Theor Appl Genet (2002) 104:1353–1361
DOI 10.1007/s00122-001-0832-x

G. Besnard · B. Khadari · P. Baradat · A. Bervillé

Olea europaea (Oleaceae) phylogeography based on chloroplast DNA polymorphism

Received: 18 July 2001 / Accepted: 25 October 2001 / Published online: 10 April 2002
© Springer-Verlag 2002

Abstract Chloroplast DNA diversity in the olive (*Olea europaea* L.) complex was studied using PCR-RFLP and microsatellite markers. Fifteen chlorotypes were distinguished. We constructed a cpDNA phylogenetic tree in which five clades were recognised and located in distinct geographic areas: clade *A* in Central and Southern Africa, clade *C* in Asia, clade *M* in North-West Africa, clade *E1* in the Mediterranean Basin and Sahara, and clade *E2* in West Mediterranea. Cultivated olive clustered with Mediterranean and Saharan wild forms (clades *E1* and *E2*). Strong genetic differentiation for cpDNA markers was observed between eastern and western Mediterranean olives, suggesting that these areas have represented different glacial refugia. Humans most likely spread one eastern chlorotype, preponderant in cultivars, across the western Mediterranean Basin. Its presence in *O. e.* subsp. *laperrinei* from the Sahara suggests a possible Mediterranean olive origin in an African population, which may have overlapped in the Southern Mediterranean during the Quaternary.

Keywords cpDNA · Olive · Phylogeography · Refugia · Spatial differentiation

Introduction

Phylogeographic studies of tree species are of particular interest for ecological and historical purposes and for the management of their genetic resources. In Europe, during the Quaternary ice ages, many species could survive only in favourable regions or refugia, and patterns of genetic differentiation among extant populations are often due to survival in different refugial zones combined with genetic drift and founder effects during re-colonisation. Molecular studies of spatial genetic structure in association with archaeological data (mainly fossil pollen) suggest different geographical routes for the diffusion of plant species northwards from their refugial zones (Huntley and Birks 1983; Taberlet et al. 1998; Hewitt 1999).

Olive (*Olea europaea* L.) is a thermophilic species characteristic of the Mediterranean Basin. Olive tree cultivation has been an important source of oil since ancient times, and humans have greatly influenced its dispersal (Solari and Vernet 1992; Terral and Arnold-Simard 1996). Consequently, olive genetic structure results probably from several factors related not only to the occurrence of refugia zones, but also to the particular biogeographic conditions of the Mediterranean Basin and to human influence.

Olive belongs to the *O. europaea* complex for which six subspecies have been defined using morphological characters (Green and Wickens 1989; Vargas et al. 2001). Each of these subspecies has a specific distribution: *O. europaea* subsp. *europaea* corresponds to the Mediterranean olive, *O. e.* subsp. *maroccana* (Greut. & Burd.) Vargas et al. is endemic to Southern Morocco, *O. e.* subsp. *laperrinei* (Batt. & Trab.) Ciferri is present in the Saharan Mountains, *O. e.* subsp. *cerasiformis* (Webb & Berth.) Kunk. & Sund. is endemic to the Madeira Island, *O. e.* subsp. *guanchica* Vargas et al. is endemic to the Canary Islands, and *O. e.* subsp. *cuspidata* (Wall.) Ciferri is distributed from South Africa to China. This latter subspecies is composed of a range of morphological types corresponding to limited geographic areas with three main species names: *Olea africana* Mill. from South to East Africa, *Olea chrysophylla* Lam. from East Africa to Arabia, and *Olea cuspidata* Wall. from Iran to China. All these taxa are supposed to be inter-fertile with Mediterranean olive (Besnard et al. 2001a), and gene exchange between distant populations could have contributed to the evolution of the Mediterranean olive (Quézel 1978, 1995).

In the Mediterranean olive, two forms have been described: namely wild olive or oleaster [*O. e.* subsp. *euro-*

Communicated by H.C. Becker

G. Besnard · B. Khadari · P. Baradat · A. Bervillé (✉)
INRA / UR-Génétique et Amélioration des Plantes,
Bât 33, 2 Place P Viala, 34060 Montpellier cedex 1, France
e-mail: berville@ensam.inra.fr
Tel.: +33-04-99-61-22-33, Fax: +33-04-67-04-54-15

Table 1 Geographical origin of sampled trees. N is the number of studied individuals for each site

Taxa	Provenance	N	Taxa	Provenance	N
O. e. ssp. *europaea*					
Cultivars[a]	Near East-Egypt	22	Oleasters	Ostricone (Frc)	14
	Greece-Turkey	10		Bonifacio (Frc)	6
	Iran	5		Filitosa (Frc)	10
	Algeria-Tunisia	10		Ogliastro (Frc, F)	10
	Morocco	16		Mont Boron (Fr)	25
	Iberian Peninsula	11		Cap des mèdes (Fr)	22
	Italy-Yugoslavia	19		Repentence (Fr)	11
	France	50		La Gardiole (Fr)	8
Oleasters	Mont Carmel (Israel)	18		Montpeyroux (Fr, F)	5
	Al Ascharinah (Syria)	12		Pignan (Fr, F)	4
	Harem (Syria)	13		Saint Paul et Valmalle (Fr, F)	3
	Balcah (Turkey)	1		Rivesaltes (Fr)	3
	Urla (Turkey)	6	*O. e.* ssp. *laperrinei*	La Source, Hoggar (Algeria)	1
	Cyrenaique (Lybia)	15		Adriane Mount, Hoggar (Al)	21
	Zaghouan (Tunisia)	6	*O. e.* ssp. *maroccana*	Immouzzer (Morocco)	12
	Tizi Ouzou (Algeria)	13		Argana (Morocco)	2
	Chefchaouen (Morocco)	6		Mentaga (Morocco)	3
	Taounate (Morocco)	4	*O. e.* ssp. *guanchica*	La Palma (Canary Islands)	9
	Moulay Idriss (Mo, F)	12		La Gomera (Canary Islands)	1
	Khenifra (Morocco)	9		Tenerife (Canary Islands)	1
	Bin El Ouidane (Mo)	10	*O. e.* ssp. *cuspidata*		
	Asni (Morocco)	6	*O. africana*	Nairobi (Kenya)	5
	Argana (Morocco)	9		Elgon Mount (Kenya)	5
	Tamri (Morocco)	5		Timau (Kenya)	4
	Tamanar (Morocco)	13		Stellenbosch (South Africa)	5
	Torviczon (Spain)	15		Morgenster (South Africa)	3
	Sierra Crevillente (Sp, F)	1		Amalundu (Zimbabwe)	1
	Asturias (Spain)	7		La Providence (*La Réunion*)	4
	Cadiz (Spain)	1	*O. chrysophylla*	Almihwit (Yemen)	9
	Majorque (Spb)	1	*O. cuspidata*	Kerman (Iran)	9
	Castellon (Spain)	1		IRO P (India)	1
	Messina (Sicily, Italy)	17		IRO P (China)	1
	Ali (Sicily, Italy)	12	*O. woodiana*	Umzimkulu River (SA)	1

