Original article

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.)

Maria ZAHARIEVA[a], Sylvain SANTONI[a],
Jacques DAVID[b*]

[a]UMR Diversité et génome des plantes cultivées, Ensam-Inra,
34060 Montpellier Cedex, France
[b] UMR Diversité et génome des plantes cultivées, Institut national de la recherche agronomique, Domaine de Melgueil, 34130 Mauguio, France

Abstract – The wild wheat relative *Ae. geniculata* Roth (= *Ae. ovata* L.) is a preferentially selfing, allo-tetraploid species (MU genome) widely distributed around the Mediterranean Basin. Two hundred and two individuals belonging to one hundred and fifty one populations originating from different eco-geographical regions were used to study its genetic diversity. The diversity was revealed on Restriction Fragment Length Polymorphisms (RFLP). Nine clones mapped on the *Triticeae* group 2 chromosome consensus map were used as probes. The genetic variability was found to be highly structured. A fine scale study performed on populations from Bulgaria, France and Morocco suggested that 52% of the diversity was explained by differences among populations within regions. A Factorial Analysis of Correspondence (FAC) distinguished two main structures corresponding mainly to the North and South of the Mediterranean Sea. Differences among close regions separated by natural barriers were also found. Hypotheses concerning the dissemination of *Ae. geniculata* around the Mediterranean Sea are presented and examined. In order to improve the management of *Ae. geniculata* genetic resources, a methodology for core collection sampling is presented and the efficiency of RFLP markers in building this collection is discussed.

Aegilops geniculata / core collection / genetic diversity / genetic resources

Résumé – **Utilisation de marqueurs RFLP pour l'étude de la diversité génétique et l'établissement d'une *core collection* d'*Ae. geniculata* Roth (= *Ae. ovata* L.).** *Ae. geniculata* Roth (= *Ae. ovata* L.) est une espèce apparentée au blé, allotétraploïde (de génome MU), à l'autogamie prépondérante, largement distribuée sur le pourtour du Bassin Méditerranéen. Deux cent deux individus, appartenant à cent cinquante et une populations originaires de différentes régions éco-géographiques, ont été utilisés pour étudier la diversité génétique de l'espèce. L'étude a été basée sur le polymorphisme de longueur des fragments de restriction (*Restriction Fragments Length Polymorphism*, RFLP). Neuf clones, cartographiés sur

* Correspondence and reprints
E-mail: david@ensam.inra.fr

le chromosome 2 de la carte consensus des *Triticeae*, ont été utilisés comme sondes. La variabilité génétique de cette espèce s'est révélée très structurée entre populations. Une étude à petite échelle, réalisée sur des populations de Bulgarie, France et Maroc, a suggéré que 52% de la diversité s'expliquait par des différences entre populations à l'intérieur des régions. Une Analyse Factorielle des Correspondances (AFC) a permis de distinguer deux structures principales, correspondant au nord et au sud de la Méditerranée. Des différences ont également été notées entre régions proches séparées par des barrières naturelles. Des hypothèses concernant la diffusion d'*Ae. geniculata* autour de la Méditerranée sont présentées et examinées. Afin d'améliorer la gestion des ressources génétiques d'*Ae. geniculata*, une méthodologie d'établissement d'une *core-collection* est étudiée, et l'efficacité des marqueurs RFLP pour la construction de cette collection est discutée.

Aegilops geniculata / **core collection** / **diversité génétique** / **ressources génétiques**

1. INTRODUCTION

The wild species of the tribe *Triticeae*, and especially those of the *Aegilops* genus, constitute an important reservoir of genes for wheat improvement [10, 17]. Among the 22 species of the *Aegilops* genus, *Ae. geniculata* Roth (MU genome) appears particularly interesting as a source of resistance to disease and pests [6, 11, 25] and salinity [8]. It is one of the most widespread species of the genus [27]. It can be found in South Europe, North Africa and Western Asia, but is rare in central Turkey (Anatolian Plateau) and Egypt. Typically Mediterranean, this species is locally abundant in dry disturbed areas, sometimes on edges of wheat fields with which it may form natural hybrids [27]. As *Ae. geniculata* adapts to a large range of environmental constraints, it could be a potential source of alien variation for wheat improvement.

Ae. geniculata is an annual, preferentially selfing [13], allo-tetraploid species $(2n = 4x = 28)$. It is presumed to be an amphiploid of two diploid species, *Ae. umbellulata* (U genome) and *Ae. comosa* (M genome) [15]. Chloroplastic data suggest that *Ae. geniculata* is one of the oldest polyploid species of the genus [29]. The distributions of *Ae. geniculata* and its two presumed ancestor species (*Ae. umbellulata* and *Ae. comosa*) overlap in western Turkey and in Greece [15]. Different hypotheses concerning the dissemination of *Ae. geniculata* around the Mediterranean Sea can be made. A first hypothesis is that it accompanied Neolithic farmers when agriculture spread into the whole Mediterranean area, seeds or spikelets being transported with harvested cereal seeds. An alternative hypothesis assumes that it was already disseminated in the Mediterranean basin before the last glacial maximum (18 000 years ago) and that it recolonized Europe from glacial refugia as observed in the case of forest trees [23]. The situation in the south part of the Mediterranean basin is also of interest since the desert boundaries have fluctuated widely over the last 8 000 years and some areas which were connected by typical Mediterranean

species are now separated by desert [1]. If such refuge zones existed, their identification is of prime interest for *Ae. geniculata* genetic resources conservation and use. A good knowledge of the genetic diversity structure would also permit us to identify areas in which *in-situ* conservation programs could lead to a sustainable maintenance of wild genetic resources.

The valorization of these genetic resources often requires hard labor to obtain interspecific hybrids and their fertile progenies. Creating populations of domesticated wheats largely introgressed with genes coming from the wild should therefore be based on a few initial wide crosses. Thus the wild genitors used in such a crossing program should represent the largest part of the diversity of their species. Such a sample, with the smallest size and the largest diversity possible, is called a core collection [9] and can be easily stored in genebanks and used in plant breeding [4].

