New Phytologist

Research

Origin and expansion of the allotetraploid *Aegilops geniculata*, a wild relative of wheat

Nils Arrigo[1], François Felber[1], Christian Parisod[1], Sven Buerki[1], Nadir Alvarez[2], Jacques David[3] and Roberto Guadagnuolo[1]

[1]Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, 11 rue Emile-Argand, 2009 Neuchâtel, Switzerland; [2]Department of Ecology and Evolution, University of Lausanne – UNIL Sorge, Biophore Building, 1015 Lausanne, Switzerland; [3]UMR Diversité et Adaptation des Plantes Cultivées, SupAgro, 2 Place P. Viala, Domaine de Melgueil, 34130 Mauguio, France

Author for correspondence:
Nils Arrigo
Tel: +41 (0)32 718 2338
Email: nils.arrigo@unine.ch

Received: *29 March 2010*
Accepted: *26 April 2010*

New Phytologist (2010)
doi: 10.1111/j.1469-8137.2010.03328.x

Key words: *Aegilops geniculata* (=*Ae. ovata*), amplified fragment length polymorphism, chloroplast DNA sequencing, human-driven migration, introgression, phylogeography, polyploidy, postglacial recolonization.

Summary

- This study reconstructs the phylogeography of *Aegilops geniculata*, an allotetraploid relative of wheat, to discuss the impact of past climate changes and recent human activities (e.g. the early expansion of agriculture) on the genetic diversity of ruderal plant species.
- We combined chloroplast DNA (cpDNA) sequencing, analysed using statistical parsimony network, with nonhierarchical *K*-means clustering of amplified fragment length polymorphism (AFLP) genotyping, to unravel patterns of genetic structure across the native range of *Ae. geniculata*. The AFLP dataset was further explored by measurement of the regional genetic diversity and the detection of isolation by distance patterns.
- Both cpDNA and AFLP suggest an eastern Mediterranean origin of *Ae. geniculata*. Two lineages have spread independently over northern and southern Mediterranean areas. Northern populations show low genetic diversity but strong phylogeographical structure among the main peninsulas, indicating a major influence of glacial cycles. By contrast, low genetic structuring and a high genetic diversity are detected in southern Mediterranean populations. Finally, we highlight human-mediated dispersal resulting in substantial introgression between resident and migrant populations.
- We have shown that the evolutionary trajectories of ruderal plants can be similar to those of wild species, but are interfered by human activities, promoting range expansions through increased long-distance dispersal and the creation of suitable habitats.

Introduction

The influence of geological and evolutionary processes on the distribution of organisms is becoming increasingly well understood thanks to the development of phylogeographical studies focusing on the genetic signal left by past events (e.g. Wiens & Donoghue, 2004). Accordingly, the impact of the last glacial maximum (i.e. 15 000 yr ago) on intraspecific genetic variation, as well as on the location of glacial refugia and postglacial recolonization pathways, has been reported for various species across diverse areas (e.g. Taberlet *et al.*, 1998; Hewitt, 2004; Hampe & Petit, 2005; Médail & Diadema, 2009). By contrast, little is known about the influence of humans on past events, such as the survival, dispersal or range expansion–contraction of wild species.

Agricultural practices emerged in the eastern Mediterranean area (i.e. the Fertile Crescent) *c.* 10 000 yr ago, and subsequently spread rapidly around the Mediterranean Sea and across Europe (Zeder, 2008). Neolithic farmers might have influenced the current distribution of plant or animal populations in multiple ways, acting as dispersal agents through exchanges or interfering with natural recolonization by modifying the landscape (e.g. deforestation). Given that the Mediterranean region has a long history of farmer occupation, ancient as well as recent human activities have probably had a significant impact on the genetic diversity of plants associated with agroecosystems,

such as *Abies* sp. (Parducci *et al.*, 2001), *Castanea sativa* (Fineschi *et al.*, 2000), *Festuca pratensis* (Fjellheim *et al.*, 2006), *Lolium* sp. (Balfourier *et al.*, 2000), *Nigella arvensis* (Bittkau & Comes, 2005) and *Olea europaea* (Besnard *et al.*, 2007).

Aegilops geniculata Roth is a wild relative of wheat and is currently abundant in agricultural landscapes (e.g. in field edges, along roadsides, in pastures or in disturbed areas) of the Mediterranean Basin (Van Slageren, 1994). The species is annual and largely autogamous (Hammer, 1980). The morphological features of spikes (e.g. awns and hairs) ensure an efficient zoochorous dispersal of seeds, which remain intact in the soil for years, supplying a consistent seed bank. Accordingly, human activities, such as pastoralism of goats and ovine livestock or cereal cultivation, might have contributed to long-distance seed dispersal in *Ae. geniculata* (Zaharieva *et al.*, 2001a). The species thus offers a good model to explore how humans might have interfered with natural expansion. Schematically, two contrasting hypotheses can be postulated to explain the present distribution of this species in the Mediterranean area: *Ae. geniculata* spread across the region before the last glacial maximum and recolonized its current distribution range from glacial refugia during the postglacial period (i.e. the last 15 000 yr), or *Ae. geniculata* accompanied Neolithic farming and dramatically expanded its range during the spread of agriculture. Unfortunately, no evidence yet available clearly favours or dismisses one of these hypotheses (Zaharieva *et al.*, 2001b).

Aegilops geniculata is an allotetraploid species ($2n = 4x =$ 28 chromosomes; genome formula MMUU), whose ancestors are *Ae. comosa* Sm. in Sibth. & Sm. var. *comosa* (genome MM) and *Ae. umbellulata* Zhuk. (genome UU; Van Slageren, 1994; Friebe *et al.*, 1999). Based on the current distribution of the diploid progenitor species, most authors have suggested an eastern Mediterranean origin of *Ae. geniculata* (Chennaveeraiah, 1960; Waines & Barnhart, 1992; Van Slageren, 1994; Resta *et al.*, 1996). Previous molecular studies have also corroborated this hypothesis (Zaharieva *et al.*, 2001a). Nevertheless, its precise origin remains elusive and might have involved multiple polyploidy events (Meimberg *et al.*, 2009). As *Ae. geniculata* constitutes an important reservoir of genes for wheat improvement (Zaharieva *et al.*, 2001a, 2003), a better understanding of the origin of the species and the evolutionary processes shaping genetic variation across its native distribution range is critical for genetic resource conservation. Accordingly, the present study investigates the spatiotemporal dynamics of native populations of *Ae. geniculata* over the whole Mediterranean area. In particular, using both chloroplast DNA (cpDNA) sequencing and amplified fragment length polymorphism (AFLP), we aim to describe the distribution of genetic variation across the Mediterranean Basin to characterize the origin of *Ae. geniculata*, unravel the species' expansion patterns and the putative human imprints, and highlight the evolutionary processes acting on genetic diversity. We report a strong phylogeographical pattern, suggesting that the polyploid *Ae. geniculata* evolved multiple times in the area of the Bosphorus Strait, and that independent lineages expanded naturally towards the north or south of the Mediterranean Sea. Accordingly, human-mediated migration was probably limited, but immigrated individuals seem to have influenced genetic variation locally in resident populations through considerable introgression.

