Molecular Ecology (2001) **10**, 1371–1383

# Spatial effects and rare outcrossing events in *Medicago truncatula* (Fabaceae)

ISABELLE BONNIN,* JOELLE RONFORT,† FREDERIC WOZNIAK* and ISABELLE OLIVIERI‡
*Laboratoire de Génétique et Evolution des Populations Végétales, Bâtiment SN2, Université de Lille 1, 59655 Villeneuve d'Ascq cedex, France, †INRA Montpellier, Domaine de Melgueil, 34130 Mauguio, France, ‡Institut des Sciences de l'Evolution, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier cedex 05, France*

**Abstract**

**In order to elucidate the mechanisms underlying the large amount of RAPD polymorphism found in 1990 in a population of the selfing annual *Medicago truncatula* GAERTN. (Fabaceae), we have analysed most of the individuals ($n$ = 363) from the same population 6 years later using microsatellite loci. We confirm the result of the earlier study, namely that this population is very polymorphic and highly subdivided, with approximately 37% of the variance distributed among subpopulations, only 50 m apart one from another. We use standard *F*-statistics analyses, linkage disequilibria, minimum spanning network, multilocus assignment tests and spatial autocorrelation analyses to test the hypotheses that spatial structure and outcrossing events are involved in maintaining the large amount of genetic diversity at the level of each subpopulation. Interestingly, fine-scale spatial structure could be observed in only one subpopulation suggesting that other mechanisms are acting elsewhere. To the best of our knowledge, this is the first study of fine spatial genetic structure in a predominantly selfing species.**

*Keywords*: *Medicago truncatula*, microsatellites, population structure, RAPD, selfing rate, temporal variation

*Received 23 August 2000; revision received 1 January 2001; accepted 1 January 2001*

## Introduction

As shown by empirical studies based on isozymes and molecular markers, self-fertilizing plant species tend to have less genetic variation within population than outcrossing plant species, and greater among population differentiation (Loveless & Hamrick 1984; Hamrick & Godt 1990, 1996). These two features can be viewed as direct consequences of the mating system. Indeed, inbreeding reduces the effective population size, and consequently enhances the effect of genetic drift. High selfing rates also reduce the effective recombination rate, so that neutral alleles are drawn by selectively favoured or deleterious alleles, a process that ultimately reduces variation at neutral loci (Charlesworth *et al*. 1993; Nordborg & Donnelly 1997).

Correspondence: Isabelle Bonnin. Fax: 33 (0)3 20 43 69 79; E-mail: bonnin@univ-lille1.fr

A less well understood characteristic of self-fertilizing species is their greater population to population variation in gene diversity compared to outbreeders (Brown & Schoen 1991), some populations exhibiting levels of gene diversity resembling those generally observed in outcrossing populations (Bonnin *et al*. 1996a; Awadalla & Ritland 1997; Kuittinen *et al*. 1997). Although some connections with the population's ecological characteristics have been reported, the factors and mechanisms responsible for the high level of genetic diversity maintained in such populations remain unclear (Brown & Schoen 1991). Under predominant selfing, a population is expected to fracture into essentially isolated lineages. Assuming limited seed dispersal distances, and rare but recurrent outcrossing events, a self-fertilizing population is thus likely to be subdivided into small neighbourhoods that consist of single differentiated lineages. Such subdivision is expected to reduce the effect of drift and could thus play a major role in the maintenance of genetic diversity at the whole population level (Barton & Whitlock 1997). Empirical supports for this

hypothesis require some knowledge on the genetic relatedness between lineages and on how lineages are distributed and arranged within populations. Such information is, however, generally not available in published data on the genetic variation of individual populations of selfing species.

In this paper, we examine the fine spatial scale organization of genetic variation in a single population of the selfing annual *Medicago truncatula* GAERTN (Fabaceae). A former study of population structure in this species, using 21 random amplified polymorphic DNA (RAPD) markers and one morphological marker showed a very high level of polymorphism in one of the four study populations (Bonnin *et al*. 1996a). Given the sampling strategy that was used (each natural population was subdivided into three subpopulations 50 m apart, in which one plant was collected every metre alongside a 30 m transect), it was also possible to detect a high level of spatial structure in this population. Using a multilocus approach, only one multilocus genotype was found to be common in two subpopulations, indicating a very low amount of seed dispersal between subpopulations. The within-subpopulation variability was large and different mechanisms have been suggested to explain these results: (i) small neighbourhood sizes within subpopulations; (ii) higher outcrossing events than commonly recognized in this species; and (iii) dispersal in time *via* the seed bank (Bonnin *et al*. 1996a).

In order to examine the level and patterns of 'within subpopulation' structure (previously difficult to observe given the sampling strategy used), the entire population was collected in 1996 (six years later), and each sampled plant was mapped. Five microsatellites loci were used instead of RAPD. We find that, six years later, the revealed genetic diversity is still large and highly structured among the three subpopulations. Despite a high level of diversity in all three subpopulations, only one exhibits an important within subpopulation spatial structure. For the two others, different (or supplementary) mechanisms must be invoked to explain the maintenance of a large amount of genetic diversity despite predominant self-fertilization.

## Materials and methods

### *Population studied*

The population of Aude is located in the French Mediterranean region. Subpopulation 1 (30 m × 3.5 m) and subpopulation 3 (32 m × 5 m) are bordering a vineyard which is ploughed and harvested every year. They are approximatively 30 m apart. Subpopulation 2 (30 m × 6 m) is bordering an old abandoned vineyard and is located 50 m from the others (Bonnin *et al*. 1996a). Respectively, 139, 127 and 114 flowering individuals were observed in subpopulation 1, 2, and 3 at the time of sampling. Leaves were collected on 365 individuals out of 381 found in the total population.

### *Microsatellite analysis*

DNA extractions were performed on dried leaves as described in Bonnin *et al*. (1996a). Polymorphism was assayed on each DNA sample at five microsatellite loci (MTSA5, MTSA6, MT660252, MT660456, and MT660538) developed by Baquerizo *et al*. (2001). MTSA5 and MTSA6 are dinucleotide repeat loci, MT660538 is a dinucleotide imperfect repeat locus, and MT660252 and MT660456 are trinucleotide repeat loci. The polymerase chain reaction (PCR) conditions were mainly as described in Baquerizo *et al*. (in press). DNA samples (10 ng of each) were amplified using 200 μM dNTP, 2 pmol of each primer (one of which was fluorescence labelled either with IRD-700 or IRD-800), 0.75 units of AmpliTaq DNA polymerase (Perkin Elmer), 1X *Taq* polymerase buffer, and 2 mM of $MgCl^{2+}$ in a total volume of 20 μL. Two co-amplifications were possible: TSA6 and 252 on the one hand, TSA5 and 538 on the other. Initial denaturation was 4 min at 94 °C followed by 30 cycles (94 °C for 45 s, 53 °C for 45 s and 75 °C for 45 s), and 8 min at 72 °C in a Perkin-Elmer Gene-Amp system 9600. Amplification products, loaded on 7% denaturing polyacrylamide gels, were analysed on a Li-Cor automated DNA sequencer 4200. Allelic sizes were determined with M13 sequence used as a size control.