Since *Ae. geniculata* could be a good candidate for massive introgression into cultivated wheats, the aim of the present work was to describe the genetic diversity structure in populations originating from different eco-geographical regions of the species distribution and to study the feasibility of establishing a core collection using markers. The sampling was realized according to these two objectives. To compare the level of intra-population and inter-population diversity, populations of three countries, Bulgaria, France and Morocco, were represented by three individuals. The goal was not to achieve a precise vision of the intra-population diversity but to obtain a reasonably accurate average value. The sample was then extended to the whole Mediterranean basin to obtain a better representation of the genetic diversity of the species.

RFLP (Restriction Fragment Length Polymorphism) markers provide a reliable means of exploring the genetic variation in populations. They are codominant genetic markers allowing an estimation of the degree of relatedness between individuals without influence of environmental variation. RFLPs reveal polymorphism at mapped loci with a high specificity and were found to be a potentially powerful tool in crop improvement for monitoring introgressions [16]. As *Ae. geniculata* is an allo-tetraploid species and as collinearity is strong among Triticeae genomes, a RFLP clone from a wheat library is expected to reveal two (or more) homeologous loci. Therefore, we primarily decided to use RFLP clones mapped on only one wheat homeology group and thus to obtain two independent sets of linked loci on which we could contrast the effect of physical linkage on the structure of diversity.

After the description of the genetic structuration at the geographical level, the data were used to verify if molecular markers can be used to sample core collections by maximizing diversity at known loci, considering that they are in linkage disequilibrium with unknown loci, due to the difference in evolutionary trajectories of the studied populations.
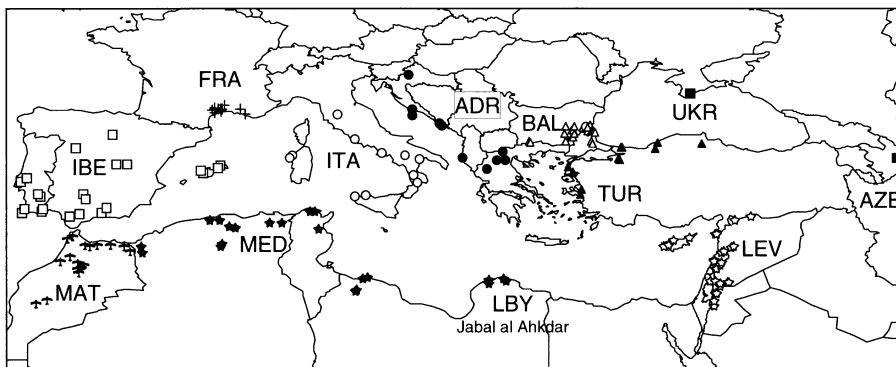
Figure 1. Geographical distribution of *Aegilops* geniculata accessions used in this study.

2. MATERIAL AND METHODS

2.1. Plant material and sampling

Two hundred and two *Aegilops geniculata* individuals belonging to 151 populations were used in this study. These populations originated from twenty countries mostly located around the Mediterranean basin: Portugal (8), Spain (14), France (13), Italy (12), Croatia (4), Albania (1), Greece (4), Bulgaria (15), Ukraine (1), Azerbaijan (1), Turkey (15), Cyprus (3), Syria (10), Lebanon (6), Jordan (6), Palestine (1), Libya (7), Tunisia (4), Algeria (9) and Morocco (17). The 45 populations from Bulgaria, Morocco and France were collected by M. Zaharieva [31], O. Benlhabib and M. Zaharieva [2] and J. David and M. Zaharieva, respectively. The other 106 populations were provided by Dr. J. Valkoun (ICARDA, Syria), Dr. K. Hammer (Zentralinstitut für Genetik und Kulturpflanzenforschung, Gatersleben, Germany) and Dr. M. Ruiz (INIA, Spain).

Populations were grouped into regions separated by eco-geographical barriers such as seas, mountains or deserts (Fig. 1). These *a priori* defined regions were the Iberian peninsula (IBE), France (FRA), Italy (ITA), countries of the east coast of the Adriatic Sea, *e.g.*, Croatia, Albania and Greece (ADR), the Balkan region including Bulgaria and the European part of Turkey (BAL), the eastern part of Turkey (TUR), Near East countries of the Levant including Syria, Cyprus, Jordan, Lebanon, Palestine (LEV), the Mediterranean Maghreb including Libya, Tunisia, Algeria, East Morocco (MED) and the Atlantic Maghreb constituted by West Moroccan populations (MAT).

2.2. RFLP analysis

Total DNA was extracted according to Tai and Tanksley [24] from frozen leaves and digested with individual restriction endonucleases (*EcoRI* and

*Hind*III). About 7.5 μg of digested DNA was separated on 0.8% agarose gel and transferred to charged Nylon membrane (Hybond N+, Amersham-Pharmacia) using alkaline transfer.

In order to take into account possible migration distortions, a set of molecular-weight markers of 24 kb (obtained from XbaI restriction of lambda DNA) and 1.5 kb (from specific priming PCR amplification on lambda DNA) were added to each DNA sample just before loading on gels. The lamba DNA digested by *Hind*III which generates restriction fragments of 2.0, 2.3, 4.3, 6.5 and 9.4 kb was used as external molecular weight standard.

Probes were radiolabeled by random priming (50 μCi of ^{32}P αdCTP (3000 Ci/mmol), (1 Ci = 37 Gbq) per probe). Membranes were hybridized overnight at 65 °C in a 7% SDS, 0.5 M NaHPO$_4$ (pH 7.2) buffer [5] and then washed three times in 40 mM NaHPO$_4$ (pH 7.2), 1% SDS (20 min at 65 °C). Membranes were then exposed to X-ray film (Hyperfilm MP, Amersham-Pharmacia) at −80 °C for seven days.

RFLP bands on autoradiography for each clone-enzyme combination were visually scored according to their approximate migration distance estimated by comparison with λ-phage markers. Only clearly visible bands were taken into consideration. The presence or absence of a band was coded as 1 or 0, respectively.

Twenty clones from oat, barley and wheat genomic libraries mapped on the *Triticeae* group 2 chromosome consensus map, provided by Dr. P. Leroy (INRA, Clermont-Ferrand), were hybridized to DNAs from three populations (Bulgaria, France and Morocco) digested with with *EcoRI* or *Hind*III restriction enzymes. Nine probes were selected on the basis of the clearness of the hybridization pattern and their ability to reveal a high level of polymorphism among the three reference populations and were then used to analyzed the genetic structure of the 151 studied populations (Tab. I).