Materials and Methods

Sampling area and plant material investigated

The sampling area covers the whole Mediterranean Basin (Fig. 1) and has been further divided into six natural biogeographical regions defined on the basis of natural frontiers according to Presti & Oberprieler (2009): the Iberian Peninsula (IBP), the Franco-Italian Peninsula (FIP), the Balkans (HR), the Bosphorus Strait (BST), the Levantine (LEV) and, finally, Northern Africa (NAF).

A total of 411 *Aegilops geniculata* Roth individuals was sampled in 146 locations. Half of the samples (194 individuals) were collected as seeds in 22 naturally occurring populations from Spain, France, Italy, Croatia and Turkey (nine samples per population on average), and were cultivated at the Botanical Garden of the University and the City of Neuchâtel, Switzerland. The sampling was completed with 217 DNA samples from germplasm accessions (deposited at IAV, Rabat, Morocco; ICARDA, Aleppo, Syria; IIPGR, Sadovo, Bulgaria; INIA, Madrid, Spain; INRA, Montpellier, France; and IPK, Gatersleben, Germany) collected in 121 different locations (two samples per population on average; Supporting Information Table S1). DNA was extracted from 10 mg of silica-dried leaves using a standard cetyltrimethylammonium bromide (CTAB) protocol (Chen & Ronald, 1999), and the concentration was standardized at 10 ng ml^{-1}.

Chloroplast DNA sequences were characterized in a subset of 125 individuals, including 115 samples of *Ae. geniculata* (one to three individuals per population) as well as 10 outgroup samples (*Ae. neglecta* Req. ex Bertol., *Ae. biuncialis* Vis., *Ae. triuncialis* L., *Triticum aestivum* L. and *T. turgidum* L.). Available sequences (GenBank accessions EU012758–EU012763; EU012766–EU012770; EU012967; EU012801–EU012803; Meimberg *et al.*, 2009) from *Ae. comosa* and *Ae. umbellulata*, the diploid progenitor species of *Ae. geniculata* (Friebe *et al.*, 1999), were also included. AFLP analyses were performed on the complete set of *Ae. geniculata* samples (411 individuals). A complete description of the samples, as well as cpDNA and AFLP, is presented in Table S1.

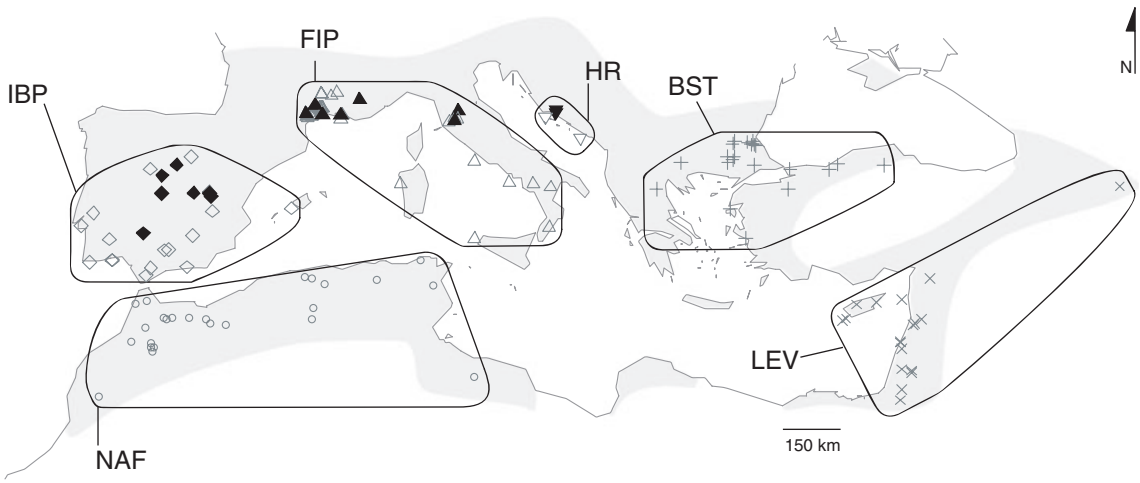

Fig. 1 Sampled populations of *Aegilops geniculata* across the Mediterranean area. Populations are labelled according to the geographical area of collection (BST, Bosphorus Strait; FIP, Franco-Italian Peninsula; HR, Balkans; IBP, Iberian Peninsula; LEV, Levantine; NAF, Northern Africa). The sampling effort for each location is denoted by distinct symbols (grey, less than five samples per location; black, more than five samples per location). The species' natural range is shaded in grey (simplified from Van Slageren, 1994).

cpDNA sequences

Three contiguous loci of the small single-copy region of cpDNA were sequenced: the *ndhF-rpl32* intergenic spacer (IGS), the *rpl32* gene and the *rpl32-trnL* IGS. The whole region was amplified with the trnL(UAG) and ndhF primers (Shaw *et al.*, 2005, 2007) and sequenced using *Aegilops*-specific primers (internF, 5′-CCAATTCTTATCTCTTT-CTGAAAG-3′; internR, 5′-GCTTTGCCCAATAGAA-ACACA-3′). PCR amplifications were performed on 30 ng of genomic DNA in a 40 µl solution (1 × PCR buffer, 0.2 mg ml^{-1} BSA, 1.5 mM $MgCl_2$, 200 mM deoxynucleoside triphosphates (dNTPs), 0.2 mM primers and 2 U Taq polymerase (Promega)) using the '*rpl16*' PCR cycle proposed by Shaw *et al.* (2005, 2007). The sequencing was performed on an ABI 3730XL (Applied Biosystems, Foster City, CA, USA; service provided by Macrogen Inc. Seoul, South Korea) and the base calling was checked using ChromasPro (version 1.34, Technelysium Ltd. Helensvale, Qld, Australia). Sequences were deposited in GenBank under accession numbers GQ250449–GQ250573.

cpDNA haplotype network

The 125 cpDNA sequences produced in this study, as well as 15 additional sequences of *Ae. comosa* and *Ae. umbellulata* from GenBank (Meimberg *et al.*, 2009), were aligned using ClustalX (Thompson *et al.*, 1997). The alignment was finally adjusted manually in BioEdit (Hall, 1999). All three cpDNA regions were analysed as a single partition to produce a total evidence phylogenetic analysis (*sensu* Kluge, 1989). We performed a statistical parsimony network (Templeton *et al.*, 1992) with TCS version 1.21 (Clement *et al.*, 2000). Substitutions were given a weight of unity. Indels (recoded as a fifth state character using FastGap version 1.2; Borchsenius, 2009) and inversions were given a weight of two.

AFLP amplification and scoring

AFLP reactions were performed following the protocol of Gugerli *et al.* (2008). Two selective PCR primer pairs were used (*Eco*RI-ACT/*Mse*I-CTG and *Eco*RI-AGT/*Mse*I-CAT), with FAM-labelled *Eco*RI primers. PCR products were mixed with a 500 LIZ size ladder and analysed with an ABI 3730XL capillary sequencer (service provided by Macrogen Inc.). In order to detect and calculate the size of AFLP bands, raw electropherograms were analysed using PeakScanner (ABI) with default parameters except a light peak smoothing. A binary matrix of AFLP band presence (1) – absence (0) was built using the automated scoring RawGeno package (R CRAN; Arrigo *et al.*, 2009) with the following parameters: scoring range, 100–400 bp; minimum intensity, 100 rfu; minimum bin width, 0 bp; maximum bin width, 2 bp. Closely sized bins were eliminated. As recommended by Vekemans *et al.* (2002), the correlation between AFLP band size and frequency among samples was assessed to check for potential homoplasy.