[a] Genotypes previously discriminated with RAPDs; native range of populations is indicated in brackets: Al = Algeria; Mo = Morocco; Sp = Spain; Spb = Balearic Islands, Spain; Fr = Continental France; Frc = Corsica, France; SA = South Africa. The letter F, following the name of the country, indicates populations supposedly feral. IRO, P = "Institute for Olive Research", CNR, Perugia, Italy

paea var. *sylvestris* (Mill.) Lehr.] and cultivated olive (*O. e.* subsp. *europaea* var. *europaea*). In many areas, these two forms co-exist. The cultivated olive was one of the first trees to be domesticated (Zohary and Spiegel-Roy 1975). Olive picking is known to have occurred in the Near East since at least the Upper Palaeolithic (Kislev et al. 1992), and in Spain the trimming of olive trees has occurred since at least the Neolithic (Terral and Arnold-Simard 1996). During ancient times, the development of olive cultivation was associated with numerous commercial exchanges around the Mediterranean Basin and with the appearance of structures (tools and industry) for olive exploitation. The dissemination of the present Mediterranean wild olive has thus likely been associated with the presence of exploited olives and, more recently, with cultivars that were propagated by cuttings or grafting. Furthermore, individual trees have a great longevity, some living more than one 1,000 years. To some extent it is presumed that wild olives could thus in fact be ancient cultivated forms; hence, some genotypes corresponding to very old cultivars may still occur. To more clearly understand the process of olive domestication, it is important to clarify the origins and the migration of the wild Mediterranean form. Because of its allogamous mode of reproduction, a high level of heterozygosity has been maintained in olives, and domestication has been primarily due to the vegetative multiplication of selected individuals with favourable allele combinations (Zohary and Spiegel-Roy 1975). The exact origin of cultivated olive is still debated, and two distinct origins of cultivars, i.e. the east and west parts of the Mediterranean Basin, are possible (Zohary and Spiegel-Roy 1975; Terral and Arnold-Simard 1996; Besnard and Bervillé 2000).

As a first step towards documenting the history of the *O. europaea* complex, we analysed the present structure of the genetic diversity of cultivars and oleaster populations using cpDNA polymorphisms, which are maternally inherited in olive. To test the hypothesis that olive trees have recently migrated, and whether they were introduced or indigenous to the Mediterranean Basin, we chose to compare chloroplast DNA variation in the Mediterranean olive with that in the related species belonging to the *O. europaea* complex.

Table 2 Chloroplast polymorphism revealed in *Olea* species by microsatellites and PCR-RFLPs. Polymorphism codes correspond to the fragment code and to the polymorphism type (RS = Restriction Site; indel = insertion – deletion). For indel, the sizes of the bands associated with each variant are given. Restriction site absence or presence is noted as – and +, respectively

Polymorphism code	Restriction enzyme	Chlorotype															
		CE1	CE2	COM1	COM2	CCK	CCE1	CCE2	CCE3	CA1	CA2	CA3	CA4	CA5	CC1	CC2	CWO
ccmp5-indel		116	117	115	116	117	117	118	119	116	117	117	116	116	116	114	116
ccmp7-indel		129	129	130	130	129	129	129	129	129	129	129	129	129	128	128	128
FV-RS-1	*Taq*I	–	–	–	–	–	–	–	–	+	+	+	+	+	–	–	–
FV-RS-2	*Taq*I	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–
FV-RS-3	*Taq*I	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–
K2Q-indel	*Hpa*II	600	600	600	600	600	600	600	600	600	600	600	600	600	300	300	600
QR-RS-1	*Taq*I	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–
QR-RS-2	*Hae*III	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+
QR-indel-1	*Taq*I	228	228	228	228	228	228	228	228	230	230	230	228	230	228	228	235
QR-indel-2	*Taq*I	218	218	214	214	218	218	218	218	218	218	218	218	218	218	218	–
QR-indel-3	*Hae*III	282	282	285	285	288	290	290	290	290	290	290	290	290	290	290	295
TC-RS	*Taq*I	+	+	+	+	+	+	+	+	+	–	+	+	+	+	+	+
TC-indel	*Taq*I	244	244	244	244	244	244	244	244	244	244	244	244	246	246	246	240

Materials and methods

Plant materials

We analysed 143 cultivated forms and 334 oleasters from 37 sites from around the Mediterranean Basin (Table 1). For each cultivated clone previously discriminated using RAPDs (Besnard et al. 2001b; Khadari and Bervillé 2001), we analysed one tree. The wild samples were collected from sites distant from olive groves in 11 different regions. More precision about the plant material is available upon request to the authors.

For four other subspecies, we analysed 11 to 47 individuals (Table 1). Furthermore, *Olea woodiana* Knobl. was chosen as an outgroup. This species belongs to the section *Ligustroides* Benth. & Hook., which is the section closest to the section *Olea* corresponding to the *O. europaea* complex (Besnard et al. 2002a). In our presentation of the different taxa we followed the classification of Green and Wickens (1989) and Vargas et al. (2001), but we considered three taxa for the subspecies *cuspidata* (*O. africana*, *O. chrysophylla*, *O. cuspidata*) because of their different geographic locations.

PCR-RFLP and microsatellite procedures

DNA preparation was performed according to Besnard et al. (2000). PCR amplification of chloroplast DNA fragments was performed according to Dumolin-Lapègue et al. (1997b). Five primer pairs [FV, K2, QR, TC and VL, codes refered in Dumolin-Lapègue et al. 1997b] were used to amplify chloroplast DNA fragments. Approximately 1 µg of amplified DNA was restricted (2 U) according to the providers recommendation, separately by *Alu*I, *Dde*I, *Eco*RI, *Hae*III, *Hpa*II, *Nde*II, *Rsa*I or *Taq*I. Restriction products were separated in a 2.2% agarose gel. To detect the short indels (1 to 10 bp) in cpDNA fragments, we separated *Taq*I and *Hae*III restriction fragments on a 6% acrylamide gel after migration at 60 W for 1.5 h. In addition, cpDNA microsatellite polymorphism was studied using eight primer pairs defined by Weising and Gardner (1999) ccmp1 to 7 and ccmp9. PCR fragments were separated on a 6% acrylamide gel after migration at 60 W for 1.5 h.