2.3. Fst statistics

We assessed the structuration of the genetic diversity for each marker on different geographical scales for samples from Bulgaria, France and Morocco. Using the analysis of variance on the 0/1 genotypes at the marker [30], Bulgaria, France and Morocco were declared as three levels of a region factor, the population factor was nested into region, and the three individuals per population provided an estimation of the within-population variance. For each marker, the variances of the different factors (region σ_R^2, population(region) $\sigma_{P/R}^2$ and within-population residual σ_w^2 were computed by the VARCOMP procedure [21]. Fst$_R$ and Fst$_{P/R}$ were calculated as Fst$_R = \sigma_R^2/(\sigma_R^2 + \sigma_{P/R}^2 + \sigma_w^2)$ and Fst$_{P/R} = \sigma_{P/R}^2/(\sigma_R^2 + \sigma_{P/R}^2 + \sigma_w^2)$ [30]. The relative part of the within-population variation was $\sigma_w^2/(\sigma_R^2 + \sigma_{P/R}^2 + \sigma_w^2)$.

Table I. Characteristics of the nine clones used in RFLP analyses and the polymorphism revealed with *EcoRI* and *HindIII* restriction enzymes.

Probe	Restriction enzyme	Map position (cM)	Number of polymorphic bands	Allele size (kb)	Number of patterns	Number of polymorphic bands/pattern
bcd 348	*HindIII*	24.7 2S	13	1.6-14.5	21	6
bcd 1184	*HindIII*	42.7 2S	11	3.1-13.5	15	3
cdo 405	*EcoRI*	111.7 2S	25	4-14.5	69	6
fba 345	*EcoRI*	*2BL	16	2.5-12.5	51	5
bcd 266	*EcoRI*	154.8 2L	1	6.22	2	1
cdo 678	*EcoRI*	177.6 2L	11	5-15.5	25	5
fba 385	*HindIII*	187 2L	6	4.4-15.5	10	3
fba 314	*HindIII*	209.5 2L	14	4-14	42	2
bcd1231	*HindIII*	211.4 2L	19	2.2*14.5	91	6

* localized on homoeologous group 2B chromosomes of wheat

2.4. FAC coordinates phenotypes and mapping

On the scale of the whole Mediterranean basin, all the populations were included, Moroccan, Bulgarian and French accessions being represented by one of their three individuals, drawn at random. Factorial Analysis of Correspondences (FAC) was computed on the matrix of individual RFLP data to obtain the principal coordinates (CORRESP procedure SAS, 1988). Significant geographic structuration was then sought on the different canonical axes declaring the nine *a priori* defined regions as a controlled factor in ANOVA. The determination coefficient of the model (R^2) estimated the part of the variation carried by an axis that could be explained by this *a priori* geographic structuration.

2.5. Geostatistics

A variogram was computed for genetic distance between pairs of individuals according to their geographic distance. Geographic distance was subdivided into 20 classes and restricted to half of the maximum distance to allow every accession to be equally represented in each distance class [28]. The experimental variogram $g^*(h)$ at the distance class h was calculated as:

$$g^*(h) = \frac{1}{W(h)} \sum_{l<k} w(y_l, y_k, h) Z(y_l, y_k)$$

where $Z(y_l, y_k)$ was the Excoffier genetic distance, $\sum_{i=1}^{55}(y_{l,i} - y_{k,i})^2$, calculated for the genotypes $y_{l,i}$ at each i RFLP band between individuals y_l and y_k,

$$W(h) = \sum_{l<k} w(y_l, y_k, h)$$

where $w(y_l, y_k, h) = 1$ if populations y_i at point i and y_k at point k are in the distance class h, and 0 otherwise [18].

To determine the distance classes for which a significant correlation existed between genetic and geographic distances, genotypes were randomly permuted 200 times among the geographical locations and $g^*(h)$ was computed. This defined a 95% confidence interval of $g^*(h)$ assuming the null hypothesis of no spatial correlation. The average value of the variogram assuming the null hypothesis, $g^*(Ho)$, was used to calculate the ratio of the total genetic variance that could be found between two individuals separated by distance h, $g^*(h)/g^*(Ho)$. Even if the Mediterranean basin cannot be considered as a continuous area, we used the Euclidean distance. Other geographic distances were tested and indicated the same trends (results not shown).

2.6. Testing the efficiency of the *M* strategy to build a core collection

The Maximization strategy or *M* strategy [22] was chosen to build a core collection. This method seeks to maximize the allelic richness retained in a core collection at a set of marker loci. Its efficacy is based on linkage disequilibrium between markers. It supposes that when evolutionary trajectories are different from one population to another, the whole genome is affected by events like isolation by distance, migration or selection. Molecular markers can be used to unravel these events and identify the most complementary populations from an evolutionary point of view. A gain in the capture of diversity on anonymous sites in the genome is then expected from the use of markers. Simulation [2] showed that this method was efficient compared to other methods. The simulations suggested that the more restricted the gene flow and the higher the selfing rate, the more efficient the method. We therefore explored the efficiency of the *M* strategy for our data set: *Ae. geniculata* is a strong selfer [13] and the *M* strategy was expected to be efficient on a scale as large as the Mediterranean basin.

Individuals with no missing data were used. We randomly split at the RFLP bands into two sets. The first is hereafter called the markers set. Core collections of different sizes (N) were built by maximizing the allelic richness for these selected bands (each band was coded 0 or 1, so that the minimum allelic richness by band was 1 and the maximum 2). The remaining bands constituted the set of targeted bands hereafter called the targets. They represented a sample of anonymous sites in the genome on which the efficiency of the *M* strategy will be tested. We used the Mstrat software [12] (available at http://www.ensam.fr/gap/resgen88) for sampling core collections. Mstrat uses as first criterion the maximization of allelic richness of a subsample and can also use as a secondary criterion the maximization of heterozygosity. This option was selected. As a baseline for comparisons, Mstrat generated core collections by random drawing of N accessions in the whole collection. For both core collections established with the markers or by random choice, Mstrat computed the allelic richness for the targets, $Mark_n$ or $Rand_n$ respectively. The *M* strategy was considered to be efficient when the allelic richness captured for targets using markers was higher than that obtained by random sampling. The relative efficiency of the *M* strategy over the random strategy, hereafter R_n, was measured for a core collection of size N by the ratio of the scores: $R_n = (Mark_n - Mark_1)/(Rand_n - Rand_1)$, where $Mark_1$ and $Rand_1$ were equal to the number of targeted bands, which was the diversity when only one individual was sampled. Note that the maximum allelic richness was 2 $Mark_1$ in that case. When the random sampling captured one new allele, R_n were captured using the markers.

