentiation in the Mediterranean and Saharan olives are analysed to test for recent admixture between these taxa.

**Conclusions** This study reports for the first time admixture between Mediterranean and Saharan olives. Although its contribution remains limited, Laperrine’s olive has been involved in the diversification of cultivated olives.
nuclear and plastid genomes) on both Mediterranean and Saharan subspecies. In addition, such knowledge may have implications for the management of olive genetic resources.

According to Rubio de Casas et al. (2006), the genetic differentiation between subspecies *O. europaea* and *laperrierei* is highly significant ($F_{ST} = 17.2\%$; assessed with amplified fragment length polymorphisms (AFLPs)) as expected for allopatric taxa. Early generations of hybrids between two genetically differentiated taxa can be detected theoretically with genetic population analyses (e.g. admixture detection; Beaumont et al., 2001). Such a method has already been successfully used in olive to detect hybrids between *O. europaea* and *cuspidata* in the invasive Australian range (Besnard et al., 2008, 2013b), and for detecting admixture between Mediterranean gene pools (Baldoni et al., 2006; Breton et al., 2006; Belaj et al., 2007). This approach was recently applied to document the secondary domestication history of the apple (Cornille et al., 2012) and almond trees (Delplancke et al., 2013).

In the present study, we describe patterns of genetic differentiation in the Mediterranean and Saharan olives, and we tested for admixture between these taxa. Nuclear microsatellites and plastid haplotypes from different studies were compiled and the genetic characterization of wild populations was completed. The maternal haplotype of all individuals was first used to detect possible seed-propagation exchanges between taxa. A Bayesian clustering method was applied for the detection of main gene pools, and simulations were used for assessing the power of our approach to detect admixture. Based on our results, the human-mediated diffusion of the olive culture over the Mediterranean basin and the contribution of *O. europaea* subsp. *laperrierei* to the cultivated olive diversification are discussed.

**Materials and Methods**

**Plant material**

In the present study, we analysed plastid and nuclear genetic data sets generated on 1130 olive accessions (see below). This sample encompasses Mediterranean cultivars and wild olives from the Mediterranean basin and the Saharan mountains. Mediterranean accessions were *a priori* classified according to three pre-defined zones (Supplementary Data Tables S1 and S2): East (from Croatia to the Levant, including Egypt), Central (from Italy to France and Libya to Algeria) and West (Iberian Peninsula and Morocco).

**Olive cultivars** (*Olea europaea* subsp. *europaea* var. *europaea*). All cultivated accessions considered in this study were first characterized with nuclear microsatellite loci (or simple sequence repeats (SSRs)) by Besnard et al. (2007b), Khadari et al. (2008) and Haouane et al. (2011). When accessions were distinguished based on only one or two SSR alleles, the distinction was doubtful and could be attributed to mutations (Lopes et al., 2004; Baali-Cherif and Besnard, 2005; Khadari et al., 2008). To avoid considering variants of the same accession in the data analyses but also to avoid including families of very closely related individuals that can bias Bayesian clustering analyses (Rodriguez-Ramillo and Wang, 2012), we decided to keep (at random) only one tree for such closely related accessions (i.e. distinguished by one or two alleles per ten loci). A total of 470 accessions were analysed. Four hundred and thirty-five cultivated accessions were from the Olive World Germplasm Bank of Marrakech (OWGB-M). Additionally, 15 cultivated accessions were from the World Olive Germplasm Bank of Cordoba (WOGB-C) and 20 were sampled in the field (Supplementary Data Table S1). All these accessions have also been characterized with a plastid genomic profiling method (Besnard et al., 2013a).

**Oleasters populations**. Wild populations from the Mediterranean Basin (*O. europaea* subsp. *europaea* var. *sylvestris*) have been recently sampled for the study of ptDNA variation toward the Mediterranean basin (Besnard et al., 2013a). In the present study, a sub-set of 390 trees from 45 locations was considered for nuclear microsatellite characterization (Supplementary Data Table S2).