Screening of the cpDNA fragment / enzyme combination and the cpDNA microsatellite primers showing polymorphism were performed on a subset of 18 individuals (details upon request): *O. e.* subsp. *maroccana* Immouzzer, *O. chrysophylla* Yemen, *O. africana* La Réunion, Stellenbosch, Elgon, Timau and Zimbabwe, *O. cuspidata* Iran, India and China, *O. e.* susbp. *guanchica*, *O. e.* subsp. *laperrinei* La Source, two cultivars ("Chemlal" and "Toffahi"), and Oleasters from Ostricone, Ali and Harem plus *Olea woodiana*. These samples, representing all taxa of the *O. europaea* complex and originating from distant populations, were selected on the basis of RAPDs to represent, as much as possible, the whole range of genetic diversity in the olive. Four of the five cpDNA fragments revealed 11 polymorphic sites, which corresponded to six point mutations and five length variants (Table 2), and the use of the eight cpDNA microsatellite primers allowed us to identify two polymorphic loci, and the length of variant fragments are displayed in Table 2. Fragment / enzyme combinations showing polymorphism were further used to characterise the remainder of the individuals.

cpDNA phylogeny reconstruction

From cpDNA PCR-RFLP and microsatellite data, a matrix of multi-state characters was constructed. Phylogenetic relationships among chlorotypes were analysed using the Wagner parsimony method, which minimises the total number of character state changes in the tree. This analysis was performed with the PAUP Software (Phylogenetic Analysis Using Parsimony version 4.0; Swofford 1998). All characters were of unordered type and had equal weight. After a heuristic search, the equally most-parsimonious trees and one strict consensus tree were obtained. The consistency index (CI), which measures the overall amount of homoplasy, was then calculated. Reliability of the resulting trees was tested by bootstrap analysis (Felsenstein 1985) of 1,000 replicates with PAUP.

Genetic differentiation of the Mediterranean olive

The genetic differentiation among Mediterranean olive was assessed using F_{st} values, calculated over the four chlorotypes present in oleaster populations (considering one locus with four alleles), according to the Weir and Cockerham method adapted to haploid data (Weir 1990) and extending the algorithm to a three-way nested statistical model with unbalanced data (Searle et al. 1992). Such extensions are implemented in the OPEP software (Baradat and Labbé 1995), which was used for the computations. The sampled regions represented three great geographic zones [East (Turkey, Near-East), Cyrenaique (Libya) and West (Maghreb, France, Italy, Spain)]. For the computations, three subdivisions were considered: geographic zone, country and population. The ANOVA model corresponding to the studied variability is written below:

$$y_{ijkl} = \mu + s_i + c_{ij} + p_{ijk} + e_{ijkl}.$$

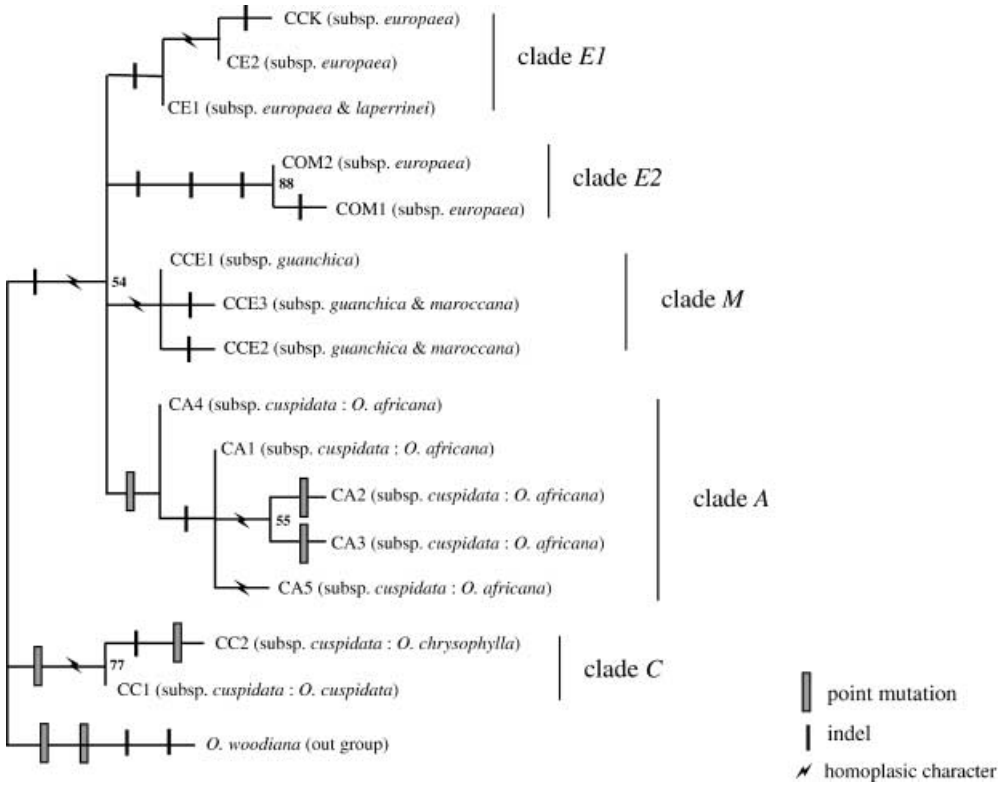

Fig. 1 Chlorotype consensus phylogram constructed with Wagner's parsimony. Each bar represents a mutational event

In this model s_i is the effect of the geographic zone *i*, c_{ij} is the effect of region *j* in the geographic zone *i*, p_{ijk} is the effect of population *k* in region *j*, and e_{ijkl} is the effect of individual *l* in the population *k*, with the corresponding variances: σ_s^2, $\sigma_{c|s}^2$, $\sigma_{p|c}^2$ and σ_e^2.