Individuals were randomly distributed in 96-well plates in order to produce a reliable AFLP dataset. Five samples and one blank control were included on each plate to check that between-plate variability was minimized during PCRs. In addition, 58 samples (representing 14% of the final dataset) were randomly chosen from each plate and replicated to calculate the error rate (Bonin *et al.*, 2004). Bands that were clearly not reproducible were discarded from further analyses.

Patterns of AFLP diversity

The diversity patterns of *Ae. geniculata* were investigated by computing the Shannon index, as recommended by Bussel (1999). Spatial patterns of genetic diversity were inferred using a 'sliding window' that considered a 100 km grid over the whole sampling area (i.e. one grid point for each 100 km, in latitude and longitude) and computed the Shannon index by considering samples located within a 150 km perimeter around each grid point. In order to provide an unbiased Shannon index under unequal sampling among areas, computations were bootstrapped by a 1000 times resampling of 10 samples per grid point. Computations were performed using custom R scripts (R Development Core Team, 2009, script available on request). Finally, the regional Shannon index was computed by averaging the values obtained from the moving window analysis for each sampling area (i.e. IBP, FIP, HR, BST, LEV and NAF). These regional diversities were compared with a *post hoc* Tukey honest significant differences test.

Isolation by distance was tested over the whole sampling area, as well as within the different geographical districts, by Mantel tests between the genetic and geographical distances among sampled individuals using R (packages 'stats' and 'vegan'; R Core Development Team, 2009). Genetic distances were assessed by the Jaccard distance on the binary AFLP matrix, and geographical distances were calculated as the Euclidean distance based on spherical projection centred on the sampling area of the geographical coordinates. Significance was assessed by 1000 permutations.

AFLP genetic structure

Because *Ae. geniculata* is an autogamous allotetraploid species, Hardy–Weinberg equilibrium is unexpected. We therefore investigated the genetic structure using a nonmodel-based algorithm. Nonhierarchical *K*-means clustering (Hartigan & Wong, 1979) was performed using R (package 'stats'; R Core Development Team, 2009, script available on request). This technique assigns individuals to a defined number of genetic groups (hereafter *K*) in order to maximize the intergroup variance (measured here as the inertia; Legendre & Legendre, 1998). It has already been successfully applied in a phylogeographical framework based on AFLP markers (Burnier *et al.*, 2009). Here, we performed 100 000 independent runs (i.e. starting from random points) for each assumed value of *K* (ranging between 1 and 11) and recorded the intergroup inertia of each run. We adapted the strategy proposed by Evanno *et al.* (2005) to select the most likely number of groups using intergroup inertia as a proxy of clustering accuracy (see details in Notes S1). The proportion of individuals assigned to each *K*-means group within populations was displayed on a map. As a complementary representation of genetic structure, a principal coordinates analysis was computed on a Jaccard distance matrix between individuals and labelled with the *K*-means groups.

Comparison of cpDNA and AFLP phylogeographical signals

Association between datasets was investigated through a contingency table by comparing the occurrence of cpDNA haplotypes and AFLP groups defined with *K*-means. Associations were evaluated using chi-squared (χ^2) tests.

Results

cpDNA haplotypes in *Ae. geniculata*

The whole cpDNA dataset, including the outgroups and the diploid progenitor species, resulted in a sequence alignment of 795 bp (the *rpl32-trnL* IGS (519 bp), the *rpl32* gene (195 bp) and the *ndhF-rpl32* IGS (81 bp)), with a total of 20 potentially parsimony informative characters (PPIc) among 24 variable characters (16 substitutions, one inversion and seven indels). The *rpl32-trnL* IGS, the *rpl32* gene and the ndhF-rpl32 IGS showed 10, six and four PPIc, respectively (out of 11, seven and six variable characters, respectively). Considering the 115 *Ae. geniculata* samples only, six PPIc out of seven variable characters (four substitutions, one inversion and two indels) were left. As a whole, the cpDNA sequences allowed 17 distinct haplotypes to be distinguished, with nine haplotypes restricted to *Ae. geniculata*.

The statistical parsimony network (Fig. 2b) distinguished two major clades (hereafter referred to as N and S haplotypes), discriminated by a 6 bp inversion located in the *rpl32-trnL* IGS. Accordingly, S haplotypes occurred in 44 *Ae. geniculata* individuals (i.e. 36% of the sampling) with four distinct S haplotypes (i.e. S1–S4), whereas N haplotypes accounted for 76 individuals (64% of the sampling) and five haplotypes were detected (N1–N5). The position of the most external outgroup individuals (*Triticum* sp.) indicated that N haplotypes were derived from S haplotypes. The diploid progenitor species *Ae. comosa* (M genome) presented only S haplotypes. By contrast, the other progenitor *Ae. umbellulata* (U genome) showed both S and N haplotypes. Finally, closely related *Aegilops* species (i.e. *Ae. neglecta*, *Ae. biuncialis* and *Ae. triuncialis*) had N and S haplotypes, respectively.

Considering only *Ae. geniculata*, the statistical parsimony network presented a star-like topology, with rare and geographically restricted variants (N2–N5 and S2–S4) radiating from the most commonly occurring haplotypes (N1 and S1, respectively). The BST area showed the highest regional diversity, with five different haplotypes (Fig. 2b). Although the N and S haplotypes were equally frequent in the BST area, the N haplotypes occurred significantly more