### *Statistical analyses*

*Monolocus analyses.* Microsatellites loci probably evolve under a stepwise mutation model (SMM), although an infinite alleles model (IAM) cannot always be rejected (review in Jarne & Lagoda 1996). Rousset (1996) showed that the estimate of selfing rates using the classical formula $F_{IS} = S/(2 - S)$ holds regardless of the mutation model. Differently, genetic differentiation can be influenced by the mutation process, given that the relationship found between $F_{ST}$ and $Nm$ in a finite island model at equilibrium under IAM is inappropriate under SMM (Slatkin 1995; Rousset 1996). In this case, the use of the statistic $R_{ST}$, based on variance in allelic size, is recommended (Slatkin 1995). However, given that the accuracy of $R_{ST}$ estimators may be influenced by homoplasy (Goodman 1997) and alleles differing by a number of bases which is not a multiple of the repeat unit (as for locus MT660456 and MT660538 in our study), we did not use this estimator of genetic differentiation.

For each locus and each subpopulation, allelic frequencies, gene diversity ($H_E$), and deviations from Hardy–Weinberg expectations were calculated using FSTAT 2.8 (Goudet 1995). The significance of $F_{IS}$ value, estimated as in

Weir & Cockerham (1984), was tested by 1000 permutations of alleles within samples. The selfing rate ($S$) in each subpopulation was estimated using the formula $F_{IS} = S/(2 - S)$, which assumes genetic equilibrium under a mixed-mating model and no spatial structure. Genotypic linkage disequilibria within subpopulation, and differentiation among subpopulations were tested using the package GENEPOP, version 3.2 (Raymond & Rousset 1995). $F_{ST}$ estimates were computed according Weir & Cockerham (1984) and their significance determined by Fisher exact tests.

Within subpopulation structure was examined using AUTOCORF2, a program developed by Hardy & Vekemans (1999), which calculates conditional kinship coefficients between individuals according to geographical distances. Each subpopulation was subdivided into 23 distance intervals of 0.50 m, corresponding to a minimum of 100 pairs of individuals analysed in each interval. Kinship coefficients were calculated as described in Loiselle *et al*. (1995) and Ritland (1996), using individual multilocus genotypes, to obtain a more powerful test for spatial genetic structure (Heywood 1991; Smouse & Peakall 1999). Loiselle *et al*. (1995) weight estimates from each locus by their heterozygosity, while Ritland (1996) uses the number of alleles at the locus to obtain the optimal per locus weight. The method of Ritland should provide the estimators with the lowest sampling variance, so that they should be most powerful to detect a genetic structure (Sweigart *et al*. 1999). However, those estimates suffer downward bias when at least one allele occurs at low frequency (inferior to about 0.05; Ritland 1996). Given that this situation is encountered for most loci in our study, and because conclusions were identical whatever the method used, we chose to present Loiselle's estimator only.

*Multilocus analyses* Multilocus genotype diversity within subpopulation was measured by the Simpson index corrected for finite samples, D (Pielou 1969; Peet 1974):

$$D = 1 - [\Sigma n_i (n_i - 1)/N (N - 1)]$$

where $n_i$ is the number of individuals of genotype $i$ and $N$ is the total number of individuals. D ranges from zero in a population composed of a single genotype to one in a population where every individual sampled has a different genotype. This quantity is closely related to the probability that two randomly chosen haplotypes are different in the sample (Nei 1987).

A minimum spanning network of the 98 observed multilocus genotypes was drawn, using MINSPNET, a program provided by the software ARLEQUIN version 2.000 (Schneider *et al*. 2000). The network was computed from the distance matrix containing the number of differences among each pair of multilocus genotypes. Its construction assumes that each genotype is linked by a single or a series of mutational events to all other genotypes through a unique pathway. Given that monolocus analyses suggest very low levels of outcrossing in *Medicago truncatula*, as well as a high level of spatial structure, we used this approach to identify rare migration/recombination events among subpopulation as in Bonnin *et al*. (1996a). Assuming that most multilocus genotypes should be grouped according to their subpopulation of origin, those that do not fall into their own geographical group, are considered as immigrants into this group.

Assignment tests were employed to try to determine the origin of particular genotypes. They were implemented in the software ARLEQUIN version 2.000 (Schneider *et al*. 2000) using a method described in Paetkau *et al*. (1997) and Waser & Strobeck (1998). This method consists in determining the log-likelihood of each individual multilocus genotype in each population sample, assuming that the individual comes from that population. The likelihood is estimated by the probability of the individual's genotype at a particular locus being drawn at random from a population: $P = x_i^2$ for homozygotes and $P = 2\, x_i\, x_j$ for heterozygotes, with $x_i$ and $x_j$ the allele frequencies observed in the population for that locus. These values are multiplied across loci to give the likelihood of each individual multilocus genotype. Clearly, this method is not appropriate in the case of predominantly selfing species. However (and given that there is no other software available to perform assignment tests), it might allow to identify genotypes that are potentially the result of migration followed by several recombination events, rather than pure migration through seeds. More importantly, should the migrants identified on the network (assuming a selfing mating system) correspond well to those identified using the assignment tests (assuming complete outcrossing), we will be able to conclude that, provided a sufficient genetic differentiation among groups, the method of assignment test is robust to the mating system.