According to this model, we computed the three following parameters:

$$F_{st}^{(1)} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{c|s}^2 + \sigma_{p|c}^2 + \sigma_e^2}$$

which measures the differentiation due to the geographic zone,

$$F_{st}^{(2)} = \frac{\sigma_s^2 + \sigma_{c|s}^2}{\sigma_s^2 + \sigma_{c|s}^2 + \sigma_{p|c}^2 + \sigma_e^2}$$

which measures the differentiation due to the geographic zone and the region,

$$F_{st}^{(3)} = \frac{\sigma_s^2 + \sigma_{c|s}^2 + \sigma_{p|c}^2}{\sigma_s^2 + \sigma_{c|s}^2 + \sigma_{p|c}^2 + \sigma_e^2}$$

which measures the differentiation due to the geographic zone, the region and the population.

Within the subspecies *europaea*, the pattern of variability due to the presence-absence of a particular chlorotype was also measured. The measures of allele contributions to differentiation were the intra-class correlation coefficients computed on the 0 / 1 variates. They are referenced as $\theta_{st}^{(1)}$, $\theta_{st}^{(2)}$ and $\theta_{st}^{(3)}$, corresponding to $F_{st}^{(1)}$, $F_{st}^{(2)}$ and $F_{st}^{(3)}$, respectively. Appendix 1 shows the relationships between F_{st} and θ_{st}

Standard errors on the estimates of F_{st} and θ_{st} were computed by the Jackknife method (Shao and Tu 1995). We used the sampling strategy recommended by Weir (1990) for *F*-statistics, when inferences are to be drawn from the particular set of populations sampled, each new estimate being found by omitting one individual at a time from the complete data set.

Results

cpDNA polymorphism

Considering all polymorphic sites, a total of 15 chlorotypes was identified in the *O. europaea* complex (Table 2). They were distinguished with six cpDNA fragment-restriction enzyme combinations. However, the QR fragment restricted by *Taq*I and *Hae*III plus the ccmp5 and ccmp7 microsatellite polymorphisms were sufficient to distinguish the chlorotypes identified in the subspecies *europaea*, *laperrinei*, *guanchica* and *maroccana*. These were used to analyse the 527 individuals from the Mediterranean and North Africa (Canary Islands, South Morocco and Hoggar). In olive and oleaster, five chlorotypes were recognised: CE1, CE2, COM1, COM2 and CCK.

Phylogenetic relationships between the chlorotypes

Two point mutations and six multi-state length variants were phylogenetically informative because they shared at least two chlorotypes (Table 2). After a heuristic search, 18 most-parsimonious trees were obtained using the Wagner parsimony analysis. The CI value was 0.9 when excluding uninformative characters (autapomorphies). Homoplasy was detected for two characters corresponding to the length variants (ccmp5 and TC-indel). The consensus tree is presented and only bootstrap values higher than 50% were added on the nodes (Fig. 1).

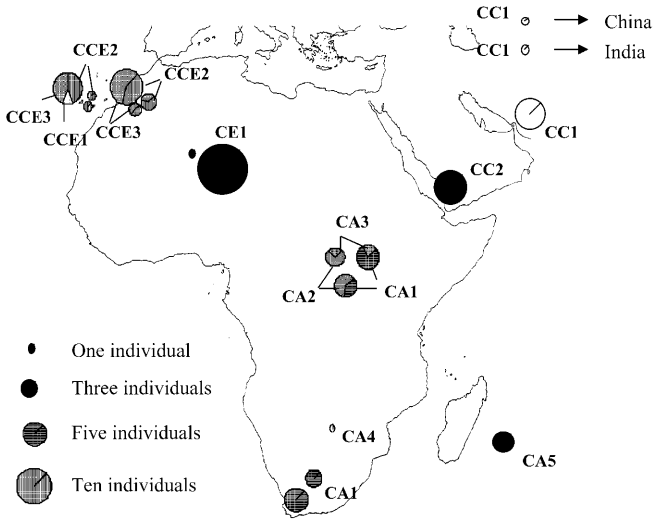

Fig. 2 Distribution of the chlorotypes in the *O. europaea* taxa belonging to the subspecies *cuspidata*, *laperrinei*, *maroccana* and *guanchica*

Five clades of chlorotypes can be distinguished: (1) the clade *C* characterising the taxa *O. cuspidata* and *O. chrysophylla* with two chlorotypes, (2) the clade *A* with five chlorotypes characterising the taxon *O. africana*, (3) the clade *M* characterising the subspecies *guanchica* and *maroccana*, the clade *E1*, with three chlorotypes characterising Mediterranean olives (subsp. *europaea*) and Saharan olives (subsp. *laperrinei*), (4) and the clade *E2*, with two chlorotypes characterising West Mediterranean olives.

Bootstrap analysis supports the distinction of clade *C*, which was based on two mutation events: one indel of about 300 bp (K2Q-indel) and a ccmp7 microsatellite polymorphism (Table 2, Fig. 1). Clade *A* was separated by one mutation event from the remaining chlorotypes. Clade *E2* differed from the other chlorotypes by three specific length mutations (QR-indel-2, QR-indel-3 and ccmp7-indel; Table 2). In contrast, the clades *E1* and *M* are not well-supported. Indeed, the phylogenetic position of CCK, with CE1 and CE2, remains to be verified due to the ccmp5 homoplasic character plus one autapomorphy for the QR-indel-3 character (Table 2). The lineage CCK could be also attributed to the clade *M*. (Fig. 1, Table 2).

Geographic distribution of *O. europaea* chlorotypes

Each of the five clades was located in a specific geographic zone: the chlorotypes of clade *A* were detected in Central and Southern Africa, and on *La Réunion* Island, the chlorotypes of clade *C* were found in Asia, whereas the clades *E1*, *E2* and *M* are specific from the Mediterranean area and from North Africa (Figs. 2, 3). For the clade *A*, the CA1 chlorotype was found in South Africa and Kenya. Each of the four other chlorotypes occurred in a specific area: CA2 and CA3 in Central Africa (Kenya), CA4 in Zimbabwe and CA5 within *La Réunion* (Fig. 2). Similarly, for clade *C*, two other chlorotypes were found in Asia and spread over specific areas: CC2 in Yemen, CC1 in Iran, in Northern India and in China.