**Laperriere’s olive populations**. Populations of *O. europaea* subsp. *laperrierei* (Batt. & Trab.) Cif. have been previously sampled in the Ahaggar (Hogggar and Tassili n’Ajjer, South Algeria) and the Air (Tamgak, Bagezane and Egalah, North Niger). Most individuals were characterized with nuclear SSR loci (seven of them shared with the present study) in order to identify genotypes and analyse the population dynamics (Baali-Cherif and Besnard, 2005; Besnard et al., 2007a). Here, 270 distinct genotypes were considered: 48 from the Bagezane-Egalah, 44 from the Tamgak, 45 from the Tassili n’Ajjer and 133 from the Hogggar.

**Genetic characterization**

Ten nuclear SSR loci were used to characterize all individuals: PA(ATT)2 (Saumitou-Laprade et al., 2000), ssrOeUA-DCA01, ssrOeUA-DCA05, ssrOeUA-DCA09, ssrOeUA-DCA14, ssrOeUA-DCA15, ssrOeUA-DCA18 (Seef et al., 2000), GAPU71A (Carriero et al., 2002) and EM003 (de la Rosa et al., 2002). We completed the molecular analyses for all accessions (in particular for wild accessions) using the methods previously described (Baali-Cherif and Besnard, 2005; Haouane et al., 2011). The congruence between initial data sets was checked by comparing 35 genotypes (both wild and cultivated accessions; data not shown) that were independently characterized with the same set of markers in two different laboratories. We also showed that the unnamed cultivar from Morocco that harbours the ptDNA haplotype L1-1 (reported by Besnard et al., 2011) displays exactly the same nuclear SSR profile as ‘Dhokar’ (Tunisia; OWGB Marrakech) and can thus be considered as the same accession.

Laperriere’s olive samples were analysed with ptDNA markers as described by Besnard et al. (2011). These data and available plastid profiles for 1797 Mediterranean olive trees (Besnard et al., 2013a) – including all Mediterranean accessions here analysed with nuclear SSRs – were compiled for further analyses.

**Data analyses**

Genetic diversity parameters were first estimated for each group (i.e. cultivars, oleasters and Laperriere’s olive). In order to account for different sample sizes (minimum 270 individuals), allelic richness ($R_s$) was estimated for all nuclear SSR loci and ptDNA haplotypes using FSTAT v.2.9.4 (Goudet 1995). The total genetic diversity ($H_T$; Nei, 1987) was also estimated for each locus. A Wilcoxon signed-ranks test (one sided) was used to evaluate the significance of the difference of allelic richness and gene diversity at nuclear loci between cultivars, oleasters and Laperriere’s olives.
Haplotypes networks on the whole plastid data set or at a clade level (lineages E1 and L1) were reconstructed with the reduced-median method implemented in NETWORK V.4.1.12 (Bandelt et al., 1999). For the allele coding, we followed exactly the same procedure as previously described by Besnard et al., (2013a). A ptDNA haplotype was defined by the combination of alleles from the 45 polymorphic loci.

The inter- and intrasubspecies structure of the genetic diversity was then investigated. Pairwise $F_{ST}$ between the three taxa were computed based on nuclear SSRs using FSTAT. The significance of pairwise differentiation was assessed using standard Bonferroni corrections. The number of genetically homogeneous clusters ($K$) in the whole data set and for each pair of taxa (i.e. cultivars/oleasters, cultivars/Laperrine’s olive, oleasters/Laperrine’s olive) was then determined using a model-based clustering method implemented in STRUCTURE v.2.3.4 (Pritchard et al., 2000). Bayesian analysis was run under the admixture model for 1 000 000 generations after a burn-in period of 200 000, assuming a correlation among allele frequencies. Analyses were run for $K$ between one and ten clusters with ten iterations for each $K$ value. The most likely number of clusters was determined using the ad-hoc measure $\Delta K$ (Evanno et al., 2005) with the R program (R Development Core Team, 2010), whereas the similarity index between ten replicates for the same $K$ clusters ($H'$) was calculated with CLUMPP v1.1.2 (Greedy algorithm; Jakobsson and Rosenberg, 2007). For each $K$ value that was retained, each accession was assigned to their respective clusters with a posterior membership coefficient ($p$).