As many core collection of size N with identical allelic richness could be obtained, the efficiency results were based on 50 independent core collections. To avoid intrinsic correlation in allelic richness of the two sets due to allelism, we

included all the bands generated by a single RFLP probe in the same category, either as markers or targets.

2.7. Minimum size core collection

On 100 repeats using all the RFLP bands as markers, we sought the minimum size, N^*, for which some core collections over the 100 counted all the marker alleles contained in the whole collection. Five hundred independent core collections of size N^* were then constructed by Mstrat and the N^* plants that were present the most frequently over the 500 repeats were kept as the basis of the core collection. As these N^* plants could not match all the allelic richness because they were selected from independent core collections, plants were added until the total allelic richness was obtained.

3. RESULTS

The nine probes used in combination with *EcoRI* and *Hind*II enzymes revealed a total of one hundred and sixteen polymorphic restriction fragments (Tab. I) resulting in a mean of 27 polymorphic bands per clone and in 2 to 69 different hybridization patterns. Of the total number of polymorphic fragments detected, 55 were present at frequencies higher than 5% and were kept for the genetic structure analysis. Unfortunately, we were not able to detect the allelic relationships between the markers and we only used the presence/absence information.

3.1. Fst statistics

Individuals with too many missing data were discarded and the analysis was carried out on 42 populations and 55 polymorphic bands. The regions (France, Bulgaria and Morocco) and populations within these regions showed significant effects on the structuration of the genetic diversity. Fisher tests from ANOVA indicated that the region effect was significant at the level 5% for 30 out of the 55 bands, and the populations within regions effect was significant 51 times out of 55. The Fst_R mean value was 0.15, but the most of the differences were due to populations within regions ($F_{P/R} = 0.52$). The genetic variation appeared therefore to be highly structured (Fst total $= 0.68$) and only 32% of the variation was found at the intrapopulation level. The distributions of the 55 Fst values are given in Figure 2. Fst_R were mainly distributed between 0 and 0.2 but some of them had values higher than 0.3. A value of 0.7 was even found, corresponding to a band almost specific to the France populations (frequency of 0.77 in this country but less than 0.05 in Bulgaria and Morocco). The Fst between-population values obtained from the 55 markers were widely distributed with few points lower than 0.15 (Fig. 2.)
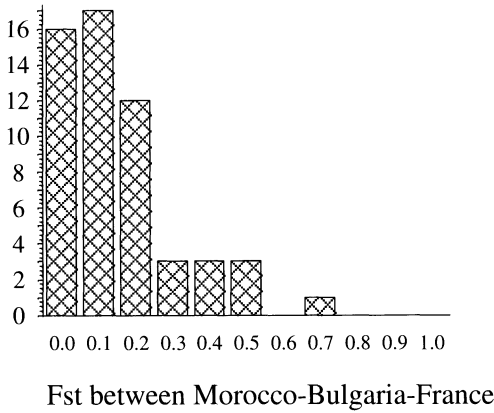
Fst between Morocco-Bulgaria-France

Figure 2. Distribution of Fst and the relative part of within-population diversity on 55 RFLP bands. Fst were calculated according to [30] using ANOVA with a nested model: Region + population (Region). In the text, Fst between regions is Fst_R, Fst between populations within regions is $Fst_{P/R}$. Three regions Bulgaria, France and Morocco were represented by a total 42 populations of three individuals.
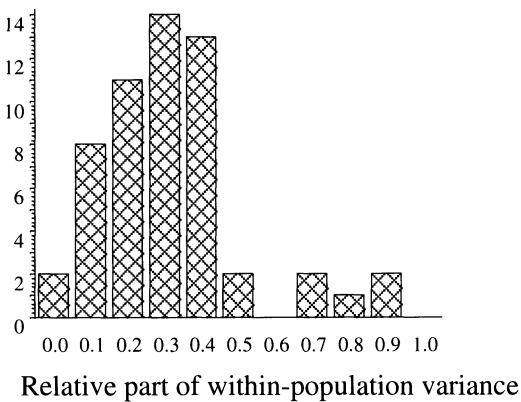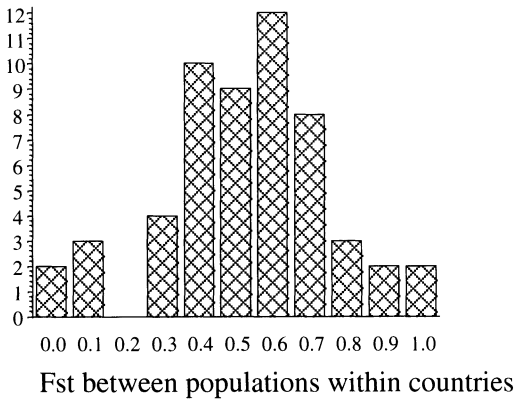

Fst between populations within countries

Relative part of within-population variance

Table II. Inertia and geographic structuration observed on the first eight principal component axes derived from Factorial Analysis of Correspondence of 55 RFLP markers in 147 individuals of *Ae. geniculata* originating from the Mediterranean basin. ANOVA was carried out on the individual FAC co-ordinates by declaring *a priori* defined eco-geographical regions as controlled factor. P values are given by the Fisher test of ANOVA, and R^2 is the determination coefficient of the model.

Axis	Inertia	Geographic structuration	
		R^2	P *value*
Axis 1	8.91	0.47	$P < 0.0001$
Axis 2	8.00	0.16	$P < 0.01$
Axis 3	5.85	0.26	$P < 0.0001$
Axis 4	5.27	0.27	$P < 0.0001$
Axis 5	4.87	0.17	$P < 0.001$
Axis 6	4.51	0.19	$P < 0.001$
Axis 7	4.27	0.14	$P < 0.01$
Axis 8	3.87	0.14	$P < 0.01$

3.2. Multivariate analysis

Four populations, originating from the Libyan region of Jabal al Akhdar, showing a number of specific bands, were discarded from the data. The FAC analysis was then performed on the 147 remaining populations and the 55 RFLP markers. The first eight axes explained 45.6% of the total diversity. Their inertia and the determination coefficient of the ANOVA declaring *a priori* regions are given in Table II.