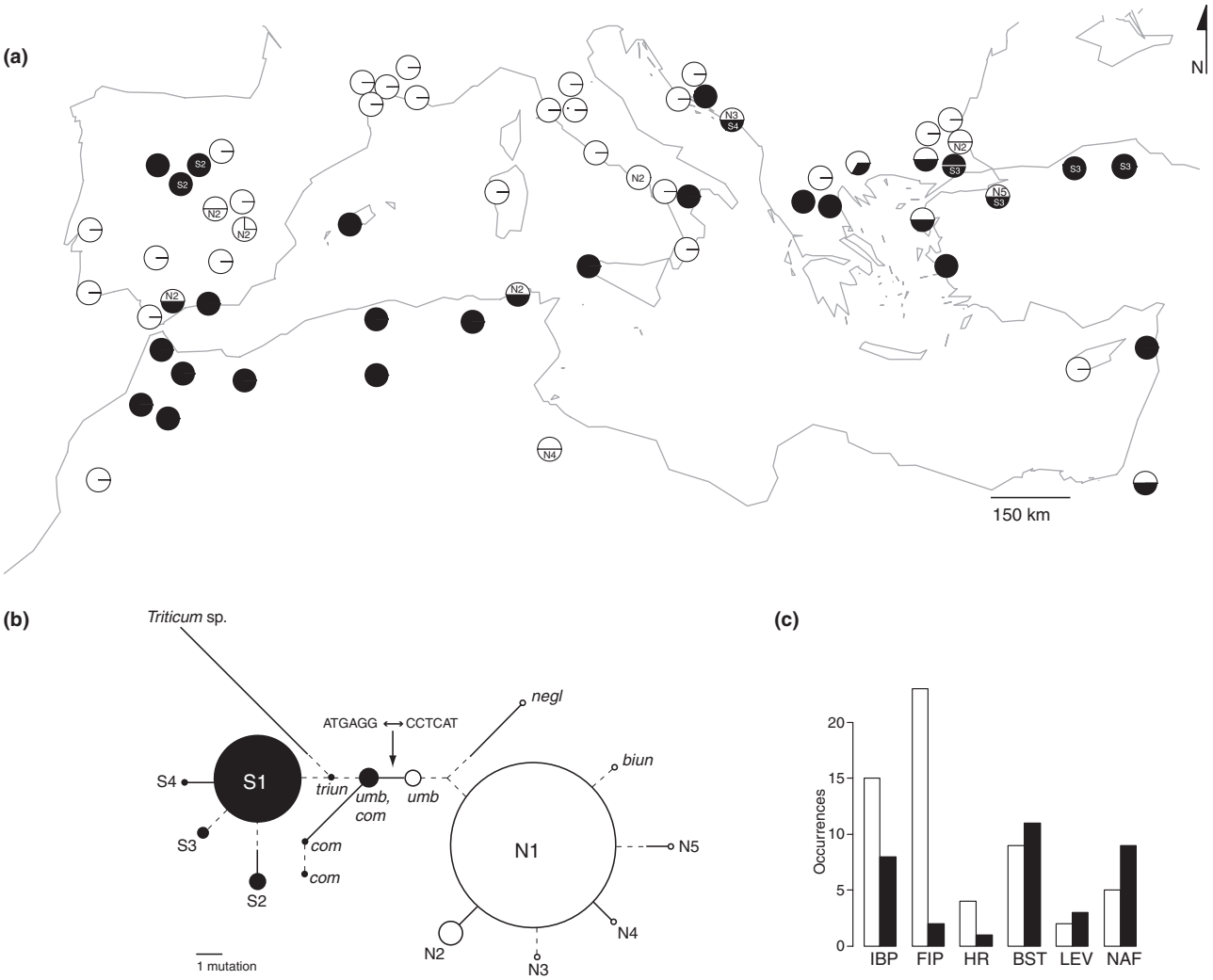

Fig. 2 Distribution of chloroplast DNA (cpDNA) haplotypes in *Aegilops geniculata* across the Mediterranean area. (a) Proportion of haplotypes in populations. The two most frequent haplotypes are displayed in black (S1) and white (N1). Locally derived haplotypes (S2–S4 and N2–N5) are indicated. (b) Statistical parsimony network, with haplotypes coded as in (a) (S, southern haplotypes; N, northern haplotypes). Outgroup haplotypes are coded as acronyms (biun, *Ae. biuncialis*; com, *Ae. comosa*; negl, *Ae. neglecta*; triun, *Ae. triuncialis*; umb, *Ae. umbellulata*; *Triticum* sp., *Triticum durum* and *T. aestivum*). Substitutions and inversions are represented using full lines on the network; indels are displayed using broken lines. The circle size is proportional to the number of individuals for each haplotype. (c) Census of haplotypes belonging to either the N (open bars) or S (closed bars) lineage in the different geographical areas (see Fig. 1).

frequently than S haplotypes in populations from the IBP, FIP and HR areas, and the opposite pattern was observed for populations located in the LEV and NAF areas (Fig. 2a,c). However, notable exceptions to this clear-cut pattern were apparent and suggested long-distance migration events. The S1 haplotype (i.e. prevailing in the NAF and LEV areas) was observed in some populations north of the Gibraltar Strait (i.e. in the south and centre of the IBP area), on Mediterranean islands (i.e. Mallorca and Sicily) as well as southern Italy (i.e. south of the FIP area). In the centre of the IBP area, both S1 and its closely related S2 haplotypes showed a restricted distribution in three nearby populations surrounded by populations harbouring the common N haplotypes. Correspondingly, the N1–N5 haplotypes were scattered in the southern Mediterranean range (i.e. BST, LEV and NAF areas), outside of their commonly reported range (i.e. IBP, FIP and HR areas).

AFLP and patterns of genetic diversity

The AFLP analysis produced a total of 649 bands (352 and 297 bands for *Eco*RI-ACT/*Mse*I-CTG and *Eco*RI-AGT/*Mse*I-CAT, respectively), with an average of 170 bands per individual. The measured error rate was 5.8%. No significant correlation was measured between the size and frequency of AFLP bands ($P = 0.69$), excluding a major influence of

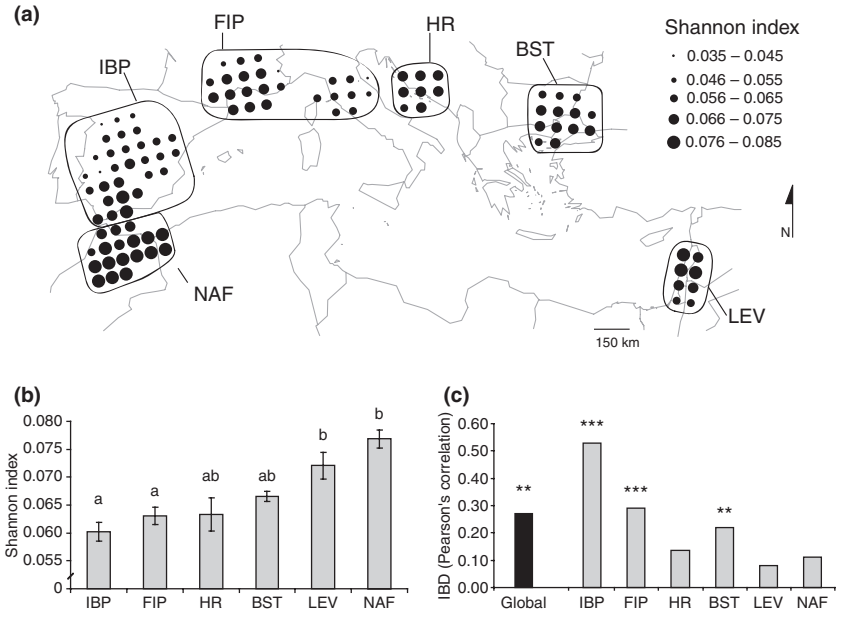

Fig. 3 Patterns of genetic diversity within *Aegilops geniculata* in the different Mediterranean areas: IBP, Iberian Peninsula; FIP, Franco-Italian Peninsula; HR, the Balkans; BST, the Bosphorous Strait; LEV, the Leventine; NAF, Northern Africa. (a) Local genetic diversity (Shannon index) measured using a sliding window on a 100 km grid. (b) Regional genetic diversity, measured as the averaged local genetic diversities, for each area. Error bars represent the standard deviation. Different letters indicate areas that are significantly different at $\alpha = 0.05$ (*post hoc* Tukey honest significant differences test). (c) Mantel tests between the inter-individual genetic and geographical distances (1000 permutations; ***significant at $\alpha = 0.001$; **significant at $\alpha = 0.01$). IBD, isolation by distance.

homoplasy on the results (as suggested in Vekemans *et al.*, 2002).