## Results

### *Monolocus analyses*

Allelic frequencies, expected heterozygosities and $F_{IS}$ values are given in Table 1. All microsatellite loci, except locus MT660252, showed six or seven alleles, with a high level of polymorphism at the whole population level (the expected heterozygosities per locus were 0.73, 0.75, 0.16, 0.67, and 0.59 for locus MTSA5, MTSA6, MT660252, MT660456, and MT660538, respectively). Overall, subpopulation 1 was slightly less variable than subpopulations 2 and 3 (Table 1). $F_{IS}$ estimates were large for the three subpopulations and varied slightly among loci (Table 1). Estimated selfing rates based on $F_{IS}$ values varied between 95% and 97% depending on the subpopulation.

| Locus | Allele size (pb) | A1 | A2 | A3 | Aude |
|---|---|---|---|---|---|
| MTSA5 | | ($n$ = 130) | ($n$ = 125) | ($n$ = 108) | |
| | 215 | 0.688 | 0.420 | 0.028 | 0.399 |
| | 235 | 0.088 | 0.016 | 0.523 | 0.193 |
| | 239 | — | 0.016 | 0.009 | 0.008 |
| | 241 | 0.158 | 0.104 | 0.426 | 0.219 |
| | 243 | 0.027 | — | — | 0.010 |
| | 245 | 0.012 | 0.444 | 0.014 | 0.161 |
| | 251 | 0.027 | — | — | 0.010 |
| | $H_E$ | 0.495 | 0.620 | 0.549 | |
| | $F_{IS}$ | 0.891 | 0.961 | 0.933 | |
| MTSA6 | | ($n$ = 122) | ($n$ = 122) | ($n$ = 103) | |
| | 136 | 0.061 | 0.029 | 0.010 | 0.035 |
| | 138 | 0.025 | 0.053 | — | 0.027 |
| | 140 | 0.783 | 0.025 | 0.039 | 0.295 |
| | 148 | 0.037 | 0.496 | 0.058 | 0.205 |
| | 152 | — | 0.012 | — | 0.004 |
| | 160 | 0.020 | 0.008 | 0.330 | 0.108 |
| | 164 | 0.074 | 0.377 | 0.563 | 0.326 |
| | $H_E$ | 0.379 | 0.612 | 0.574 | |
| | $F_{IS}$ | 0.848 | 0.920 | 0.932 | |
| MT660252 | | ($n$ = 130) | ($n$ = 125) | ($n$ = 109) | |
| | 109 | 0.031 | 0.020 | 0.225 | 0.085 |
| | 112 | 0.969 | 0.980 | 0.775 | 0.915 |
| | $H_E$ | 0.060 | 0.039 | 0.352 | |
| | $F_{IS}$ | 1.000 | 0.797 | 0.922 | |
| MT660456 | | ($n$ = 130) | ($n$ = 123) | ($n$ = 104) | |
| | 90 | — | 0.224 | 0.029 | 0.085 |
| | 93 | 0.008 | 0.049 | — | 0.020 |
| | 99 | 0.058 | — | 0.495 | 0.165 |
| | 102 | 0.781 | 0.313 | 0.457 | 0.525 |
| | 123 | — | 0.256 | 0.010 | 0.091 |
| | 132 | 0.154 | 0.159 | 0.010 | 0.113 |
| | $H_E$ | 0.366 | 0.765 | 0.550 | |
| | $F_{IS}$ | 0.979 | 0.979 | 0.948 | |
| MT660538 | | ($n$ = 131) | ($n$ = 125) | ($n$ = 109) | |
| | 132 | — | 0.012 | — | 0.004 |
| | 135 | 0.038 | 0.004 | — | 0.015 |
| | 143 | 0.076 | 0.064 | 0.009 | 0.052 |
| | 214 | 0.737 | 0.008 | 0.096 | 0.296 |
| | 216 | 0.111 | — | 0.101 | 0.070 |
| | 224 | 0.038 | 0.912 | 0.794 | 0.563 |
| | $H_E$ | 0.440 | 0.165 | 0.354 | |
| | $F_{IS}$ | 0.913 | 0.952 | 0.974 | |
| All locus | $H_E$ | 0.348 | 0.440 | 0.476 | |
| | $F_{IS}$ | 0.910 | 0.950 | 0.940 | |
| | S | 0.950 | 0.970 | 0.970 | |

**Table 1** Allelic frequencies, expected heterozygosities ($H_E$) and $F_{IS}$ in each subpopulation. $F_{IS}$ values tested by permutations of alleles within samples are significantly different from zero. The selfing rate [S = $2^*F_{IS}/(1 + F_{IS})$ ] is given for all locus in each subpopulation

| **Locus** | Total population | A1–A2 | A2–A3 | A1–A3 |
|---|---|---|---|---|
| MTSA5 | 0.32 | 0.19 | 0.37 | 0.40 |
| MTSA6 | 0.40 | 0.47 | 0.21 | 0.49 |
| MT660252 | 0.15 | 0 | 0.18 | 0.16 |
| MT660456 | 0.23 | 0.23 | 0.22 | 0.26 |
| MT660538 | 0.56 | 0.68 | 0.06 | 0.55 |
| **All locus** | 0.37 | 0.41 | 0.25 | 0.42 |
| All locus – MTSA5 | 0.38 | 0.47 | 0.19 | 0.43 |
| All locus – MTSA6 | 0.36 | 0.39 | 0.26 | 0.40 |
| All locus – MT660252 | 0.38 | 0.42 | 0.25 | 0.44 |
| All locus – MT660456 | 0.41 | 0.46 | 0.26 | 0.45 |
| All locus – MT660538 | 0.31 | 0.31 | 0.27 | 0.38 |

**Table 2** $F_{ST}$ for each locus among subpopulations, and between each pairwise subpopulation. Genotypic differentiation tested by Fisher exact tests is very highly significant except for locus MT660252 (A1–A2)