The variation on axis 1 was the most linked to the geography ($R^2 = 0.47$). The Fisher test contrasted significantly a northern group constituted by French and Italian populations and a southern Mediterranean group of LEV, MED, MAT and TUR populations. The ADR, IBE and BAL regions appeared to be intermediate (Fig. 3a). Axis 2 discriminated IBE from MED and BAL regions (Fig. 3a). On axes 3 and 4, significant differentiation appeared between close regions (Fig. 3b). On axis 3, the Balkan region contrasted significantly TUR and LEV. When mapping coordinates (figure not shown), the populations from the Black sea coast appeared to be different from the populations of the other regions. On axis 4, populations from Iberian peninsula clearly differed from those located on the other side of Gibraltar straits (MAT and MED regions).

3.3. Relationships between genetic and geographical distances

Experimental $g^*(h)$ values (Fig. 4) indicated a significant correlation at the 5% confidence level between genetic and geographical distance for the first two distance classes, 100 and 200 km on average. As $g^*(100) = 0.182$ and $g^*(Ho) = 0.203$, 90% of the total diversity could be found at a distance of
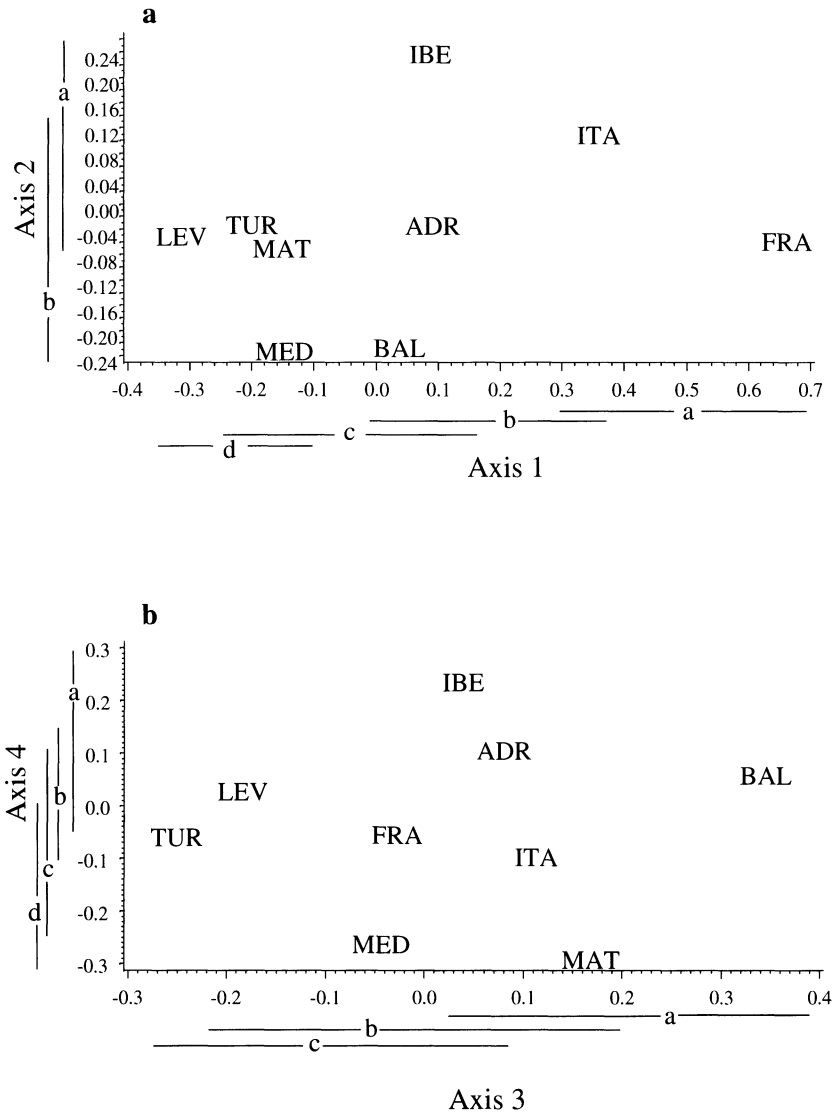
a

b

Figure 3. Mean values of the nine *a priori* defined eco-geographical regions for the four first axes resulting from the Factorial Analysis of Correspondence of 55 RFLP. On the bottom and the side of graphs, the lines and their associated letter indicate the groups of mean values that were not significantly different. These differences were tested by the Ryan Eynot Gabriel Welsh test ($\alpha = 5\%$) and ANOVA, with eco-geographical regions as a controlled factor. IBE: Iberian peninsula, FRA: France, ITA: Italy, ADR: Croatia, Albania and Greece, BAL: Bulgaria and European part of Turkey, TUR: Eastern part of the Turkey , LEV: Syria, Cyprus, Jordan, Lebanon, Palestine, MED: Libya, Tunisia, Algeria, East Morocco, MAT: West Morocco.
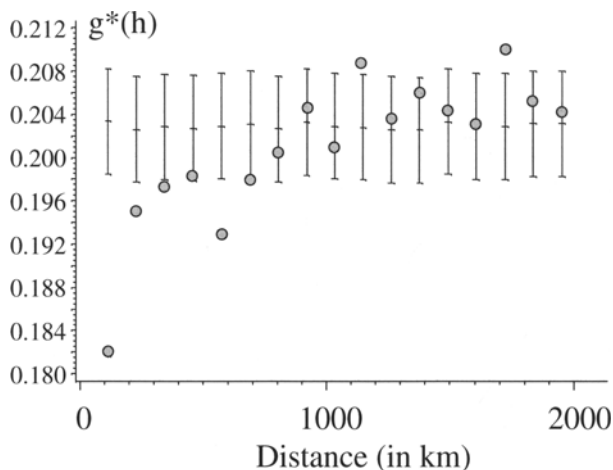
Figure 4. Experimental variogram $g^*(h)$, calculated from the Excoffier genetic distance for 55 RFLP bands in 20 distance classes (see text). Grey circles are observed $g^*(h)$ values. Vertical lines show the 95% envelope of $g^*(h)$ distribution assuming the null hypothesis of no spatial correlation, obtained after 200 random permutations of the genotypes among the prospected sites.