The power of model-based clustering analyses to detect admixed individuals issued from intersubspecies hybridization or successive backcrosses was assessed by simulations based on the ten nuclear microsatellite loci used in our study. Two objectives were tackled: first, we assessed the level of assignment that allows a confident detection of admixture between Mediterranean and Saharan gene pools. We tested on pairs of simulated panmictic populations (that can thus be considered as non-admixed) the level of assignment to two genetic clusters: subspecies $e u r o p e a e$ (E) and $l a p e r r i n e i$ (L). In these simulated data sets, we thus estimated the level of assignment to the wrong subspecies that occurs by chance. In practice, panmictic populations of 1000 genotypes were generated for the Laperrine’s olives (L) and oleasters (E) with the software HYBRIDLAB (Nielsen et al., 2006). Alleles were randomly sampled in each taxon assuming neutrality, linkage equilibrium and random mating. Allele frequencies in our simulated matrices were constrained to be similar to our observations. The procedure was replicated ten times, and simulated genotypes were used to carry out admixture analyses with STRUCTURE as described above (considering $K = 2$ clusters). The level of partial admixture between Mediterranean and Saharan olive was then investigated in the whole data set. The distribution of the posterior membership coefficient ($p$) values was plotted. The maximum value for a wrong assignment that may happen by chance was determined. For the second objective, we estimated the number of generations of backcrosses that occurred for accessions that were confidently detected as advanced generation of hybrids (see below). We randomly selected genotypes of 50 individuals of Laperrine’s olive (from Tassili n’Ajjer and Hoggar), 50 oleasters (from the west-central Mediterranean gene pool) and 50 olive varieties from the central Mediterranean gene pool. Aiming to exclude recent hybrids, these genotypes were selected among individuals assigned to the Laperrine’s olive or oleaster populations with $p > 0.99$ and to olive varieties with $p > 0.95$ (see below). We then generated 1000 genotypes of each F1 and successive backcrossed genotypes (from BC1 to BC5) with HYBRIDLAB as described above. The procedure was replicated ten times and the simulated genotypes were used to carry out admixture analyses with STRUCTURE as described above. The level of partial admixture between Mediterranean and Saharan olive was then investigated and directly compared with the observed value on admixed individuals from our data set.

**RESULTS**

**Comparison of genetic diversity parameters between Mediterranean and Saharan olives**

No significant difference is observed between cultivars and the Laperrine’s olive for the allelic richness and genetic diversity (Wilcoxon signed-ranks tests, $P = 0.441$ and 0.759, respectively; Table 1). In contrast, we observe a marginally significantly higher allelic richness in oleaster samples than in the Laperrine’s olive (Wilcoxon signed-ranks test, $P = 0.059$), and a significantly higher genetic diversity in oleasters than in the Laperrine’s olive (Wilcoxon signed-ranks test, $P = 0.046$), but these results could just reflect samplings on different geographic scales for these two taxa. Oleasters also show significantly higher allelic richness and genetic diversity than Mediterranean cultivars (Wilcoxon signed-ranks test, $P < 0.01$ for $R_S$ and $H_T$, Table 1). A similar trend is observed on ptDNA haplotypes (but not statistically testable; Table 1).