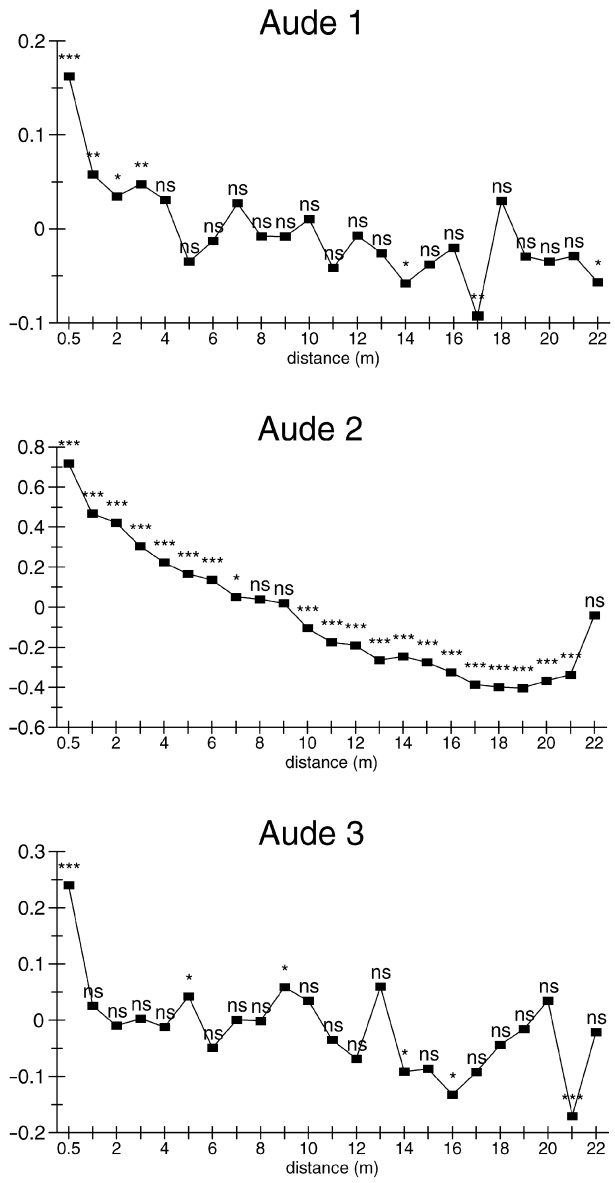


**Fig. 1** Loiselle kinship coefficients calculated over all loci. Non-significant coefficients, significant coefficients at $P < 0.05$, $P < 0.01$ and $P < 0.001$ are indicated by: ns, *, **, and ***, respectively.

Genetic differentiation among subpopulations is shown in Table 2. The value of $F_{ST}$ calculated at the population level was high, indicating a strong spatial structure within population. Overall, Aude 1 was as much differentiated from Aude 2 as from Aude 3 ($F_{ST}$ = 0.41 and 0.42, respectively), but $F_{ST}$ estimates were more variable among loci for Aude 1—Aude 2 (Table 2). The less differentiated subpopulations ($F_{ST}$ = 0.25) were also the most distant ones geographically, Aude 2 and Aude 3 being located 50 apart.

Analysing the fine-scale genetic structure showed that Aude 2 was highly structured in comparison to Aude 1 and Aude 3 (Fig. 1). In this subpopulation, a significant positive autocorrelation ($P < 0.05$) was found among individuals located up to 7 m apart (the kinship coefficient decreasing from 0.7 to 0.05), while the maximal distances showing significant positive autocorrelation in Aude 1 and Aude 3 were 3 m and 0.50 m (kinship coefficients $r$ of 0.05 and 0.24, respectively). Within the smaller distance interval (0.50 m), the mean kinship coefficient was maximal in Aude 2 (f = 0.71) and much greater than in Aude 3 and Aude 1 (f = 0.16 and 0.24, respectively). The same results were obtained considering each locus separately (not shown). Overall, subpopulation 2 showed a higher negative regression coefficient between kinship coefficients (genetic distances) and geographical distances (Fig. 1).

Pairwise linkage equilibria were tested for each pair of loci, in each subpopulation. Seven pairs showed highly significant linkage disequilibria in Aude 2 and Aude 3 ($P < 0.001$). They involved the five loci studied. In Aude 1, six combinations were found statistically associated ($P < 0.001$), locus MT660252 being in linkage equilibria with the other loci (data not shown). At the population level, the five loci were in strong linkage disequilibria ($P < 0.001$).

### *Multilocus analyses*

Using a multilocus approach, 98 genotypes were found in the sample of 334 individuals (on average each genotype was represented by roughly three individuals). Multilocus genotype diversity as calculated by the Simpson index (Pielou 1969; Peet 1974) was slightly lower in subpopulation 1 (0.71) than in subpopulations 2 (0.89) and 3 (0.86), although the number of multilocus genotypes (43) was larger in Aude 1 compared to the two other subpopulations (32 and 30 for Aude 2 and Aude 3, respectively). As expected from single locus analyses, a strong spatial structure was observed at the whole population level (Fig. 2). Out of the 98 multilocus genotypes, only six were shared by two subpopulations (five were common to Aude 1 and Aude 3, and one to Aude 1 and Aude 2), indicating few seed migration events among subpopulations. However, the minimum spanning network obtained from the number of pairwise differences matrix between multilocus genotypes showed migration/recombination events among subpopulations (Fig. 3), as some multilocus genotypes were genetically more closely related from genotypes of another subpopulation than to genotypes of the subpopulation from which they were sampled. For subpopulations 1 and 2, the majority of the genotypes were included in the main group of their subpopulation (42/43 of Aude 1, 20/31 of Aude 2), while most genotypes from Aude 3 (19/30) were interconnected with genotypes from Aude 2. This result is consistent with the lowest value of $F_{ST}$ found between subpopulations 2 and 3. Other genotypes from Aude 2 and Aude 3 (respectively 11/31 and 11/30) were observed in the main group of Aude 1, suggesting that this subpopulation is a