Considering clade *E1*, one chlorotype (CE1) in *O. e.* subsp. *laperrinei* from the Hoggar (Fig. 2) was also found in oleasters of the Eastern Mediterranean (Fig. 3A). In contrast, in the Central and Western Mediterranean region (from Libya to Spain and Morocco) oleaster had four chlorotypes (CE1 and CCK from clade *E1*; COM1 and COM2 from clade *E2*). COM1 was the most-frequent chlorotype except in some populations (see Fig. 3A). COM2 was restricted to one population from Southern Sicily. CCK was predominant in Algeria and Tunisia but was also found in France, Spain and Morocco. Western populations sampled in proximity to olive groves had the CE1 chlorotype. In cultivated forms, this chlorotype was predominant in all sampled sites (Fig. 3B). Only some cultivars from the Western Mediterranean displayed the COM1 or CCK chlorotype. Lastly, the three chlorotypes belonging to clade *M* (CCE1, CCE2 and CCE3) characterised the subspecies *guanchica* and *maroccana* and were found only in the Canary Islands and in Southern Morocco (High-Atlas). In Southern Morocco, *O. e.* subsp. *maroccana* individuals were in many places located in parapatry with oleasters and cultivated olives.

When comparing chlorotype distribution in cultivars and oleasters, a discordance can be observed (Fig. 3): CE1 is the only chlorotype found in the oleasters from the East whereas it is almost absent in the West except when the material was sampled close to olive groves; i.e. Moulay-Idriss, Montpeyroux or Ogliastro. Conversely, in cultivars, CE1 was frequent whatever their origin, but was less frequent in the West due to the presence of COM1 and CCK. Lastly, CE2 was detected only in a few cultivars in the East and West Mediterranean.

Genetic structure of Mediterranean olive trees

The oleasters exhibited a large total differentiation [$F_{st}^{(3)}$ = 0.72], with a prevailing influence of the geographic zone [$F_{st}^{(1)}$ = 0.54] (Table 3). Table 3 also shows very contrasting contributions of the four alleles for this genetic differentiation. CE1 showed marked differentiation due to an East-West cline in gene frequency. At the extreme opposite, COM2 reflected variability concentrated at the between-population level. The COM1 allele showed an intermediate pattern with a major role in the great geographic zone. The variation of CCK was equally explained by region and population.

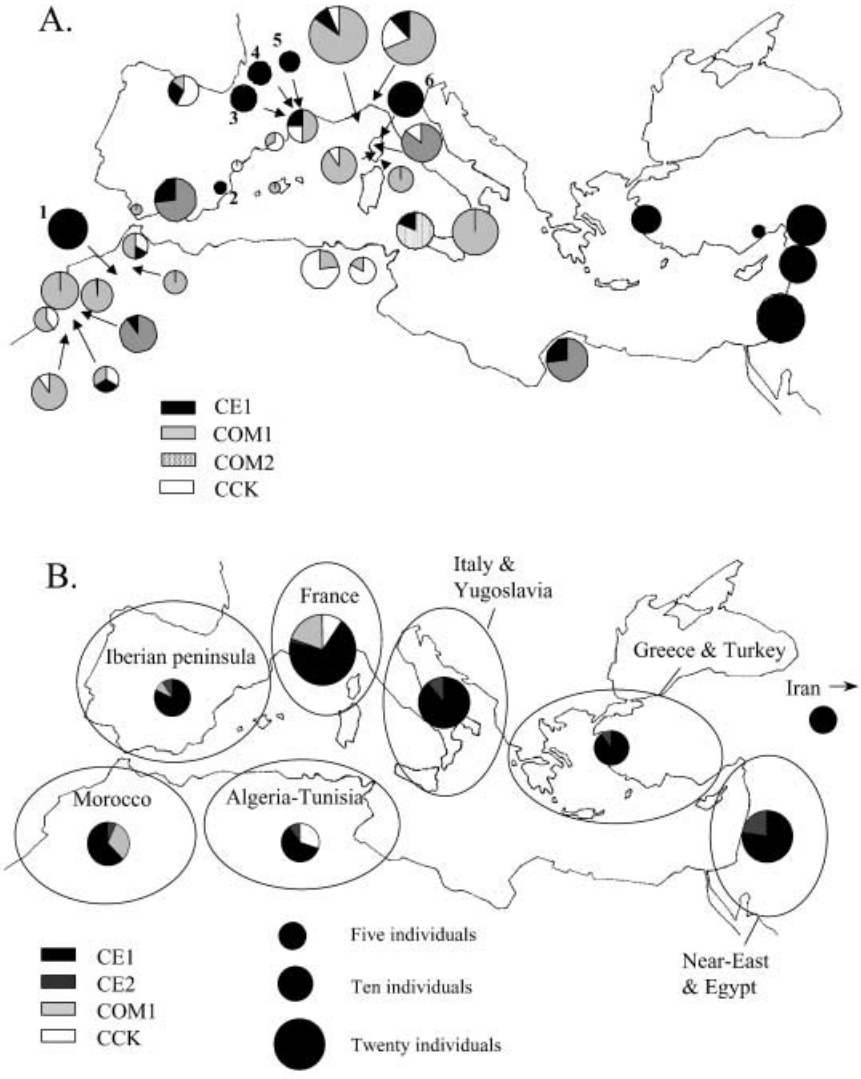

Fig. 3 Distribution of the chlorotypes in the oleasters (**A**), and in the cultivated forms (**B**). The oleaster populations 1 to 6 are considered as feral populations

Table 3 Olive chloroplast DNA differentiation in the subspecies *europaea* according to geographic zone, region, or population for the four chlorotypes and for each chlorotype. $F_{st}^{(1)}$: differentiation due to geographic zone, $F_{st}^{(2)}$: differentiation due to geographic zone, and region, $F_{st}^{(3)}$: differentiation due to geographic zone, region and population. $\theta_{st}^{(1)}$: allele contribution to differentiation (geographic zone), $\theta_{st}^{(2)}$: allele contribution to differentiation (geographic zone and region), $\theta_{st}^{(3)}$: allele contribution to differentiation (geographic zone, region and population). In brackets are given the lower and upper limits of the confidence interval (95% level)

Subsp. *europaea* differentiation			
Differentiation based on 4 alleles	$F_{st}^{(1)}$	$F_{st}^{(2)}$	$F_{st}^{(3)}$
	0.538^{***} (0.450–0.616)	0.565^{***} (0.483–0.647)	0.720^{***} (0.653–0.787)
Differentiation for each allele	$\theta_{st}^{(1)}$	$\theta_{st}^{(2)}$	$\theta_{st}^{(3)}$
CE1	0.818^{***} (0.747–0.889)	0.818^{***} (0.747–0.889)	0.831^{***} (0.768–0.894)
COM1	0.474^{***} (0.378–0.570)	0.474^{***} (0.378–0.570)	0.703^{***} (0.478–0.928)
COM2	0.000	0.036 *ns*	0.840^{***} (0.614–1.000)
CCK	0.000	0.228^{*} (0.043–0.412)	0.413^{***} (0.228-0.597)

***: P <0.001; *: P <0.05; *ns*: non significant

Discussion

Phylogenetic relationships between all the *O. europaea* chlorotypes

We recognised five clades in the cpDNA phylogram (Fig. 1), while Lumaret et al. (2000) and Besnard and Bervillé (2000) detected only three clades using cpDNA RFLPs. The clades *E2* and *M* are new and point out the complex origins for olive.