**Distinction of maternal lineages in Saharan and Mediterranean olives**

A total of 67 ptDNA haplotypes (Supplementary Data Table S3) are detected; 48 are found in the Mediterranean olive and 20 in the Laperrine’s olive. The phylogenetic analysis reveals four clusters of closely related ptDNA haplotypes, or ptDNA lineages, on the whole data set (Fig. 1A). Lineages E1, E2 and

<table>
<thead>
<tr>
<th>Locus</th>
<th>Cultivars</th>
<th>Oleasters</th>
<th>Laperrine’s olive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuclear SSR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ssrOEA-U-DCA1</td>
<td>39</td>
<td>17-6</td>
<td>0.62</td>
</tr>
<tr>
<td>ssrOEA-U-DCA5</td>
<td>27</td>
<td>13-1</td>
<td>0.52</td>
</tr>
<tr>
<td>ssrOEA-U-DCA8</td>
<td>29</td>
<td>19-2</td>
<td>0.83</td>
</tr>
<tr>
<td>GAPU71A</td>
<td>30</td>
<td>12-5</td>
<td>0.47</td>
</tr>
<tr>
<td>EM003</td>
<td>25</td>
<td>13-0</td>
<td>0.81</td>
</tr>
<tr>
<td>Mean</td>
<td>27-7</td>
<td>14-1</td>
<td>0.71</td>
</tr>
<tr>
<td>ptDNA</td>
<td>51</td>
<td>12-1</td>
<td>0.34</td>
</tr>
</tbody>
</table>
E3 have been previously reported in the Mediterranean basin (Besnard et al., 2013a). Lineage L1 is sister to E1, from which it was previously not separated (Besnard et al., 2007c). A minimum of seven cpSSR mutation steps are detected between E1 and L1 haplotypes (Supplementary Data Table S3), attesting for a clear divergence of these two lineages. In the wild pool, three ptDNA lineages (namely E1, E2 and E3) are observed exclusively in oleasters, while the fourth lineage (L1) is only revealed in the Saharan populations. All cultivars, except one, show a Mediterranean haplotype. A Saharan haplotype (L1.1) is detected in ‘Dhokar’ (Fig. 1; Supplementary Data Table S2).

Identification of genetic clusters in Saharan and Mediterranean olives

The genetic differentiation based on nuclear SSRs between cultivars, oleasters and the Laperrine’s olives is highly significant (Table 2) and particularly high between the Mediterranean and Saharan olives (e.g. $F_{ST} = 18.9\%$ between the Laperrine’s olive and oleasters). The Bayesian clustering with STRUCTURE confirms these results, with a clear distinction of the Saharan and Mediterranean taxa (Figs 2 and 3).

The $K$ value selected following the criterion defined by Evanno et al. (2005) is 2 (Fig. 2), but here we also show results for three and four clusters because they reflect sub-structures in the Mediterranean olive data set that are supported in separate analyses of Mediterranean accessions (see Supplementary Data Figs S1 and S2). At $K = 2$, the two clusters correspond to subspecies laperrinei and europaea (named L and E, respectively), and all individuals are assigned to their respective subspecies with a mean membership coefficient ($p > 0.7$). Assignments to clusters E and L is, however, not always categorical ($p < 0.9$ for ‘Taroudant no 6’ and a few accessions from the Tassili n’Ajjer). At $K = 3$, two clusters of oleasters (namely E-I and

---

**Fig. 1.** Plastid DNA haplotype networks reconstructed with the reduced-median method implemented in NETWORK (Bandelt et al., 1999). Haplotypes are represented by yellow or green circles, while the missing, intermediate nodes are indicated by small red dots. The length of branches is proportional to the number of mutation steps. (A) Network based on the whole data set. The Mediterranean lineages E1, E2 and E3 and the Saharan lineage L1 are indicated. (B) Haplotype network for the sister lineages E1 and L1. On this network, the name of each haplotype is indicated, and the circle size is relative to the number of observed occurrences for each haplotype. Haplotypes E1.1, E1.2 and E1.3 are the most frequent haplotypes in cultivars (Besnard et al., 2013a). Haplotype L1.1 (in green) is shared between Saharan and Mediterranean olives, but lineage L1 is primarily distributed in the Saharan Mountains. L1.1 was observed once in the Mediterranean basin (‘Dhokar’: Morocco–Tunisia).
E-II) are distinguished (see also Supplementary Data Figs S1 and S2). Oleasters from the western and central Mediterranean are mainly assigned to cluster E-I, while cultivars and the remaining oleasters (from Croatia to the Levant) are mainly assigned to cluster E-II. Note that 463 (98.5%) and 390 (83%) cultivated olive accessions are assigned to cluster E-II with \( p > 0.5 \) and 0.8, respectively. At \( K = 4 \), most of the western and central Mediterranean cultivars are distinguished from eastern Mediterranean accessions (both oleasters and cultivars; see also Supplementary Data Figs S1 and S2). A general trend for the assignment of genotypes from a given Mediterranean zone to a genetic cluster (e.g. EII-b for eastern accessions or E-I for western oleasters; see also Fig. 3) is thus observed. A separate analysis of all Mediterranean accessions also supports a soft distinction of cultivated and wild gene pools according to their geographical origin (Supplementary Data Fig. S1).