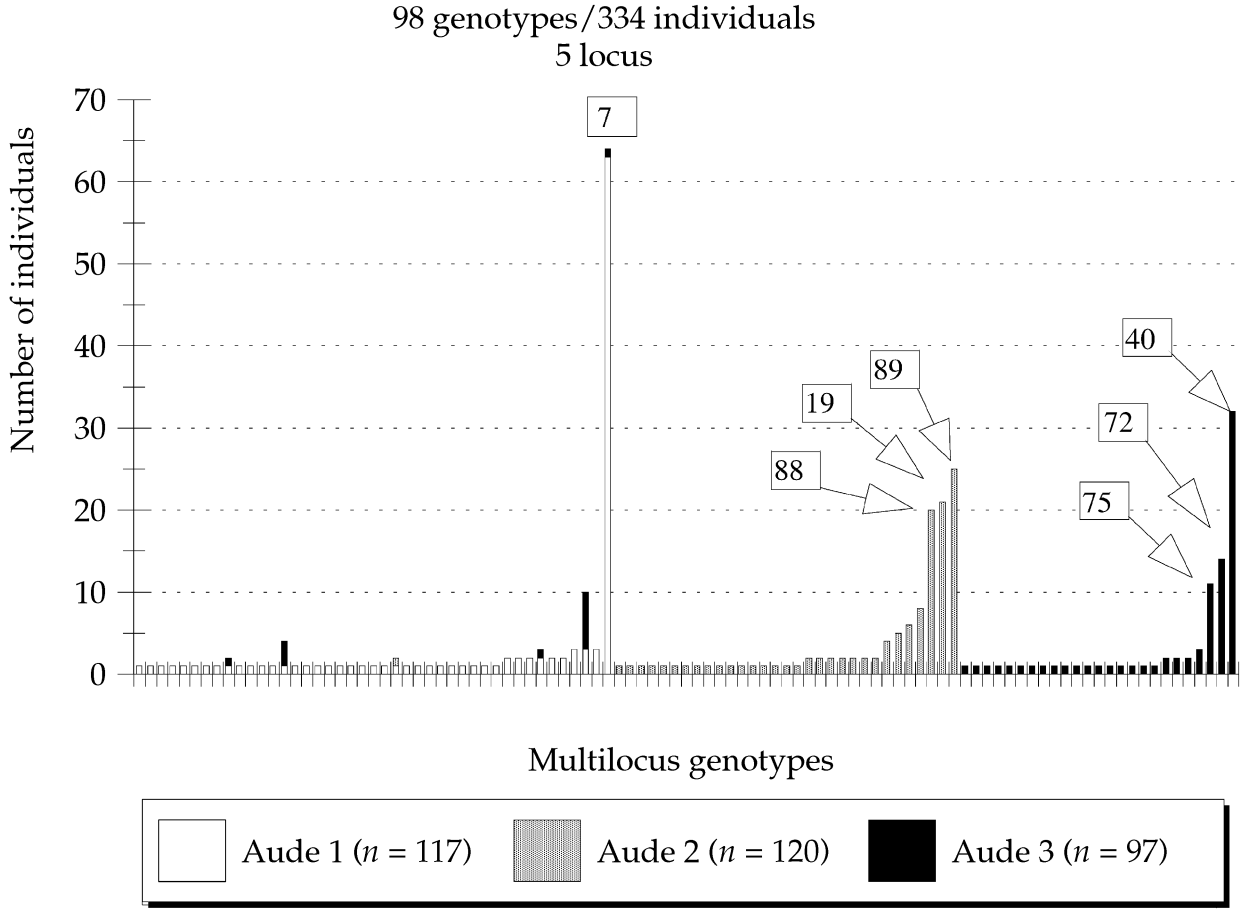


**Fig. 2** Distribution of the multilocus genotypes among subpopulations.

source of migrants for the other subpopulations. For example, it is likely that the five genotypes which are found to be common in subpopulations 1 and 3 in Fig. 2 (genotypes number 7, 13, 37, 38, 39), are immigrants from Aude 1 to Aude 3 (Fig. 3).

Figures 4, 5 and 6 show the spatial distribution of multilocus genotypes into their respective subpopulation. Subpopulation 1 possessed a very frequent genotype (number 7 with a frequency of 54%) distributed all over the subpopulation (Fig. 4), explaining the absence of within subpopulation structure observed using spatial autocorrelation analyses. Moreover, this most frequent genotype was close to the centre of the subpopulation network (Fig. 3), suggesting that it could be at the origin of most of the genotypes from Aude 1. Conversely, the more structured subpopulation, Aude 2, showed a very patchy distribution of multilocus genotypes (Fig. 5), with the most frequent genotypes (numbers 89, 19, and 88 with frequencies of 21%, 18% and 17%, respectively) observed both at the beginning and at the end of the subpopulation. As shown by the network, these most frequent genotypes are found in the main group of genotypes from Aude 2 (Fig. 3). Interestingly, genotypes from the middle of subpopulation 2 (numbers 46, 49, 52, 44, 55, 32, 63, 31, and 56; Fig. 5), which are less frequent, are also those which are located into the main group of genotypes from Aude 1 in Fig. 3, suggesting dispersal events from this subpopulation. In Aude 3, the three most frequent genotypes (numbers 40, 72, and 75 with frequencies of 33%, 14%, and 11%, respectively) are observed everywhere in the subpopulation (Fig. 6), which is in accordance with the low level of spatial structure previously found. Moreover, the genotype 40 seems to be originated from Aude 1, while genotype 72 and 75 are connected to the main group of Aude 2 (Fig. 3).

Results of assignment tests are shown in Fig. 7. As previously observed using monolocus analyses, they suggest few migration events among subpopulations. Overall, only 17 out of 98 multilocus genotypes were assigned to another subpopulation than the one from where they had been sampled: nine genotypes in Aude 1 (genotypes numbers 24, 30, 37, 38, 39, 51, 65, 74, and 86), four in Aude 2 (genotypes numbers 31, 46, 52 and 55), and four in Aude 3 (genotypes numbers 7, 13, 81 and 96). Except for Aude 1, these genotypes are those which we previously identified as migrants using the network pattern. With regard to the origin of migrant genotypes, six genotypes came from