A lower genetic diversity was observed in northern Mediterranean (i.e. IBP, FIP and HR) relative to southern Mediterranean (i.e. LEV and NAF) and BST areas (Fig. 3a,b). Indeed, the IBP, FIP, HR and BST areas presented an averaged Shannon index of 0.060–0.067, whereas values of 0.072 and 0.078 were reported in LEV and NAF areas. The genetic diversity was apparently not evenly distributed within geographical areas, with an elevated Shannon index in areas in which N and S haplotypes co-occurred, around the Gibraltar and the Bosphorus Straits (i.e. between the IBP and NAF areas and in the BST area).

Significant isolation by distance patterns, as revealed by correlations between genetic and geographical distances, were observed over the whole sampling area ($r = 0.28$, $P < 0.01$) and in most of the northern Mediterranean areas (IBP = 0.53, $P < 0.001$; FIP = 0.29, $P < 0.001$; but not HR = 0.12, $P > 0.05$), as well as the BST area (BST = 0.22 $P < 0.01$; Fig. 3c). By contrast, the correlation was not significant in the southern Mediterranean areas (LEV = 0.08, $P > 0.05$ and NAF = 0.11, $P > 0.05$).

AFLP genetic structure

According to nonmodel hierarchical *K*-means clustering, the most likely genetic structure among the 411 individuals characterized with AFLP was obtained when seven groups were considered (designated as *A*–*G*; intergroup inertia = 13.5% of the total inertia; Notes S1). Congruent results were obtained using Bayesian clustering, as implemented in STRUCTURE version 2.2 (data not shown; Falush *et al.*, 2007). These groups were unevenly frequent and presented dissimilar distributions across the sampling area (Fig. 4). Individuals were frequently assigned to groups *A* (19%), *C* (21%) and *G* (34%), whereas the other groups (*B*, *D*, *E* and *F*) accounted for only 26% of the total.

Groups *A* and *B* occurred almost exclusively in the IBP and FIP areas, *A* being largely spread over Spain and Portugal, with inclusions in France and Italy, and *B* restricted to two neighbouring populations in central Spain. Groups *C*, *D* and *E* were mainly confined to the FIP area. As an exception, a few individuals from the widespread group *C* were observed in IBP and BST areas. Individuals from group *D* were restricted to a single population in southern France, and group *E* was represented in three populations from northern Italy. Individuals from the rare group *F* showed a fuzzy geographical structure, with scattered populations in various distant areas (IBP, FIP, HR and NAF). Individuals from group *G* were clearly prevalent in the BST, NAF and LEV areas, but were also observed at lower frequencies in all geographical areas. Indeed, 24 individuals from group *G* (hereafter designated as G_{Im}) were reported in the HR area, the south of the FIP area, in the IBP area and on most Mediterranean islands.

Principal coordinates analysis on AFLP profiles (Fig. 5) was largely congruent with nonhierarchical *K*-means clustering. Indeed, the two first axes clearly distinguished individuals sampled in the IBP area (i.e. groups *A* and *B* that appear to be closely related), in the FIP area (i.e. groups *C*, *D* and *E*) and individuals collected in the eastern and southern parts of the Mediterranean area (i.e. BST, LEV and NAF, corresponding to group *G*). Individuals of group *F* were intermediate between these three main clusters.

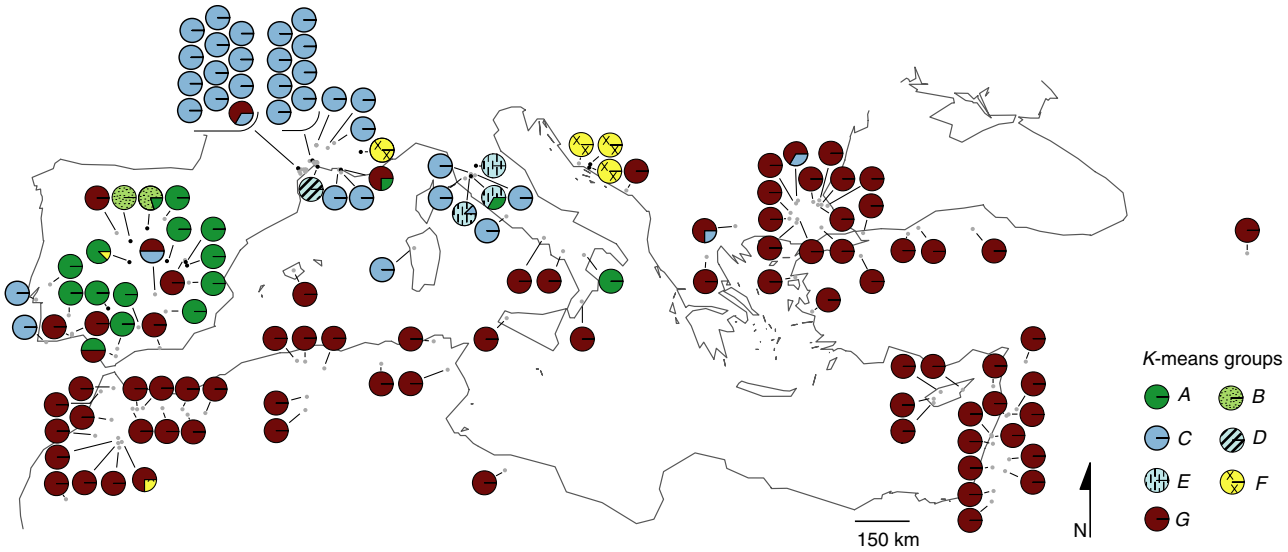

Fig. 4 Geographical origin of the 143 analysed populations of *Aegilops geniculata* and their grouping according to nonhierarchical *K*-means clustering on the amplified fragment length polymorphism (AFLP) dataset. Pie charts represent the proportion of individuals belonging to each of the seven detected groups (*A*–*G*).

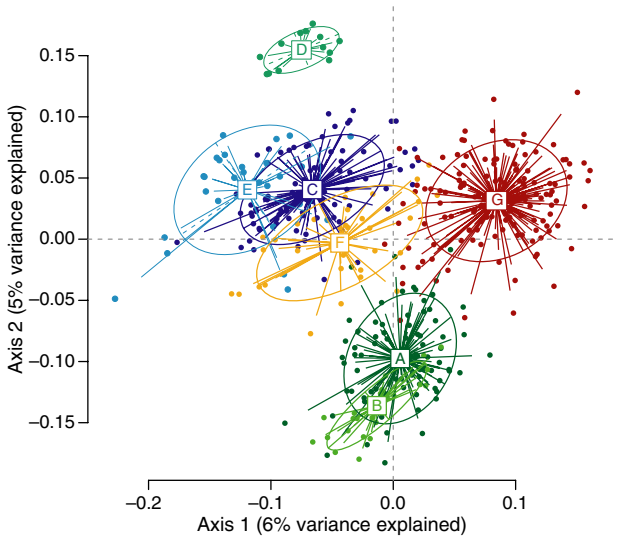

Fig. 5 Principal coordinates analysis based on the Jaccard distance between individual amplified fragment length polymorphism (AFLP) profiles of *Aegilops geniculata*. Individuals are labelled according to AFLP groups detected by nonhierarchical *K*-means clustering.