### Nuclear evidence for a few admixture events between Mediterranean and Saharan olives

Based on the analysis of the whole data set, no obvious admixture between cultivars and the Laperrine’s olive is revealed (see above), but the complex genetic structure in wild and cultivated Mediterranean olives could mask subtle relationships between these taxa and the Laperrine’s olive (as recently shown in apple; Cornille et al., 2012). For this reason, we chose to analyse the genetic structure between oleasters and the Laperrine’s olive (Fig. 3A) and between cultivars and the Laperrine’s olive (Fig. 3B). In these two separate analyses, most accessions of subspecies *laperrinei* (263/270) and *europaea* (852/860) are assigned to clusters L and E, respectively, with a mean membership coefficient \( (p) > 95\% \) (Fig. 3; see Supplementary Data Figs S1 and S2).

#### Table 2. Pairwise differentiation (\( F_{ST} \)) between the Mediterranean and Saharan olive taxa based on nuclear data

<table>
<thead>
<tr>
<th></th>
<th>Oleasters</th>
<th>Cultivars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laperrine’s olive</td>
<td>7.36***</td>
<td>26.61***</td>
</tr>
<tr>
<td>Cultivars</td>
<td>18.91***</td>
<td></td>
</tr>
</tbody>
</table>

* *** \( P < 0.001 \)*.

![Fig. 2](image_url)  
**Fig. 2.** Inference of population structure based on ten nuclear SSRs using model-based Bayesian clustering implemented in STRUCTURE (Pritchard et al., 2000) on the whole data set (1,130 individuals). Each vertical bar represents an individual. The membership coefficient of assignment \( (p) \) of each individual to different gene pools is shown for \( K = 2 \), \( K = 3 \) and \( K = 4 \) clusters. \( H' \) represents the similarity coefficient between ten runs for each \( K \), and \( \Delta K \) is the ad-hoc measure of Evanno et al. (2005). The graph at the bottom right gives \( \Delta K \) plotted against \( K \). The three taxa (i.e. Laperrine’s olive, oleasters and cultivars) are indicated, and three pre-defined regions are recognized for both oleasters and cultivars (Supplementary Data Tables S1 and S2). ‘West Mediterranean’ corresponds to accessions from Morocco and the Iberian Peninsula, ‘Central Mediterranean’ corresponds to accessions from Algeria to Libya and France to Continental Italy, and ‘East Mediterranean’ corresponds to accessions from Croatia to the Levant. The most likely genetic structure model is \( K = 2 \) clusters, according to \( \Delta K \) and \( H' \) \( (\Delta K = 19824.9 \) and \( H' = 0.999 \)). At \( K = 3 \), each cluster corresponds to subspecies *laperrinei* and *europaea* (L and E, respectively), whereas at \( K = 4 \), numerous western and central oleasters (that mostly belong to cluster E-I) were shown to be distinguished from cultivars and eastern oleasters (that mostly belong to cluster E-II). At \( K = 4 \), most eastern oleasters and cultivars were distinguished from western/central cultivars (clusters E-IIa vs. E-IIb, respectively).
This clear assignment was expected since the genetic differentiation between both subspecies is high (Table 2). Nevertheless, a few accessions were assigned to their respective subspecies with a mean membership coefficient ($p$), 0.90. Considering the high genetic differentiation between Mediterranean and Saharan olive populations, such patterns may reflect admixture between gene pools E and L, but this hypothesis needs to be statistically tested (see below).