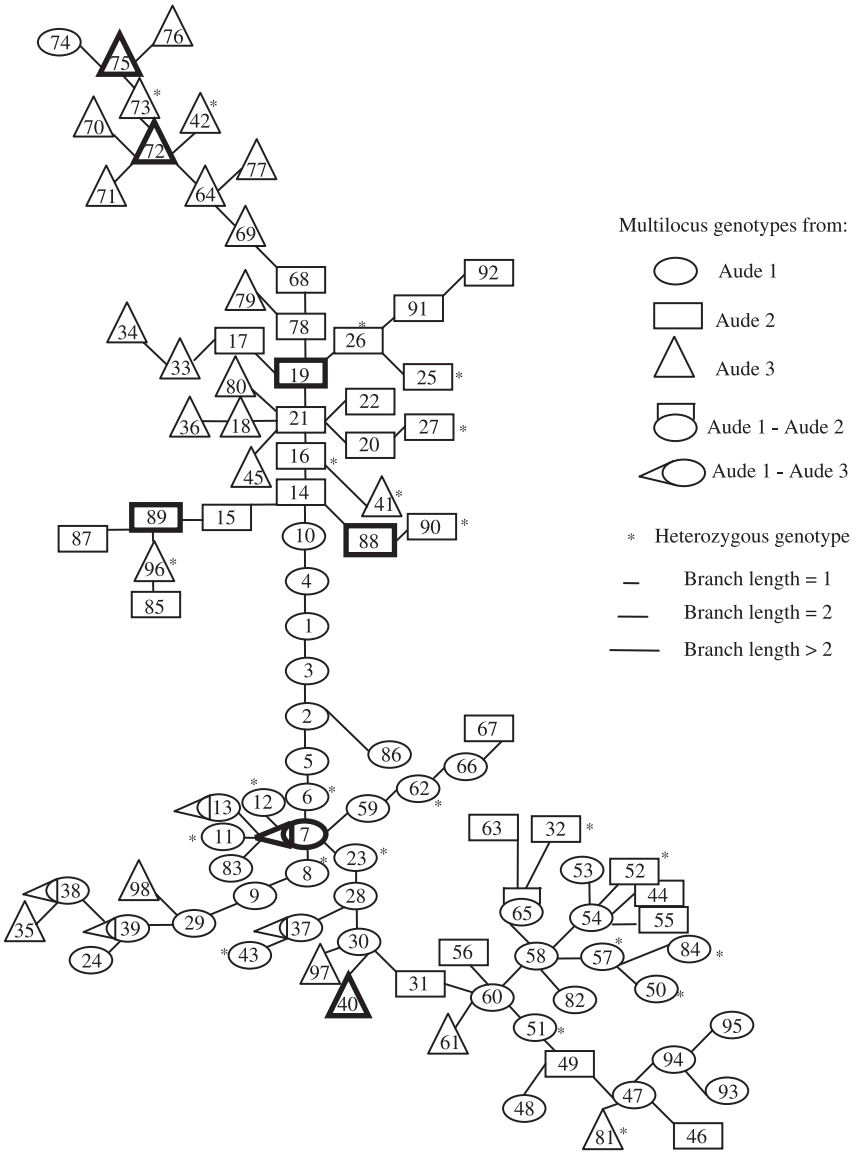


**Fig. 3** Minimum spanning network obtained from the number of pairwise differences matrix between multilocus genotypes. Thick figures represent genotypes showing frequencies higher than 10% in their own subpopulation.

Aude 1, four from Aude 2, and seven from Aude 3. Separately from the results obtained using the network, Aude 1 did not appear as the main source of migrants, but rather as a sink population. It is likely that the migrants which were not revealed by the network analysis in Aude 1, such as genotypes 24, 30, 37, 38, and 39 which are assigned to Aude 3 in Fig. 7, are issued from pollen migration from this subpopulation. Probably due to subsequent reproductive events in Aude 1, they appeared as more genetically related to genotypes from Aude 1 than genotypes from Aude 3 in the network analysis (Fig. 3).

## Discussion

As shown in 1990 using RAPD markers (Bonnin *et al*. 1996a), the highly selfing population of *Medicago truncatula* referred to as the 'Aude' population still displayed a high level of genetic variability and spatial structure for microsatellite loci in 1996 (Table 3). Multilocus genotype diversity was very similar between the two studies. The number of microsatellite loci used was, however, not sufficient to obtain an asymptotic number of multilocus genotypes, i.e. to reveal all existent lines (data not shown), in contrast with the RAPD study (Bonnin *et al*. 1996a). We may conclude, therefore, that population diversity in 1996 has probably been underestimated and would have proved larger if more loci had been used. In 1990, we probably revealed all existing multilocus genotypes for RAPD's. However, it could well be that microsatellites would have revealed more diversity. We thus cannot compare the diversity in 1990 to that in 1996 with our present data set. As the 1996 sample was exhaustive, we would expect to find more

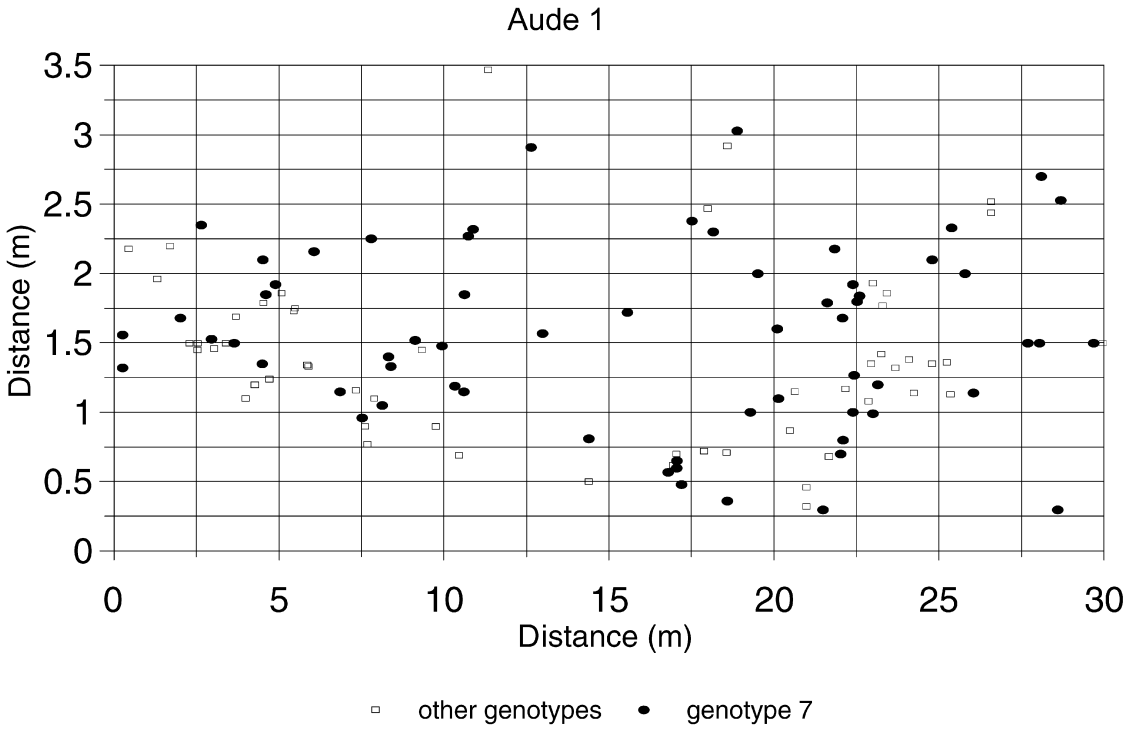


**Fig. 4** Spatial localization of the multilocus genotype 7 in subpopulation 1. Each full circle referred to an individual of genotype 7 while empty square represents sampled individuals of another multilocus genotype.

