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

Nils Arrigo¹, François Felber¹, Christian Parisod¹, Sven Buerki¹, Nadir Alvarez², Jacques David³ and Roberto Guadagnuolo¹

¹Laboratory of Evolutionary Botany, Institute of Biology, University of Neuchâtel, 11 rue Emile-Argand, 2009 Neuchâtel, Switzerland; ²Department of Ecology and Evolution, University of Lausanne – UNIL Sorge, Biophore Building, 1015 Lausanne, Switzerland; ³UMR 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


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,

*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 zoohorous 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 colonized 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.
cpDNA sequences

Three contiguous loci of the small single-copy region of cpDNA were sequenced: the \textit{ndhF-rpl32} intergenic spacer (IGS), the \textit{rpl32} gene and the \textit{rpl32-trnL} IGS. The whole region was amplified with the \textit{trnL(UAG)} and \textit{ndhF} primers (Shaw \textit{et al.}, 2005, 2007) and sequenced using \textit{Aegilops}-specific primers (\textit{internF}, 5\textquoteleft\textprime-CCAATTCTTATCTCTTTCTGAAAG-3\textquoteleft; \textit{internR}, 5\textquoteleft\textprime-GCTTTGCCCAATAGAACACA-3\textquoteleft). PCR amplifications were performed on 30 ng of genomic DNA in a 40 l l solution (1 \cdot PCR buffer, 0.2 mg ml\textsuperscript{-1} BSA, 1.5 mM MgCl\textsubscript{2}, 200 mM deoxynucleoside triphosphates (dNTPs), 0.2 mM primers and 2 U Taq polymerase (Promega)) using the \textquoteleft\textit{rpl16}\textquoteright PCR cycle proposed by Shaw \textit{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 \textit{Ae. comosa} and \textit{Ae. umbellulata} from GenBank (Meimberg \textit{et al.}, 2009), were aligned using ClustalX (Thompson \textit{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 (\textit{sensu} Kluge, 1989). We performed a statistical parsimony network (Templeton \textit{et al.}, 1992) with TCS version 1.21 (Clement \textit{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 \textit{et al.} (2008). Two selective PCR primer pairs were used (\textit{EcoRI-CTG} and \textit{MseI-CTG}) with FAM-labelled \textit{EcoRI} 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 \textit{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 \textit{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 \textit{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 (Burniet 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 (χ²) 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
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 EcoRI-ACT/ Mse-CTG and EcoRI-AGT/ Mse-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 ($r = 0.69$), excluding a major influence of
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 $G_{mn}$) 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.
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 Gsmed) vs samples located in northern Mediterranean areas (Gim), Gsmed tended to be associated with S haplotypes, whereas Gim showed a larger proportion of N haplotypes (Table 1). Finally, $K = 3$ appeared as an alternative probable

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>S</th>
<th>$\chi^2$</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>14</td>
<td>1</td>
<td>11.27</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>0.046</td>
</tr>
<tr>
<td>C</td>
<td>16</td>
<td>0</td>
<td>16</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>E</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>0.046</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0.317</td>
</tr>
<tr>
<td>G</td>
<td>22</td>
<td>29</td>
<td>0.96</td>
<td>0.327</td>
</tr>
<tr>
<td>Gim</td>
<td>7</td>
<td>6</td>
<td>0.08</td>
<td>0.781</td>
</tr>
<tr>
<td>Gsmed</td>
<td>15</td>
<td>23</td>
<td>1.68</td>
<td>0.194</td>
</tr>
</tbody>
</table>

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 (Gim) and therefore were considered separately from the other individuals (Gsmed). Associations were tested using chi-squared ($\chi^2$) tests.
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 & Tsubanekawa, 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 (Furuita, 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 allopatic 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; Médaïl & 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
distances from their inferred centre of origin. Most migration events occurred northwards (i.e. \( G_{in} \) 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. \( G_{in} \), 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 \( G_{in} \) 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 post-glacial 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 Küpper, Matthieu Honegger, Mélanie Thobor, Anouk Béguin, Sylvain Lappe, Celia Bueno, Peter Sellers and Anaísp 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**


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.