The comparison of this new cpDNA phylogram with previous taxonomic information shows several incongruities. First, this result differs from that of Green and Wickens (1989), who have not differentiated between

Asian and African forms. Second, the CCK western chlorotype was weakly assigned to the CE1-CE2 group. This result is incongruent with mtDNA and cpDNA RFLP data, which enabled them to distinguish the CCK chlorotype as a very-distinct lineage (Besnard et al. 2000). Consequently, the phylogenetic position of this chlorotype should be verified using other polymorphisms. The haplotypes found in olive and oleaster from the Mediterranean and North African regions are placed in two distinct clades, and maybe three, assuming CCK as another distinct lineage. These clusters most likely reflect distinct ancestral olive populations. These populations could have diverged due to habitat fragmentation or resulted from different olive introductions into the Mediterranean Basin.

It is noteworthy that *O. e.* subsp. *laperrinei* from Hoggar displays the CE1 chlorotype, which is also detected in Mediterranean olives. In contrast, Mediterranean and Saharan olives can be distinguished from one another based on nuclear data (Angiolillo et al. 1999; Hess et al. 2000; Besnard et al. 2001a). These two taxa would thus appear to have a common maternal ancestor (Besnard and Bervillé 2000). The CE1 chlorotype may have originated in a population overlapping Saharan and Mediterranean areas during a Quaternary favourable period. However, Lumaret et al. (2000) based on RFLPs found that *O. e.* subsp *laperrinei* displayed the same chlorotype as *O. e.* subsp. *maroccana*, but this is probably due to confusion between these two subspecies (Besnard and Bervillé 2002).

Evidence for an ancient hybridisation zone in Eastern Africa and Arabia

Natural barriers separate the five olive clades: for instance, the Red Sea between *C* and *A*, and deserts separating *E1*, *E2* and *M* from *A* and *C*. Such barriers probably have a limited gene flow and contributed to the genetic differentiation between olives from these distinct areas. Nevertheless, for subspecies *cuspidata*, the dendrograms based on nuclear data are not consistent with cytoplasmic data. For the same individuals, AFLP and RAPD analyses revealed that *O. chrysophylla* (clade *C*) from Yemen is in an intermediary position between *O. cuspidata* (clade *C*) from Iran and *O. africana* (clade *A*) from Africa (Angiolillo et al. 1999; Besnard et al. 2001a). Moreover, *O. chrysophylla* based on RAPD is more related to *O. africana* and displays a very low number of specific markers in comparison with *O. cuspidata* and *O. africana*. Such a combination of nuclear and cytoplasmic markers in *O. chrysophylla* might be the result of hybridisation between *O. cuspidata* from Asia and *O. africana* from Africa. Furthermore, *O. e.* subsp. *laperrinei* is also suspected to be, or to have been, present in this region. This taxon is present in Jebel Marra (Sudan) and the form *O. somaliensis* Bak., in the Somalian massifs surrounding the Red Sea (Ogo and Medjourtine Mounts), has been considered to be an ecotype of *O. e.* subsp. *laperrinei* because of the white hairs on its lower leaf surface (Quézel 1995). Hence, these data suggest that the Eastern African-Arabian regions should contain hybrids between trees belonging to the *A*, *C* and *E1* cpDNA clades. Analyses of more trees originating from this area would be of a great interest to verify such a hypothesis.

Human influence on the genetic differentiation of the Mediterranean populations

To properly interpret the present-day genetic structure of *O. europaea* in the Mediterranean and in North Africa, it is important to understand what could have been the human effects on the genetic differentiation of the wild Mediterranean olive. The chlorotype distribution in oleasters shows marked spatial differentiation (F_{st} = 0.72). Three chlorotypes are specific to the Western Mediterranean (CCK, COM1 and COM2) and one to the Eastern Mediterranean (CE1). However, CE1 is sometimes also present in the Western Mediterranean, notably in wild populations geographically close (less than 1 km) to olive groves. To explain such a structure of oleaster populations, we can propose three main hypotheses: (1) the CE1 chlorotype is an ancestral lineage occurring in the East and West Mediterranean, but has been largely eliminated in the West due to genetic drift; (2) oleaster populations have migrated from West to East Mediterranean (this is not usually assumed) and tree dissemination has been associated with a reduction of chlorotype diversity (founder effect); (3) lastly, humans could have introduced the CE1 chlorotype in the West probably with the dispersal of olive cultivars from East to West Mediterranean. Based on nuclear (RAPDs) and cytoplasmic data, most of the cultivars found in the West and most of the forms sampled close to olive groves are related to oleasters from the Eastern Mediterranean (Besnard and Bervillé 2000, Besnard et al. 2001a). Thus, there is a linkage disequilibrium between the CE1 chlorotype and nuclear markers characteristic of Eastern oleasters (Besnard and Bervillé 2000), which can be explained by the common origin of these polymorphisms in a wild population from the East. This argues for the recent introduction of the CE1 chlorotype into the Western Mediterranean and refutes the two other hypotheses. Thus, western oleasters carrying CE1 could be a feral form (a naturalised form of previously cultivated olives or hybrid forms). A similar situation has also been reported for Western European honeybees (Garnery et al. 1998). Consequently, in western Europe, oleasters sampled close to olive groves are most-likely feral populations.

According to F_{st} values, chlorotype structure in wild Mediterranean olives is primarily due to large-scale between-region variation, mainly an East-West differentiation, and secondly to a mosaic distribution corresponding to a small-scale variation. Calculation of F_{st} values without including feral populations resulted in an attenuated small-scale variation (data not shown). The presence of feral populations is the result of an eastern introduction and a gene flow from olive groves towards wild

populations and may be interpreted as an artificial source of variation, which obscures the effect of migration involving initial and long-distance events.