100 km. At 200 km, this portion of variance is almost 100%. $G^*(h)$ increased rapidly with distance and the spatial correlationwas no longer significant for distances larger than 200–300 km.

3.4. Efficiency of molecular markers to sample core collection

For the 202 selected accessions, 65 bands were observed for the 5 RFLP probes used as markers. In this set, the M strategy captured the whole allelic richness ($65*2 = 130$ alleles) with only 17 individuals on average (Fig. 5). This contrasted sharply with the random strategy which needed 21 accessions to achieve 80% of the diversity of the markers (116 alleles), a level for which the M strategy required 8 individuals. It appeared clearly that the use of the markers permitted a gain in the capture of the polymorphism of the 51 targeted bands. The relative efficiency R, reached its maximum value (1.3) for values of N between 7 and 9 (Fig. 5). Then the relative efficiency decreased with N to a final value of 1.16 when all the marker bands were captured ($N = 17$). As said previously, a core collection of size $N = 1$ already contains half of the alleles (when coding $0/1$), so we computed the efficiency of the method for capturing the other half of the alleles. For $N = 17$, $\text{Mark}_{17} - \text{Mark}_1 = 34.34$ and meaning that 67% of 51 complementary targeted alleles were captured (58% only for the random strategy). This gain was highly significant (t test, $P < 0.0001$). After $N = 17$, the superiority of the marker sampling continued to be significant. Therefore, according to the criterion proposed for building a core collection,
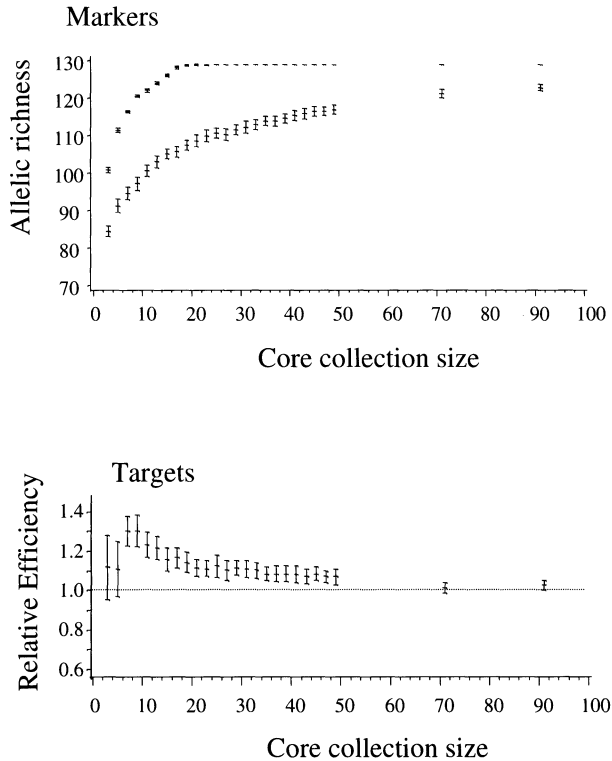
Markers

Targets

Figure 5. Efficiency of the M strategy compared to a random sample of accessions in building a core collection. Top: Allelic richness captured by the M strategy and random sampling in core collections of different sizes from a base collection polymorphic for 65 RFLP markers. Bottom: Efficiency of the M strategy relative to a random sampling for capturing allelic richness at target RFLP bands when maximizing the allelic richness for independent RFLP markers (see text).

the capture of 80% of the 51 supplementary alleles needed 40.26 accessions on average with markers (Min 23–Max 91) instead of 50.6 accessions by random sampling (Min 31–Max 91).

Using the whole set of 116 bands and with 100 repeats, core collections of size $N^* = 26$ contained all the marker alleles of the collection, and this represented 13% of the size of the base collection. The mean diversity value ($h_i = 2p_i(1-p_i)$ with p_i being the frequency of the presence of the marker i) in core collections of that size, was 0.216 on average, while it was 0.181 for the whole collection. This absolute difference was not very high but was shown to be significant by a Wilcoxon rank test on the 116 markers ($W- = 630$, $P < 0.0001$, data not shown). Of 500 core collections of size $N^* = 26$, 399 reached the maximum score of allelic richness. Of these, only 102 out of the 202 individuals were represented at least once. Of the 26 most frequently sampled accessions, the

Table III. List of accessions included in *Ae. geniculata* core collection.

Region	Accession reference - Individual code
ADR	1044, 839
BAL	89E140-1, 47733
FRA	3H-1, 95/1F-2, 95/1F-3, 95/2F-3
IBE	22734, 46698, 5516
ITA	802, 854
LBY	48178, 48190
LEV	47277, 48807
MAT	39M-1, 39M-3, 47M-1
MED	22M-2, 25M-1, 25M-3, 37M-2, 48166, 48381, 48470
TUR	47090

presence in the core collection varied from 99 times to 399 out of 399. Together, these 26 individuals made a core collection that lacked only two alleles. Thus two supplementary individuals were added to achieve the total allelic richness. The list of accessions is given in Table III. The geographic representativeness of this core collection was good. The diversity of the southern part of the Mediterranean basin was represented by a number of Maghrebian populations and by relatively few from the Near East.

4. DISCUSSION

Ae. geniculata showed a high level of structuration of its genetic variability. For Bulgaria-France-Morocco populations, most of the variation was explained by differences between populations within the same country (52%) while differences between countries accounted only for 15%. This was consistent with the variogram calculated for the whole Mediterranean basin, since at a distance as short as 100 km a large amount of diversity could be found between populations. Nevertheless, a spatial genetic correlation was demonstrated within a range of 100–200 km. For a highly selfing species, this spatial correlation at short distance can be interpreted as the result of current migration/drift processes and the high level of diversity found between populations as a direct consequence of the reproductive isolation due to preferential selfing.