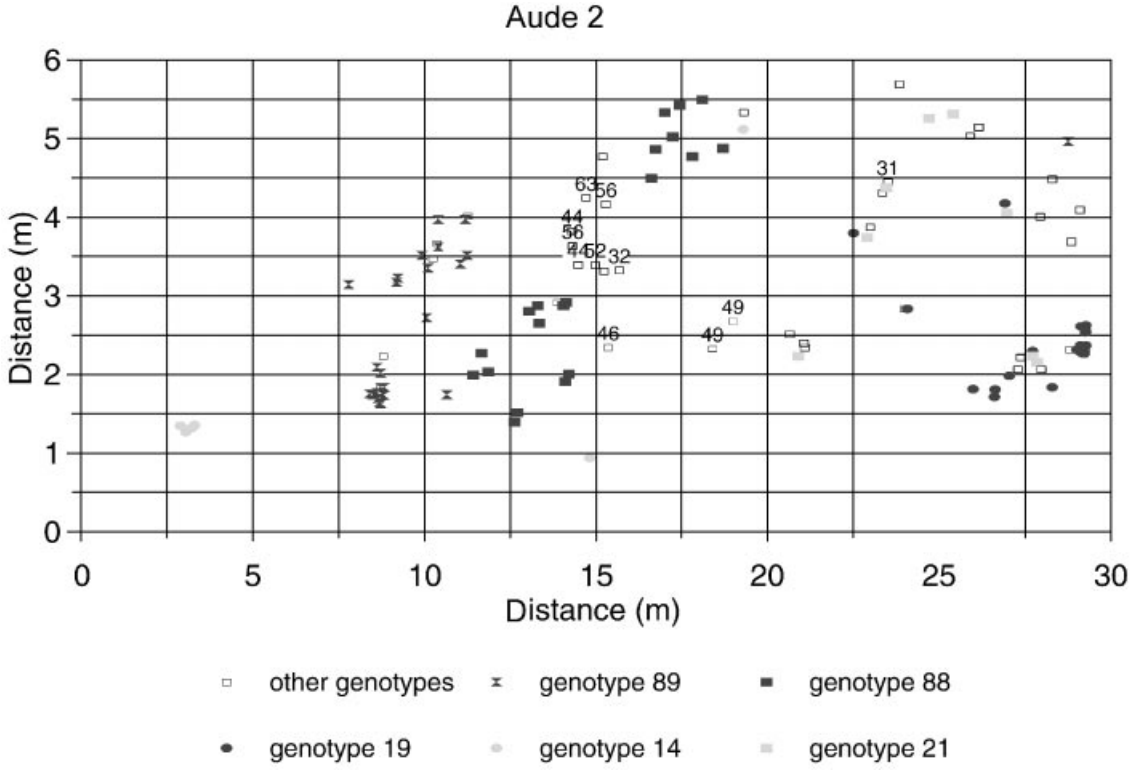


**Fig. 5** Spatial localization of multilocus genotypes in subpopulation 2. Each sampled individual is indicated by its multilocus genotype number.

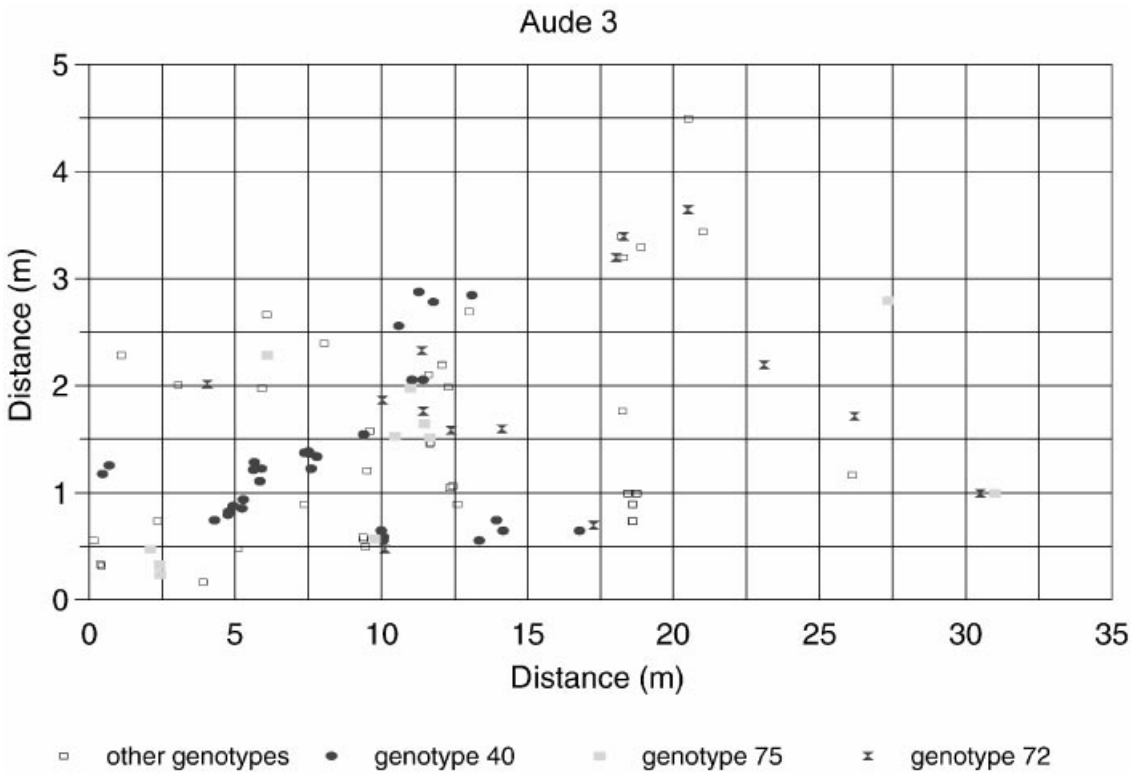


**Fig. 6** Spatial localization of multilocus genotypes in subpopulation 3. Each sampled individual is indicated by its multilocus genotype number.

**Table 3** Comparisons of the genetic diversity (Simpson's indices) and population structure (Global and pairwise $F_{ST}$) revealed in 1990 vs. 1996 using RAPD and microsatellites, respectively

| | 1990 | 1996 |
|---|---|---|
| **Multilocus genetic diversity** | | |
| Aude 1 | 0.89 | 0.71 |
| Aude 2 | 0.84 | 0.89 |
| Aude 3 | 0.85 | 0.86 |
| **Global $F_{ST}$** | **0.32** | **0.37** |
| **Pairwise $F_{ST}$'s** | | |
| A1–A2 | 0.39 | 0.41 |
| A1–A3 | 0.31 | 0.42 |
| A2–A3 | 0.28 | 0.25 |

diversity in 1996 if the same markers were used. However, as population size is rather small, we would also expect the genetic diversity to decrease under the influence of genetic drift. Conversely, population structure is quite robust to individual loci diversity (but see Fréville *et al.* 2001). $F_{ST}$ estimates from 1990 were very close to those observed in 1996 (Table 3). The less differentiated subpopulations were already the more distant ones: Aude 2 and Aude 3. Therefore, the use of dominant markers such as RAPD seems to be very suitable to analyse spatial structure in selfing species. In such species, heterozygous genotypes are so infrequent that allelic frequencies can be directly estimated from the frequencies of dominant phenotypes. However, the multilocus genotype network drawn in 1996 showed more migration/recombination events than in 1990, and differently from 1990, subpopulation 1 appeared as a source of migrants for the two other subpopulations, at least using the network analysis.