Olive refugia in North Africa and Mediterranean

We propose an interpretation of the chlorotype distribution observed in Northern Africa and the Mediterranean. First, in Northwest Africa, *O. e.* subsp. *maroccana* and *O. e.* subsp. *guanchica* have related chlorotypes, which are restricted to limited geographic areas. These populations have preserved relatively high haplotype diversity characteristic of relic populations as already shown in white oaks (Dumolin-Lapègue et al. 1997a), or *Argania spinosa* (L.) Skeels (El Mousadik and Petit 1996). The subspecies *maroccana* and *guanchica* are slightly differentiated based on these data, indicative of their common ancestry. Second, strong genetic differentiation between the Eastern and Western Mediterranean was shown for the other Mediterranean chlorotypes. These cpDNA lineages probably mark the diffusion of the species from glacial refugia as already proposed by Besnard and Bervillé (2000).

Nevertheless, the organisation genetic diversity in the Mediterranean olive has a complex biogeographical history. Environmental changes, dispersal by humans, hybridisation among different taxa and human selection have all contributed to the spread and differentiation of Mediterranean olive populations. In the future, we plan to combine mitochondrial and chloroplast DNA information to check the pertinence of the considered polymorphisms and to better interpret the cytoplasmic DNA structure in the Mediterranean area (Besnard et al. 2002b).

In conclusion, this study enabled us to reveal a strong phylogeographic structure in the olive complex. Some features witness the existence of long-distance gene flows in this species during favourable times. Ancestral gene flows between Mediterranean and Central African olives could be identified. Knowledge about the origins of oleasters is an important point for a good understanding of olive domestication. Consequently, it should be interesting to study more samples from the Mediterranean and from Central Africa and to identify the spatial and temporal differentiation of the olive complex.

Acknowledgements The authors wish to thank all correspondents who provided the samples of olive trees and of the *Olea* species analysed in this study, namely: Dr. G. Abdullah, Prof. Al Hakimi, Dr. Amirouch, Dr. L. Amsellem, Dr. A.H. Arsel, Prof. R. Assaf, Dr. L. Baldoni, Dr. M. Ben Salem, Dr. C. Costa, Dr. N. Gamoudi, Prof. P.S. Green, Dr. El Hadi Khalef, Dr. H. Hosseinpour, Dr. L. Humeau, Dr. A. Moukhli, Dr. A. Ouksili, Dr. L. Rallo, Dr. M. Sanchez, Dr. A. Santos, Dr. H. Sommerlatte, Dr. P. Vargas. They particularly thank C. Breton, Prof. F. Dosba, Dr. E. Douzéry, Dr. F. Kjellberg, Dr. J. Thompson and Prof. P. Villemur for their support and comments. This work was supported by the FAIR European project CT 95-0689.

Appendix

Decomposition of allele contributions to the global genetic differentiation of haplotypes using a three-way nested classification

Starting from the general definition of Wright's F_{st} – correlation between two random alleles present at a given locus for two individuals of the same group – we can rewrite the above three-way nested ANOVA model for a given allele, *a*, among *A* possible alleles ($1 \le a \le A$):

$$y_{ijkl}^{(a)} = \mu^{(a)} + s_i^{(a)} + c_{ij}^{(a)} + p_{ijk}^{(a)} + e_{ijkl}^{(a)}$$

Using the approach followed by Kempthorne (1957), we can explicit the three covariances between two individuals of the same group for presence-absence of the a^{th} allele:

$$\operatorname{cov}[y_{ijkl}^{(a)}, y_{ij'k'l'}^{(a)}] = \operatorname{cov}[s_i^{(a)}, s_i^{(a)}] = \sigma_s^{2(a)}$$

$$\operatorname{cov}[y_{ijkl}^{(a)}, y_{ijk'l'}^{(a)}] = \operatorname{cov}[s_i^{(a)}, s_i^{(a)}] + \operatorname{cov}[c_{ij}^{(a)}, c_{ij}^{(a)}] = \sigma_s^{2}$$

$$\operatorname{cov}[y_{ijkl}^{(a)}, y_{ijkl'}^{(a)}] = \operatorname{cov}[s_i^{(a)}, s_i^{(a)}] + \operatorname{cov}[c_{ij}^{(a)}, c_{ij}^{(a)}] + \operatorname{cov}[p_{ijk}^{(a)}, p$$

$$\text{As :} \quad \sigma^2[y_{ijkl}^{(a)}] = \sigma_s^{2(a)} + \sigma_{c|s}^{2(a)} + \sigma_{p|c}^{2(a)} + \sigma_e^{2(a)},$$

the three intra-class correlation coefficients for presence-absence of the a^{th} allele may be written:

$$\theta_{st}^{(a)1} = \frac{\sigma_s^{2(a)}}{\sigma_s^{2(a)} + \sigma_{c|s}^{2(a)} + \sigma_{p|c}^{2(a)} + \sigma_e^{2(a)}}$$

$$\theta_{st}^{(a)2} = \frac{\sigma_s^{2(a)} + \sigma_{c|s}^{2(a)}}{\sigma_s^{2(a)} + \sigma_{c|s}^{2(a)} + \sigma_{p|c}^{2(a)} + \sigma_e^{2(a)}}$$

$$\theta_{st}^{(a)3} = \frac{\sigma_s^{2(a)} + \sigma_{c|s}^{2(a)} + \sigma_{p|c}^{2(a)}}{\sigma_s^{2(a)} + \sigma_{c|s}^{2(a)} + \sigma_{p|c}^{2(a)} + \sigma_e^{2(a)}}$$

The system of four equations from mean squares expectations is given below:

$$E(MS_S) = \sigma_e^{2(a)} + k_1\sigma_{p|c}^{2(a)} + k_2\sigma_{c|s}^{2(a)} + k_3\sigma_s^{2(a)}$$

$$\mathrm{E(MS_{c|s})} = \sigma_{\mathrm{e}}^{2(\mathrm{a})} + \mathrm{k_4}\sigma_{\mathrm{p|c}}^{2(\mathrm{a})} + \mathrm{k_5}\sigma_{\mathrm{c|c}}^{2(\mathrm{a})},$$

$$\mathrm{E(MS_{p|c})} = \sigma_{\mathrm{e}}^{2(\mathrm{a})} + \mathrm{k_6}\sigma_{\mathrm{p|c}}^{2(\mathrm{a})},$$

$$\mathrm{E(MS_e)} = \sigma_{\mathrm{e}}^{2(\mathrm{a})}.$$

The estimate of within-population variance, $\hat{\sigma}_{\mathrm{e}}^{2(\mathrm{a})}$, is directly given by the fourth mean square. The vector of estimates of the first three variances may be expressed in matrix form:

$$\begin{bmatrix} \hat{\sigma}^2_s \\ \hat{\sigma}^2_{c|s} \\ \hat{\sigma}^2_{p|c} \end{bmatrix} = \begin{bmatrix} k_1 & k_2 & k_3 \\ 0 & k_4 & k_5 \\ 0 & 0 & k_6 \end{bmatrix}^{-1} \begin{bmatrix} MS_s - MS_e \\ MS_{c|s} - MS_e \\ MS_{p|c} - MS_e \end{bmatrix}.$$