The level of within-population variance (32%) was lower than that usually found for highly selfing species (c.a. 50%) [14]. During collecting explorations, mother plants could not always be separated, and in some cases individuals used to represent a population could come from the same plant. This would have reduced the level of within-population diversity. For selfing species, the maintenance of within-population diversity can result from migration. Some authors also refer to selection in a fine grained environment [20]. In our case, the spatial correlation observed between neighboring populations suggested that recurrent

migration regularly occurs between close populations. As *Ae. geniculata* is a poor pollen producer [13], most dispersion should occur by seed dispersal, over short distances by ants or possibly by grazing sheep that can disperse bearded spikes over greater distances.

Diversity was also found to be structured over great distances. The structuration observed between France, Bulgaria and Morocco and also between the *a priori* defined regions proved that isolation by distance played a role in the genetic structuration of *Ae. geniculata*. FAC analysis suggested that the colonization of the Mediterranean basin occurred *via* two routes from an initial area located around Turkey and Near East. The Gibraltar and Bosphorus straits could be considered as the limits of these two groups. For the Northern group, France and Italy appeared to be the most different from the others, while the Iberian peninsula, the Adriatic and Balkans showed some similarities with the southern group. The Balkan region exhibited a strong specificity mainly due to the populations originating from the Black Sea coast. The differentiation between populations from the coastal and central regions of Bulgaria was also revealed by RAPD markers [32]. The Iberian region also appeared different from its close neighbors, the Maghreb and to a lesser extent from France. This pattern could be explained by the presence of *Ae. geniculata* before the last glaciation in the North and the rescue of ancient populations in refugia that could have been the Balkans and Spain. It is also possible that Southern Italy played a similar role and France could have been recolonized by Italian populations. These areas have often been cited as refugia for other European plant species [7, 23, 26]. The southern group appeared less differentiated. The Near East and Turkey could be considered as a homogeneous area for *Ae. geniculata* diversity, while some differences could be detected between the Maghreb and Atlantic Morocco regions. This could mean that the periods at which the desert front isolated the West from the East of North Africa [1] had little influence on the variability of *Ae. geniculata*. Nevertheless, such an isolating effect could be seen in the existence of the very particular form of the Libyan region of Jabal al Akhdar.

The hypothesis of dissemination through agriculture appeared less evident from the data presented here but can not be definitely discarded. Agriculture spread *via* at least three pathways, one in the south from the Near East to Egypt, the second one along the northern Mediterranean coast, and the last one *via* the Danube plain towards Central Europe [19]. Accessions from France and Italy, from the Balkans and those from the southern group would be representative of these three routes of agriculture dissemination. The other regions, the Iberian peninsula, the Adriatic and part of the Balkans would then be areas of migration confluence.

The data provided good evidence that the linkage disequilibrium between loci that can be inferred from the observed genetic structuration was sufficient to permit the M strategy to capture the main part of the studied diversity. The gain in capturing allelic diversity at anonymous loci in the genome was

real. The gain of the M strategy is probably due to linkage disequilibrium between different parts of the genome, not necessarily physically linked. Indeed, part of the linkage disequilibrium is created between populations according to different evolutionary trajectories, under the effects of local genetic drift, selection, restricted gene flows and also by mutation. Such a disequilibrium can concern loci on different chromosomes. Simulations demonstrated that the M strategy was most efficient in the case of restricted gene flows and for selfing species [2], and our study on *Aegilops geniculata* provided some experimental evidence. We demonstrated that diversity was structured according to the distance between populations and that there are two major dissemination routes. This should explain a part of the success of the M strategy to capture diversity from the different eco-geographical regions. But we also observed that a part of the diversity could be found locally either between individuals in the same population or between close populations. This was illustrated in the list of the core collection by populations from Morocco and France. The maximizing strategy allowed use of all the levels of structuration of the genetic diversity whereas other methods sampling among groups, that are based either on passport data or on numerical classification, seem to be less efficient in sampling within groups. We did not compare different methods, but results from other data sets are in agreement with this assertion (Gouesnard, pers. comm.).

Concerning the ideal size of a core collection, our data suggested that it could be greater than the size at which all markers are captured. Indeed, after all markers have been captured, the addition of new accessions chosen randomly still added diversity on anonymous target loci. Some important questions remain to be investigated: how many supplementary accessions should be selected when all the revealed molecular diversity has been captured? How many markers are necessary to capture 80% of the diversity on the whole genome? The answers would certainly greatly depend on the degree of structuration between populations and on the distribution of the markers along the genetic map. In our study, markers and targets were supposed to belong to the same homeology group and the efficient capture of the targeted diversity could come from this physical relationship, but not necessarily. Indeed, our protocol planned to split markers into two independent groups of linked locus due to the allotetraploidy of *Ae. geniculata* to verify if physical linkage was important in the capture of diversity, by contrasting efficiency within and among sets of markers. Because we could not identify allelisms, this comparison remains to be done with other markers or for another species for which the genetic knowledge is important. Intuitively, one can argue that when the molecular information increases, more accessions would be required to capture the allelic richness on the markers and more sites in the genome would be represented.

ACKNOWLEDGEMENTS

REFERENCES

[1] Adams J.M., Faure H. QEN members, Review and Atlas of Palaeovegetation: Preliminary land ecosystem maps of the world since the Last Glacial Maximum. Adams J.M., Faure H (eds.) Oak Ridge National Laboratory, USA, 1997, htpp://www.esd.ornl.gov/ern/qen/adams1.html.

[2] Bataillon T., David J.L., Schoen D.J., Neutral genetic markers and conservation: simulated germplasm collections, Genetics 144 (1996) 409–417.

[3] Benlhabib O., Hmoud R., El Khlifi O., Nachit M., Baum M., Sharma H., Zaharieva M., Croisements interspécifiques pour l'élargissement de la diversité génétique des blés au Maroc, in: Birouk A, Rejdali M. (eds.), Ressources phytogénétiques et développement durable, Rabat, Maroc, 1997, pp. 183–186.

[4] Brown A.H.D., The core collection at the crossroads, in: Hodgkin T., Brown A.H.D., Van Hintum T.J.L, Morales E.A.V. (Eds.), Core collection of plant genetic resources, IPGRI, John Wiley and Sons, UK, 1995.

[5] Church G.M., Gilbert W., Genomic sequencing. Proc. Natl. Acad. Sci USA 81 (1984) 1991–1995.