Table 1 Contingency table comparing the phylogeographical signals given by chloroplast DNA (cpDNA) and amplified fragment length polymorphism (AFLP)

	N	S	χ^2	*P*
A	14	1	11.27	< 0.001
B	0	4	4	0.046
C	16	0	16	< 0.001
E	4	0	4	0.046
F	3	1	1	0.317
G	22	29	0.96	0.327
G_{Im}	7	6	0.08	0.781
G_{Smed}	15	23	1.68	0.194

The number of cpDNA haplotypes belonging to either the N or S clade is compared with the corresponding *K*-means groups inferred from AFLP data (*A*–*G*). More detailed statistics are provided for the *G* group because individuals occurred out of its native area (G_{Im}) and therefore were considered separately from the other individuals (G_{Smed}). Associations were tested using chi-squared (χ^2) tests.

Comparison of cpDNA and AFLP phylogeographical signals

As assessed using a contingency table comparing the occurrence of cpDNA haplotypes and AFLP groups defined with *K* means, phylogeographical signals contained in both datasets were globally congruent (Table 1, $\chi^2 = 33.27$, $P < 0.001$). Individuals collected in the IBP and FIP areas (i.e. northern Mediterranean) were mostly assigned to groups *A* and *C*, and were mostly associated with N haplotypes. By contrast, individuals from group *G* were mainly collected in the BST, LEV and NAF areas and presented comparable frequencies of N and S haplotypes. Noticeably, individuals from the *K*-means group *B*, which were geographically restricted to the centre of the IBP area and genetically similar to those from group *A*, were associated with a specific S haplotype (S2). In addition, when splitting group *G* into samples collected in southern Mediterranean areas (designated as G_{Smed}) vs samples located in northern Mediterranean areas (G_{Im}), G_{Smed} tended to be associated with S haplotypes, whereas G_{Im} showed a larger proportion of N haplotypes (Table 1). Finally, $K = 3$ appeared as an alternative probable

subordinate grouping, revealing broader phylogeographical structures by distinguishing an Iberian, Franco-Italian and southern Mediterranean lineage (see Notes S1).

Discussion

Origin of *Ae. geniculata*

Genetic variation in *Ae. geniculata* revealed a clear differentiation between populations from the northern and southern Mediterranean for both cpDNA and AFLP. Only the BST area presented high cpDNA haplotype diversity and comparable frequencies for N and S haplotypes, thus possibly corresponding to the centre of origin of the allotetraploid *Ae. geniculata*. The current distribution of diploid progenitor species already suggested an eastern Mediterranean origin for *Ae. geniculata* (Chennaveeraiah, 1960; Waines & Barnhart, 1992; Van Slageren, 1994; Resta *et al.*, 1996; Zaharieva *et al.*, 2001a). In addition, we observed N and S haplotypes in *Ae. umbellulata* accessions originating from Turkey. These results indicated that *Ae. geniculata* included ancestral variation from different parental lineages and therefore might have originated multiple times. Although gene flow between *Ae. geniculata* and its parental species near the origin centre cannot be formally excluded, multiple origin of polyploid lineages has already been suggested for this species (Sasanuma *et al.*, 2006; Meimberg *et al.*, 2009). Recurrent polyploidy has been attested for a variety of *Aegilops* species, such as *Ae. biuncialis*, *Ae. neglecta*, *Ae. cylindrica* and *Ae. columnaris* (Chennaveeraiah, 1960; Murai & Tsunewaki, 1986; Chee *et al.*, 1995; Vanichanon *et al.*, 2003; Caldwell *et al.*, 2004; Kadosumi *et al.*, 2005), and is considered to be likely for *Ae. geniculata*.

Expansion of *Ae. geniculata*

The N and S haplotypes of *Ae. geniculata* were largely corroborated by AFLP, suggesting that two largely independent lineages expanded from the BST area, circumventing the Mediterranean Sea through either the north or south.

AFLP in northern Mediterranean areas revealed low regional genetic diversity, but strong biogeographical structure, with significant isolation by distance and six of the seven *K*-means groups. It seems unlikely that the spread of agriculture across Europe could have promoted the early expansion of *Ae. geniculata* through this expansion route. Agriculture was indeed established *ca.* 7500 yr ago in the western Mediterranean (Wainwright & Thornes, 2004; Zeder, 2008) and human activities probably homogenized the genetic diversity of pastoral species by promoting long-distance dispersal (e.g. through biannual transhumances covering between 100 and 700 km; Blondel, 2006). In contrast with our data, neither isolation by distance nor strong genetic structure is expected under human-driven expansion. Moreover, because agriculture spread in Spain and Italy contemporaneously from south Italian settlements (Wainwright & Thornes, 2004; Zeder, 2008), no genetic differentiation should be observed between these two peninsulas. It thus seems more likely that the expansion of *Ae. geniculata* across the northern Mediterranean occurred before the last glacial maximum, which is congruent with available evidence suggesting an ancient origin of this polyploidy (Furuta, 1981; Terachi *et al.*, 1984; Wang *et al.*, 2000; Badaeva *et al.*, 2004).

Consistent with the natural expansion of *Ae geniculata* following climatic oscillations, the low genetic diversity observed across northern areas would have resulted from bottlenecks and founder effects during glacial survival in refugia and postglacial range expansion. Correspondingly, the significant isolation by distance would have been caused by progressive postglacial range expansion and/or regional equilibrium between migration and drift. In either case, patterns of genetic variation suggest a long-standing occupancy of the landscape. According to this scenario, gene pools from the IBP and FIP areas would have differentiated in allopatric refugia (e.g. in the south of the Iberian Peninsula and in Italy) and maintained their integrity thanks to the Pyrenees, currently representing a suture zone. In support of this hypothesis, the same potential refugia and recolonization routes have already been outlined for other plant and animal species (Hewitt, 2000, 2004; Médail & Diadema, 2009).

Genetic homogeneity across the southern Mediterranean area suggests limited fragmentation of the species' range during the last glacial maximum. In marked contrast with the northern pattern, we observed high levels of genetic diversity, a low biogeographical structure and no clear isolation by distance across the southern expansion route. This pattern suggests intensive postglacial migration within this area, possibly in relation to dramatic climatic oscillations that affected the North African coast during the last 10 000 yr (Kuper & Kroepelin, 2006). In addition, long-distance dispersal promoted by recent human activities, such as transhumance (Ballouche, 2003), possibly smoothed initial biogeographical patterns among *Ae. geniculata* populations. Given the low differentiation observed across the BST, LEV and NAF areas, it remains difficult to firmly assess the relative influence of climate and humans on the spatiotemporal dynamics of *Ae. geniculata* in the southern Mediterranean areas.

Human-driven migration events and introgression

As *Ae. geniculata* is frequent in pastures and field borders, human activities have probably contributed to its recent dispersal. Accordingly, our survey highlighted restricted populations with specific genotypes occurring in areas

distant from their inferred centre of origin. Most migration events occurred northwards (i.e. G_{Im} individuals), from southern Mediterranean areas to northern coasts, and broadly corresponded to human trade routes (e.g. Phoenician trade centre in the Mediterranean islands, Arabic invasions in southern Spain; Moscati, 2001).