Because partial assignment to both subspecies could be not necessarily due to admixture and to some extent may arise by chance (Vähä and Primmer, 2006; Winkler et al., 2011), we analysed pairs of simulated panmictic populations of *laperrinei* and *europaea* based on observed allelic frequencies in each taxon.
The assignment of genotypes to their respective clusters (i.e., L for the Laperrine’s olive and E for oleasters) is correct using both observed and simulated data sets (Fig. 4). The mean membership probability ($p$) on the observed data is slightly lower than on simulated data. In particular, the assignment of simulated genotypes to their cluster is rarely below 0.9 (i.e., 0.16% and 0.03% of simulated genotypes in the Laperrine’s olives and oleasters, respectively) while this occurs in 1% of observed genotypes (1.85% and 1.03% of individuals in the Laperrine’s olives and oleasters, respectively; Fig. 4). This means that when one individual is partially assigned to both subspecies with $p$ values $<0.9$, this is probably not random and may reflect admixture between laperrinei and europaea.

The assignment of genotypes to their respective cluster (i.e., L for the Laperrine’s olive and E for oleasters) is correct using both observed and simulated data sets (Fig. 4). The mean membership probability ($p$) on the observed data is slightly lower than on simulated data. In particular, the assignment of simulated genotypes to their cluster is rarely below 0.9 (i.e., 0.16% and 0.03% of simulated genotypes in the Laperrine’s olives and oleasters, respectively) while this occurs in $>1\%$ of observed genotypes (1.85 and 1.03% of individuals in the Laperrine’s olives and oleasters, respectively; Fig. 4). This means that when one individual is partially assigned to both subspecies with $p$ values $>0.1$, this is probably not random and may reflect admixture between laperrinei and europaea.

Based on an assignment threshold of 0.1, a few Laperrine’s olive individuals from the Tassili n’Ajjer (South Algeria) show evidence of admixture with the Mediterranean cluster E (Fig. 3). Five individuals from this population are assigned to E with $p > 0.1$, reaching a maximum of 0.317 in individual Tassili n° 5 (Fig. 3A). We also find five Mediterranean accessions (three cultivars and two oleasters) that are similarly assigned to cluster L with $p > 0.1$ (Fig. 3A, B): cultivars ‘Belluti’ (Turkey; $p = 0.255$), ‘Ifri’ (Algeria; $p = 0.105$), ‘Dhokar’ (Tunisia–Morocco; $p = 0.136$), Oleasters ‘Pylos no 12’ (Greece; $p = 0.101$) and ‘Taroudant no 6’ (Morocco; $p = 0.237$). All putatively admixed individuals from the Sahara and Mediterranean basin are more assigned to subsp. laperrinei and subsp. europaea, respectively, suggesting they may result from early generations of backcrosses (e.g., BC1 or BC2) with the local taxon. Our simulations of backcrosses indicate that the observed level of admixture may indeed correspond to BC1 or BC2, but the confidence interval of assignment to each gene pool is large for BC generations (see details in Supplementary Data Figs S4 and S5). This is probably due to an insufficient number of loci to assess precisely.

---

**Fig. 4.** Comparison of membership assignments ($p$) to the Laperrine’s olive and oleaster genetic clusters on natural populations (observations) and simulated panmictic data sets (simulations; ten iterations of 1000 genotypes). A model-based Bayesian clustering implemented in STRUCTURE (Pritchard et al., 2000) was used considering $K = 2$ clusters. The frequency of genotypes assigned to their respective subspecies is reported for each class of values of $p$ in the Laperrine’s olives (A) and oleasters (B). Each simulated data set is individually treated. Frequencies of assignment of genotypes to their genetic cluster for $p > 0.9$ are given on the right for simulated and observed data sets.
the contribution of each gene pool in advanced generations of backcrosses (e.g. Šiška and Primmer, 2006). This also means that our approach may not allow detection of all BC2 generations and is inappropriate for detecting more advanced generations of hybrids.