[6] Dimov A., Zaharieva M., Mihova S., Rust and powdery mildew resistance in *Aegilops* accessions from Bulgaria, in: Damania A.B. (Ed.), Biodiversity and wheat improvement, John Wiley and Sons, London, 1993, pp. 165–169.

[7] Dumolin-Lapègue S., Demesure B., Fineschi S., Le Corre V., Petit R.J., Phylogeographic structure of white oaks throughout the European continent, Genetics 146 (1997) 1475–1487.

[8] Farooq S., Shah T.M., Asghar M., Intergeneric hybridization for wheat improvement: V. Production of and metaphase 1 chromosome analysis in F1 hybrids of wheat (*Triticum aestivum*) with Aegilops *ovata* L., Cereal Res. Com. 24 (1996) 155–161.

[9] Frankel O.H., Genetic perspectives of germplasm conservation. in: Arber M., Llimensee W.K., Peacock W.J., Starlinger P. (Eds.), Genetic manipulation: Impact on Man and Society, Cambridge University Press, Cambridge, UK, 1984, pp. 173–196.

[10] Friebe B., Jiang J., Raup W.J., McIntosh R.A., Gill B.S., Characterization of wheat alien translocations conferring resistance to diseases and pests: current status, Euphytica 91 (1996) 58–87.

[11] Gill B.S., Sharma H.C., Raupp W.J., Browder L.E., Hatchett J.H., Harvey T.L., Moseman J.G., Waines J.G., Evaluation of *Aegilops* species for resistance to wheat powdery mildew, wheat leaf rust, Hessian fly and greenbug, Plant Dis. 69 (1985) 314–316.

[12] Gouesnard B., Bataillon T.M., Decoux G., Rozale C., Schoen D.J., David J.L., MSTRAT: an algorithm for building core collections by Maximizing Allelic or pherotypic richness, J. Hered. 92 (2001) 93–94.

[13] Hammer K., Vorarbeiten zur Monographischen Darstellung von Wildpflanzensortimenten: *Aegilops* L., Kulturpflanze 28 (1980) 33–180.

[14] Hamrick J.L., Godt M.J.M., Effects of life history traits on genetic diversity in plant species, Philos. Trans. R. Soc. London 351 (1996) 1291–1298.

[15] Kimber G., Sallee P.J., Feiner M.M., The interspecific and evolutionary relationships of *Triticum ovatum*, Genome 30 (1988) 218–221.

[16] Ma Z.Q., Gill B.S., Sorrells M.E., Tanksley S.D., RFLP markers linked to two Hessian fly-resistance genes in wheat (*Triticum aestivum* L.) from *Triticum tauschii* (Coss.) Schmal, Theor. Appl. Genet. 85 (1993) 750–754.

[17] Mamluk, O.F., Van Slageren, M.W., *Aegilops* spp. as sources of resistance to wheat diseases, in: Birouk, A., Rejdali, M. (Eds.), Ressources Phytogénétiques et Développement Durable, Actes, Rabat, Maroc, 1997, pp. 197–202.

[18] Monestiez P., Goulard M., Charmet G., Geostatistics for spatial genetic structures: study of wild populations of perennials ryegrass, Theor. Appl. Genet. 88 (1994) 33–41.

[19] Nesbitt M., Samuel D., From staple crops to extinction. The archaeology and history of the hulled wheats, in: Padulosi S., Hammer K., Heller J. (Eds.), *Hulled wheats*, Proceedings of the First International Workshop on Hulled wheats, Castelvecchio Pascoli, Tuscany, Italy, 1995, pp. 41–100.

[20] Nevo E., Noy-Meyer I., Beiles A., Krugman T., Agami M., Natural selection of allozyme polymorphisms: micro-geographic spatial and temporal ecological differentiations in wild emmer wheat, Isr. J. Bot. 40 (1991) 419–449.

[21] SAS®, Stat User's guide. Release 6.03 Edition, SAS Institute Inc, Cary, NC, USA, 1988

[22] Schoen D.J., Brown A.H.D., Maximising genetic diversity in core collections of wild relatives of crop species, in: Core collection : improving the management and the use of the plant germplasm collections, Workshop IBPGR/CGN/CENARGEN, Brasilia, 1994.

[23] Taberlet P., Fumagalli L., Wust-Saucy A.G., Cosson J.-F. Comparative phylogeography and postglacial colonization routes in Europe, Mol. Ecol. 7 (1998) 453–464.

[24] Tai T.H., Tanksley S.D., A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. Plant Mol. Biol. Rep. 8 (1990) 297–303.

[25] Valkoun J., Hammer K., Kucerova D., Bartos P., Disease resistance in the genus *Aegilops* L. Stem rust, leaf rust, stripe rust, and powdery mildew, Kulturpflanze 33 (1985) 133–153.

[26] Van Dijk P., Bakx-Schotman T., Chloroplast DNA phylogeography and cytotype geography in autopolyploid *Plantago media*, Mol. Ecol. 6 (1997) 345–352.

[27] Van Slageren M.W., Wild wheats: a monograph of Aegilops L. and *Amblyopyrum* (Jaub. and Spach) Eig (*Poaceae*), Wageningen, Agricultural University – Aleppo, International Center for Agricultural Research in Dry Areas, 1994.

[28] Wackernagel H., Multivariate geostatistics: An introduction with applications, Springer Verlag Berlin Heidelberg, 1995.

[29] Wang G. Z, Miyashita N.T., Tsunewaki K., Plasmon analyses of *Triticum* (wheat) and *Aegilops*: PCR-single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. Proc. Nat. Acad. Sci. USA 94 (1997) 14570–14577.

[30] Weir B.S., Genetic Data Analysis II. Methods for discrete population genetic data. Sinauer Associates, Sunderland, 1996, pp. 183–185.

[31] Zaharieva M., *Aegilops* species in Bulgaria. Their ecogeography and distribution. In: Damania A.B. (Ed.), Biodiversity and Wheat Improvement, John Wiley and Sons, Chichester, UK, 1993, pp. 369–374.

[32] Zaharieva M., David J., This D., Monneveux P., Analyse de la diversité génétique d'*Ae. geniculata* Roth en Bulgarie, Cah. Agric. 8 (1999) 181–188.