Recent migrants of *Ae. geniculata* apparently had an important impact on the evolution of resident gene pools. Our results indeed outlined several admixture events involving migrants and native populations. For instance, individuals from two populations in central Spain (group *B*) probably had an ancient southern Mediterranean origin (as attested by their S2 haplotype), followed by admixture with the resident genotypes from the northern Mediterranean area (as attested by their relatedness with group *A*). In addition, numerous introgression events involved individuals of group *G* occurring in northern areas (i.e. G_{Im}, Table 1), highlighting a contrasting situation. In this case, AFLP indicated a southern Mediterranean origin for the nuclear genome, whereas cpDNA pointed to the resident northern lineage. Such introgression was apparently frequent, as the proportion of N haplotypes in the G_{Im} samples was larger than that observed in the rest of the *G* group (Table 1). In particular, as S haplotypes were restricted to a few populations, whereas the corresponding nuclear genome (*G*) was much more widespread, migrant nuclear genomes apparently spread more efficiently than migrant chloroplasts in the IBP area. The processes driving such a preferential introgression are unclear (Currat *et al.*, 2008) and whether this particular genome is selectively advantageous remains an open question.

Evolutionary history of *Ae. geniculata* and human interferences

In this study, we have outlined the joint influence of postglacial recolonization and recent human activities on the genetic diversity of species associated with agroecosystems. In particular, we found phylogeographical patterns in *Ae. geniculata* corresponding to glacial refugia and recolonization routes reported for wild species (Hewitt, 2004; Jakob *et al.*, 2007; Schmitt, 2007; Médail & Diadema, 2009). We also reported numerous trans-Mediterranean exchanges, corroborating the phylogeographical relationships between North Africa and southern Europe evidenced in the wild flora and fauna (e.g. Costich & Meagher, 1992; Comes & Abbott, 2001; Beja-Pereira *et al.*, 2006; Obbard *et al.*, 2006; Zeder *et al.*, 2006; Ortiz *et al.*, 2008; Rodriguez-Sánchez *et al.*, 2008), which all suggested long-distance and, possibly, human-mediated dispersal. Although the evolutionary trajectories of plants associated with agroecosystems appear to be similar to those of wild species, they certainly have been impacted by recent human activities, promoting long-distance dispersal opportunities (e.g. transhumance and trade; Blondel, 2006) and/or offering suitable habitats (e.g. the opening of pastures and the introduction of cultures; Wainwright & Thornes, 2004).

Acknowledgements

The authors are grateful to the team of the Botanical Garden of the University and City of Neuchâtel for their help with plant cultivation. Philippe Küpfer, Matthieu Honegger, Mélanie Thobor, Anouk Béguin, Sylvain Lappe, Celia Bueno, Peter Sellers and Anahí Espindola provided useful comments and technical assistance. This project was funded by the National Centre of Competence in Research (NCCR) Plant Survival and the PNR-59, two research programmes of the Swiss National Science Foundation.

References

Arrigo N, Tuszynski JW, Ehrich D, Gerdes T, Alvarez N. 2009. Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BMC Bioinformatics* **10**: 33.

Badaeva ED, Amosova AV, Samatadze TE, Zoshchuck SA, Shostak NG, Chikida NN, Zelenin AV, Raupp WJ, Frijters A, Gill BS. 2004. Genome differentiation in *Aegilops*. iv. Evolution of the U-genome cluster. *Plant Systematics and Evolution* **246**: 45–76.

Balfourier F, Imbert C, Charmet G. 2000. Evidence for phylogeographic structure in *Lolium* species related to the spread of agriculture in Europe. A cpDNA study. *Theoretical and Applied Genetics* **101**: 131–138.

Ballouche A. 2003. La dimension anthropique de l'histoire de la végétation Holocène au Maghreb. In: *Le peuplement ancien de l'Algérie: sa place dans le contexte Méditerranéen*. Paris, France: Collège de France.

Beja-Pereira A, Caramelli D, Lalueza-Fox C, Vernesi C, Ferrand N, Casoli A, Goyache F, Royo LJ, Conti S, Lari M *et al.* 2006. The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences, USA* **103**: 8113–8118.

Besnard G, De Casas RR, Vargas P. 2007. Plastid and nuclear DNA polymorphism reveals historical processes of isolation and reticulation in the olive tree complex (*Olea europaea*). *Journal of Biogeography* **34**: 736–752.

Bittkau C, Comes HP. 2005. Evolutionary processes in a continental island system: molecular phylogeography of the Aegean *Nigella arvensis* alliance (Ranunculaceae) inferred from chloroplast DNA. *Molecular Ecology* **14**: 4065–4083.

Blondel J. 2006. The 'Design' of Mediterranean landscapes: a millennial story of humans and ecological systems during the historic period. *Human Ecology* **34**: 713–729.

Bonin A, Bellemain E, Bronken Edeisen P, Pompanon F, Brochmann C, Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**: 3261–3273.

Borchsenius F. 2009. *FastGap 1.2*. Denmark: Department of Biological Sciences, University of Aarhus, Aarhus, Denmark. http://192.38.46.42/aubot/fb/FastGap_home.htm.

Burnier J, Buerki S, Arrigo N, Küpfer P, Alvarez N. 2009. Genetic structure and evolution of Alpine polyploid complexes: *Ranunculus kuepferi* (Ranunculaceae) as a case study. *Molecular Ecology* **18**: 3730–3744.

Bussel JD. 1999. The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotama petraea* (Lobeliaceae). *Molecular Ecology* **8**: 775–789.

Caldwell KH, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W. 2004. Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. *Genetics* **167**: 941–947.

Chee PW, Lavin M, Talbert LE. 1995. Molecular analysis of evolutionary patterns in U genome wild wheats. *Genome* **38**: 290–297.

Chen DH, Ronald PC. 1999. A rapid DNA minipreparation method suitable for AFLP and other PCR applications. *Plant Molecular Biology Reporter* **17**: 53–57.

Chennaveeraiah MS. 1960. Some aspects of the cytotaxonomy of *Aegilops*. *Proceedings Summer School of Botany* 43–50.

Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**: 1657–1660.

Comes HP, Abbott RJ. 2001. Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). *Evolution* **55**: 1943–1962.

Costich DE, Meagher TR. 1992. Genetic variation in *Ecballium elaterium* (Cucurbitaceae): breeding system and geographic distribution. *Journal of Evolutionary Biology* **5**: 589–601.

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. *Evolution* **62**: 1908–1920.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Falush D, Stephens M, Pritchard JK. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* **7**: 574–578.

Fineschi S, Taurchini D, Villani F, Vendramin GG. 2000. Chloroplast DNA polymorphism reveals little geographical structure in *Castanea sativa* Mill. (Fagaceae) throughout Southern European countries. *Molecular Ecology* **9**: 1495–1503.

Fjellheim S, Rognli OA, Fosnes K, Brochmann C. 2006. Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. *Journal of Biogeography* **33**: 1470–1478.

Friebe BR, Tuleen NA, Gill BS. 1999. Development and identification of a complete set of *Triticum aestivum–Aegilops geniculata* chromosome addition lines. *Genome* **42**: 374–380.

Furuta Y. 1981. Chromosome structural variation in *Aegilops ovata* L. *Japanese Journal of Genetics* **56**: 287–294.

Gugerli F, Englisch T, Niklfeld H, Tribsch A, Mirek Z, Ronikier M, Zimmermann NE, Holderegger R, Taberlet P. 2008. Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation – a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics* **10**: 259–281.

Hall TA. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95–98.