If we consider the *A* alleles, as the presence of one of them excludes the *A*-1 other possibilities, all the variances written above are summed-up over these *A* possible occurrences:

$$\sigma^2_s = \sum_{a=1}^{A} \sigma^{2(a)}_s, \sigma^2_{c|s} = \sum_{a=1}^{A} \sigma^{2(a)}_{c|s},$$

$$\sigma^2_{p|c} = \sum_{a=1}^{A} \sigma^{2(a)}_{p|c}, \text{ and } \sigma^2_e = \sum_{a=1}^{A} \sigma^{2(a)}_e,$$

and the F_{st} value corresponding to each θ_{st} may be obtained by replacing the variances corresponding to the ath allele by the sum of the A variances. This global estimate can of course be obtained by summing all the above mean squares over the A alleles and by solving the corresponding equation set. This additivity also allows estimating the contribution of groups of alleles over several loci, if any. With two alleles, all their variance components are equal, and both the numerator and the denominator of F_{st} are multiplied by two, compared to those of θ_{st}, and thus $\theta_{st}=F_{st}$.

References

Angiolillo A, Mencuccini M, Baldoni L (1999) Olive genetic diversity assessed using amplified fragment length polymorphisms. Theor Appl Genet 98:411–421

Baradat P, Labbé T (1995) OPEP: un logiciel intégré pour l'amélioration des plantes pérennes. In: Traitements statistiques des essais de sélection. Stratégies de sélection des plantes pérennes, CIRAD-CP, Montpellier, pp 303–330

Besnard G, Baradat P, Chevalier D, Tagmount A, Bervillé A (2001a) Genetic differentiation in the olive complex (*Olea europaea*) revealed by RAPDs and RFLPs in the rRNA genes. Genet Res Crop Evol 48:165–182

Besnard G, Bervillé A (2000) Multiple origins for Mediterranean olive (*Olea europaea* L. subsp. *europaea*) based upon mitochondrial DNA polymorphisms. C R Acad Sci sér III 323:173–181

Besnard G, Bervillé A (2002) On chloroplast DNA variations in the olive (*Olea europea* L.) complex: comparison of RFLP and PCR polymorphisms. Theor Appl Genet (in press)

Besnard G, Breton C, Baradat P, Khadari B, Bervillé A (2001b) Cultivar identification in the olive (*Olea europaea* L.) based on RAPD markers. J Am Soc Hort Sci 126:668–675

Besnard G, Green PS, Bervillé A (2002a) The genus *Olea*: molecular approaches of its structure and relationships to other *Oleaceae*. Acta Botanica Gallica 149: (this issue)

Besnard G, Khadari B, Baradat P, Bervillé A (2002b) Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive complex (*Olea europea* L.). Theor Appl Genet DOI 10.1007/b00122-002-0868-6

Besnard G, Khadari B, Villemur P, Bervillé A (2000) Cytoplasmic male sterility in the olive (*Olea europaea* L.). Theor Appl Genet 100:1018–1024

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997a) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487

Dumolin-Lapègue S, Pemonge MH, Petit RJ (1997b) An enlarged set of consensus primers for the study of organelle DNA in plants. Mol Ecol 6:189–191

El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. Mol Ecol 5:547–555

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Garnery L, Franck P, Baudry E, Vautrin D, Cornuet JM, Solignac M (1998) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. Genet Sel Evol 30:S31–S47

Green PS, Wickens GE (1989) The *Olea europaea* complex. The Davis and Hedge Festschrift, Edinburg University Press, pp 287–299

Hess J, Kadereit JW, Vargas P (2000) The colonization history of *Olea europaea* L. in Macaronesia based on internal transcribed spacer 1 (ITS-1) sequences, randomly amplified polymorphic DNAs (RAPD), and intersimple sequence repeats (ISSR). Mol Ecol 9:857–868

Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87–112

Huntley B, Birks HJB (1983) An atlas of the past and present pollen maps of Europe: 0–13,000 years ago. Cambridge University Press, Great Britain

Kempthorne (1957) An Introduction to Genetic Statistics. Wiley, New York

Khadari B, Bervillé A (2001) Genetic diversity of Moroccan cultivated olive using RAPD markers. Acta Hort 546:439–442

Kislev ME, Nadel D, Carmi I (1992) Epipalaeolithic (19,000 BP) cereal and fruit diet at Ohalo II, Sea of Galilee, Israel. Rev Palaeobot Palynol 73:161–166

Lumaret R, Amane M, Ouazzani N, Baldoni L, Debain C (2000) Chloroplast DNA variation in the cultivated and wild olive taxa of the genus *Olea*. Theor Appl Genet 101:547–553

Quézel P (1978) Analysis of the flora of Mediterranean and Sahara Africa. Ann Missouri Bot Gard 65:479–534

Quézel P (1995) La flore du Bassin Méditerranéen: origine, mise en place, endémisme. Ecol Mediterranea 21:19–39

Searle SR, Casella G, Mc Culloch CE (1992) Variance components. Wiley, New York

Shao J, Tu D (1995) The Jackknife and the Bootstrap. Springer, New York

Solari ME, Vernet JL (1992) Late glacial and Holocene vegetation of the Corbières based on charcoal analysis at the Cova de L'Espérit (Salses, Pyrénées Orientales, France). Rev Palaeobot Palynol 71:111–120

Swofford DL (1998) PAUP, Phylogenetic Analysis Using Parsimony (and other methods), version 4. Sinauer Associates, Sunderland, Massachusetts

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7:453–464

Terral JF, Arnold-Simard G (1996) Beginnings of olive cultivation in eastern Spain in relation to Holocene bioclimatic changes. Quat Res 46:176–185

Vargas P, Muñoz Garmendia F, Hess J, Kadereit J (2001) *Olea europaea* subsp. *guanchica* and subsp. *maroccana* (Oleaceae), two new names for olive tree relatives. Ann Jard Bot Madrid 58:360–361

Weir BS (1990) Genetic data analysis. Sinauer, Sunderland

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19

Zohary D, Spiegel-Roy P (1975) Beginnings of fruit growing in the Old World. Science 187:319–327