**DISCUSSION**

In this study, we directly compared for the first time the genetic diversity in the cultivated olive, and wild populations of oleasters and Laperrine’s olives. In the following section, patterns of genetic differentiation in the Mediterranean basin are first interpreted in the light of recent studies that confirmed a primary origin of cultivated olives in the Levant (Kaniewski et al., 2012; Zohary et al., 2012; Besnard et al., 2013a). The admixture between Mediterranean and Saharan gene pools is then discussed.

*On patterns of genetic differentiation among Mediterranean olives*

Both oleaster and cultivar samplings cover the whole Mediterranean area and are representative of the different gene pools present in this region (Haouane et al., 2011; Besnard et al., 2013a). We observe a reduction of 33 % of the allelic richness and 16 % of the total gene diversity for nuclear SSRs in cultivars compared with oleasters, while a reduction of 62 % of the allelic richness and total gene diversity is revealed for plastid haplotypes in the cultivated pool. Such a lower genetic diversity in cultivars compared with oleasters has also been revealed by Lumaret et al. (2004), Belaj et al. (2010) or Besnard et al. (2011) with different marker sets. This substantial reduction of diversity on both nuclear and plastid genomes may be interpreted as a signature of genetic erosion linked to domestication and diffusion, and, here, we can definitely conclude that all the oleaster diversity was not captured in the cultivated pool despite recurrent crop–wild gene flow (Miller and Gross, 2011).

A clear genetic differentiation between eastern and western Mediterranean oleasters is also observed in our study. This result has been reported by several authors (Angiolillo et al., 1999; Besnard et al., 2001a, 2013a; Breton et al., 2006; Rubio de Casas et al., 2006). The assignment of most cultivars (83 %) to gene pool E-II (Fig. 2; Supplementary Data Fig. S1) with p > 0·8 confirms the strong eastern affiliation of the cultivated gene pool as indicated by plastid markers (Besnard et al., 2013a). Actually, no cultivar was strongly assigned to the western gene pool E-I (for p > 0·8), emphasizing its secondary contribution to the olive domestication. It is usually accepted that olive domestication has primarily started in the eastern Mediterranean basin (Kaniewski et al., 2012; Zohary et al., 2012). The human-mediated dispersal of oleiculture from the East to the West with the main civilizations was then associated with secondary diversification of cultivars in the central and western Mediterranean, as reported by several authors (Besnard et al., 2001a; Belaj et al., 2002, 2012; Owen et al., 2005; Baldoni et al., 2006; Sarri et al., 2006; Breton et al., 2008; Haouane et al., 2011; Díez et al., 2012). This long story may explain the existence of a sub-structure among cultivated olives with three main gene pools (Supplementary Data Fig. S2), as recently shown by Haouane et al. (2011), Belaj et al. (2012) and Díez et al. (2012). Additional investigations are still necessary to depict processes of cultivated olive diversification, and particularly to infer the history of olive diffusion and to determine precisely the contribution of the western gene pool.

*Evidence for admixture between Mediterranean and Saharan olives*

Based on ten nuclear SSR loci, we then used model-based clustering analyses to test for recent admixture between Saharan and Mediterranean olives (Figs 3 and 4). According to simulations on oleasters and the Laperrine’s olives, the probability to obtain assignment of one genotype to its respective cluster inferior to 0·9 does not exceed 2 × 10^{-3}. A partial assignment with p-values > 0·1 to each cluster is thus very unlikely to be reached by chance but instead indicates admixture between these taxa (Fig. 4). Based on this very conservative threshold, five Laperrine’s olive individuals and five Mediterranean accessions (two oleasters and three cultivars, namely ‘Dhokar’, ‘Ifri’ and ‘Belluti’) show assignment to both clusters L and E (Fig. 3A, B). This analysis thus confirms that the Maghreb cultivar ‘Dhokar’, that has a Saharan maternal origin (see above), results from admixture between Laperrine and Mediterranean olives, but other wild and cultivated admixed accessions are detected, and particularly in the Tassili n’Ajjer population. Such reciprocal gene flow was also strongly suspected on an AFLP profiling ( Rubio de Casas et al., 2006). These ten admixed individuals may result from early generations of backcrosses (e.g. BC1 to BC2) within the local taxon as supported by additional simulations (Supplementary Data Figs S4 and S5).