Hammer K. 1980. Vorarbeiten zur monographischen Darstellung von Wildpflanzensortimenten: *Aegilops* L. *Kulturpflanze* **28**: 33–180.

Hampe A, Petit RJ. 2005. Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters* **8**: 461–467.

Hartigan JA, Wong MA. 1979. A K-means clustering algorithm. *Applied Statistics* **28**: 100–108.

Hewitt GM. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**: 907–913.

Hewitt GM. 2004. Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**: 183–195.

Jakob SS, Ihlow A, Blattner FR. 2007. Combined ecological niche modelling and molecular phylogeography revealed the evolutionary history of *Hordeum marinum* (Poaceae) – niche differentiation, loss of genetic diversity, and speciation in Mediterranean Quaternary refugia. *Molecular Ecology* **16**: 1713–1727.

Kadosumi S, Kawahara T, Sasanuma T. 2005. Multiple origins of U genome in two UM genome tetraploid *Aegilops* species, *Ae. columnaris* and *Ae. triaristata*, revealed based on the polymorphism of a genome-specific PCR fragment. *Genes & Genetic Systems* **80**: 105–111.

Kluge AG. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Systematic Zoology* **38**: 7–25.

Kuper R, Kroepelin S. 2006. Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution. *Science* **313**: 803–807.

Legendre P, Legendre L. 1998. *Numerical ecology*. Amsterdam, the Netherlands: Elsevier.

Médail F, Diadema K. 2009. Glacial refugia influence plant diversity patterns in the Mediterranean basin. *Journal of Biogeography* **36**: 1333–1345.

Meimberg H, Rice KJ, Milan NF, Njoku CC, Mckay JK. 2009. Multiple origins promote the ecological amplitude of allopolyploid *Aegilops* (Poaceae). *American Journal of Botany* **96**: 1262–1273.

Moscati S. 2001. *The Phoenicians*. London, UK & New York, NY, USA: I.B. Tauris & Co Ltd.

Murai K, Tsunewaki K. 1986. Molecular-basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops* species. 4. ctDNA variation in *Ae. triuncialis*. *Heredity* **57**: 335–339.

Obbard DJ, Harris SA, Pannell JR. 2006. Sexual systems and population genetic structure in an annual plant: testing the metapopulation model. *The American Naturalist* **167**: 354–366.

Ortiz MA, Tremetsberger K, Terrab A, Stuessy TF, Garcia-Castano JL, Urtubey E, Baeza CM, Ruas CF, Gibbs PE, Talavera S. 2008. Phylogeography of the invasive weed *Hypochaeris radicata* (Asteraceae): from Moroccan origin to worldwide introduced populations. *Molecular Ecology* **17**: 3654–3667.

Parducci L, Szmidt AE, Madaghiele A, Anzidei M, Vendramin GG. 2001. Genetic variation at chloroplast microsatellites (cpSSRs) in *Abies nebrodensis* (Lojac.) Mattei and three neighboring *Abies* species. *Theoretical and Applied Genetics* **102**: 733–740.

Presti RM, Oberprieler C. 2009. Evolutionary history, biogeography and eco-climatological differentiation of the genus *Anthemis* L. (Compositae, Anthemideae) in the circum-Mediterranean area. *Journal of Biogeography* **36**: 1313–1332.

R Development Core Team. 2009. R: a language and environment for statistical computing, reference index version 2.8.1. Vienna, Austria: R Foundation for Statistical Computing.

Resta P, Zhang HB, Dubcovsky J, Dvorak J. 1996. The origins of the genomes of *Triticum biunciale*, *T. ovatum*, *T. neglectum*, *T. columnare*, and *T. rectum* (Poaceae) based on variation in repeated nucleotide sequences. *American Journal of Botany* **83**: 1556–1565.

Rodriguez-Sánchez F, Perez-Barrales R, Ojeda F, Vargas P, Arroyo J. 2008. The Strait of Gibraltar as a melting pot for plant biodiversity. *Quaternary Science Reviews* **27**: 2100–2117.

Sasanuma T, Kadosumi S, Kawahara T. 2006. Multiple origins of U genome in two UM genome tetraploid *Aegilops* species, *Ae. biuncialis* and *Ae. ovata*. *Wheat Info Service* **101**: 11–12.

Schmitt T. 2007. Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology* **4**, doi:10.1186/1742-9994-4-11.

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu WS, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* **92**: 142–166.

Shaw J, Lickey EB, Schilling EE, Small RL. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**: 275–288.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**:

Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.

Terachi T, Ogihara Y, Tsunewaki K. 1984. The molecular-basis of genetic diversity among cytoplasms of *Triticum* and *Aegilops*. 3. Chloroplast genomes of the M and modified-M genome-carrying species. *Genetics* **108**: 681–695.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**: 4876–4882.

Van Slageren MW. 1994. *Wild wheats: a monograph of* Aegilops *L. and* Ambylopyrum *(Jaub and Spach) Eig (Poaceae)*. Wageningen, the Netherlands: Wageningen Agricultural University Papers, 94–97.

Vanichanon A, Blake NK, Sherman JD, Talbert LE. 2003. Multiple origins of allopolyploid *Aegilops triuncialis*. *Theoretical and Applied Genetics* **106**: 804–810.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* **11**: 139–151.

Waines JG, Barnhart D. 1992. Biosystematic research in *Aegilops* and *Triticum*. *Hereditas* **116**: 207–212.

Wainwright J, Thornes JB. 2004. *Environmental issues in the Mediterranean: processes and perspectives from the past and present*. London, UK: Routledge.

Wang GZ, Matsuoka Y, Tsunewaki K. 2000. Evolutionary features of chondriome divergence in *Triticum* (wheat) and *Aegilops* shown by RFLP analysis of mitochondrial DNAs. *Theoretical and Applied Genetics* **100**: 221–231.

Wiens JJ, Donoghue MJ. 2004. Historical biogeography, ecology and species richness. *Trends in Ecology & Evolution* **19**: 639–644.

Zaharieva M, Dimov A, Stankova P, David J, Monneveux P. 2003. Morphological diversity and potential interest for wheat improvement of three *Aegilops* L. species from Bulgaria. *Genetic Resources and Crop Evolution* **50**: 507–517.

Zaharieva M, Gaulin E, Havaux M, Acevedo E, Monneveux P. 2001a. Drought and heat responses in the wild wheat relative *Aegilops geniculata* Roth: potential interest for wheat improvement. *Crop Science* **41**: 1321–1329.

Zaharieva M, Santoni S, David J. 2001b. Use of RFLP markers to study genetic diversity and to build a core-collection of the wild wheat relative *Ae. geniculata* Roth (= *Ae. ovata* L.). *Genetics Selection Evolution* **33**: S269–S288.

Zeder M. 2008. Domestication and early agriculture in the Mediterranean basin: origins, diffusion and impact. *Proceedings of the National Academy of Sciences, USA* **105**: 11597–11604.

Zeder M, Emshwiller E, Smith BD, Bradley D. 2006. Documenting domestication: the intersection of genetics and archaeology. *Trends in Genetics* **22**: 139–155.

Supporting Information

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

Table S1 Detailed plant material investigated in the present study

Notes S1 Screening for the most likely number of groups in nonhierarchical *K*-means clustering.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.