The Laperrine’s olive has thus already been involved in the cultivated olive diversification, confirming its potential use for the olive breeding (Lavee and Zohary, 2011; Besnard et al., 2012). Interestingly, due to their singular nuclear SSR profiles, cultivars ‘Dhokar’ and ‘Ifri’ were included in the core collections defined by Haouane et al. (2011), while ‘Belluti’ was not analysed in this study. Moreover, the level of admixture between subsp. *europaea* and subsp. *laperrinei* is relatively low, since no hybrid of the first generation is detected and only ten individuals out of 1130 (<1 %) are recognized as early generations of hybrids (for p > 0·1). This suggests that hybrids of the first generation are relatively rare and/or that their fitness is poor. In addition, successive backcrosses also greatly reduce the contribution of a wild progenitor, hampering its identification based on a few molecular markers as in the present study. Our approach
based on ten nuclear loci is inappropriate for detecting advanced generations of hybrids (Supplementary Data Figs S4 and S5) as previously demonstrated by Vähä and Primmer (2006). In order to assess precisely the importance of the laperrinei introgression in the cultivated olive gene pool, our results could be refined with a genome scan analysis (Scascitelli et al., 2010; Zhao et al., 2010).

We can also question the vectors of dispersal between the Mediterranean basin and the Saharan mountains. Both natural and human-mediated vectors could be involved. The aridity on the Sahara has greatly increased during the last five millennia (Maley, 2011), contributing to a high isolation of the Saharan olive populations during recent times. Ancient gene flow between the two investigated taxa could be still detectable due to the slow turnover of populations (Baali-Cherif and Besnard, 2005). Yet, the only evidence of maternal introgression was detected in a cultivar that has been collected from Tunisia to Morocco. This observation indicates that humans managed some hybrids and spread them over long distances.

Lastly, the Laperrine’s olive is a highly restricted and endangered plant (Besnard et al., 2012). Evidence of gene flow between Saharan and Mediterranean olive taxa needs to be taken into account by genetic resource managers and breeders. The Laperrine’s olive population is a putatively important wild genetic resource of the cultivated olive, and thus deserves to be protected in situ. This wild taxon should be also included in ex situ collections for further evaluation of agronomic traits.

SUPPLEMENTARY DATA
Supplementary data are available online at http://aob.oxfordjournals.org and consist of the following. Table S1: list of wild olive populations (subsp. europaea and laperrinei) characterized in the present study. Table S2: list of olive cultivars characterized in the present study. Table S3: profiles of the 67 plastid DNA haplotypes revealed in the present study using 45 polymorphic loci. Figure S1: Bayesian inference of population structure in the Mediterranean olive, for K = 2, 3 and 5 clusters. Figure S2: Bayesian inference of population structure in the oleasters and cultivated olives. Figure S3: Bayesian inference of population structure for two pairwise comparisons: (a) Laperrine’s olive–oleasters and (b) Laperrine’s olive–cultivars. Figure S4: Simulation of early and advanced generations of backcrosses between western oleasters and the Laperrine’s olive. Figure S5: Simulation of early and advanced generations of backcrosses between Central Mediterranean cultivars and the Laperrine’s olive.

ACKNOWLEDGEMENTS
This work was funded by the fellowships PIEF-GA-2008-220813 and ANR-12-AGRI-0002 (ARIMNET 2011-PESTOLIVE). It has been conducted in Silwood Park, the UMR AGAP and the lab EDB, part of the LABEX entitled TULIP (ANR-10-LABX-41). We also thank two referees for their constructive comments.

LITERATURE CITED