To explain the large genetic diversity we found in 1990, we suggested different mechanisms (Bonnin *et al*. 1996a): small neighbourhood sizes within subpopulations, more frequent outcrossing events than classically thought in this species and dispersal in time *via* the seed bank. These hypotheses still hold six years later. In 1990, it was not possible to estimate the population structure at the subpopulation level, since individuals were collected alongside transects and studied by dominant markers after having been maintained by spontaneous selfing in the greenhouse. In the present study, all the sampled individuals were analysed using codominant markers and without any generation of artificial selfing. Only 30 individuals were found to be heterozygous out of 365 analysed (25 heterozygous multilocus genotypes out of 98). Heterozygote deficiencies were very high, leading to selfing rates comprised between 95% and 97% depending on the subpopulation, in agreement with values found in other

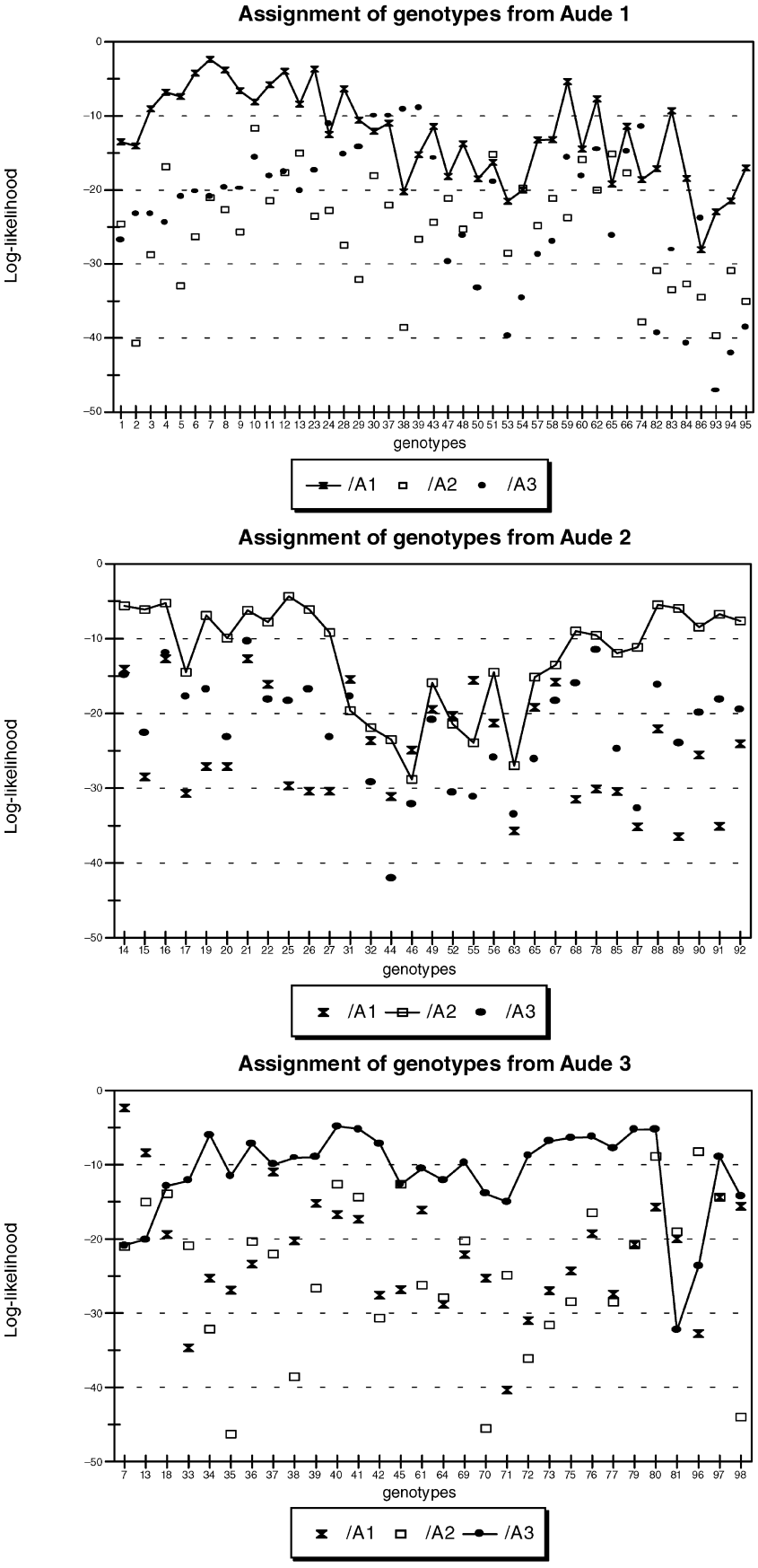


**Fig. 7** Assignment of multilocus genotypes in each subpopulation. Log-likelihoods of genotypes into their own subpopulations are shown by lines.

populations of *M. truncatula* (Chaulet & Prosperi 1994). Such large selfing rates could however, be entirely explained by subpopulation structure (that is, $F_{IS}$'s would be $F_{ST}$'s). Only the study of maternal families will allow us to determine that part of inbreeding which is due to selfing and not localized gene flow.

In plants, autocorrelation spatial analyses have been mainly used in outcrossing species to detect short distance seed or pollen dispersal events (Berg & Hamrick 1995; Loiselle *et al*. 1995; Doligez & Joly 1997; Streiff *et al*. 1998; Gehring & Delph 1999; Williams & Waser 1999). They have rarely been applied to selfing species, where limited pollen dispersal and strong hitchhiking effects are expected, leading to patchy spatial distribution. Such a patchy spatial distribution could retain more genetic diversity than expected in a nonsubdivided and highly inbreeding population (and hence of reduced effective population sizes). Our results suggest that this mechanism is likely to occur in the population studied here but only in one of the three subpopulations; different (or supplementary) mechanisms have to be invoked to explain the maintenance of a large diversity in the two other subpopulations.

Indeed, a high level of structure was detected only in Aude 2 (a low one in Aude 3, and none in Aude 1) which is also the less disturbed subpopulation. Selection could thus act at the microhabitat level in subpopulation 2, favouring different genotypes in different parts of the site. However, homogeneous selection pressures were previously suggested to occur among subpopulations from Aude so that local adaptation is unlikely to be responsible for this patchy spatial distribution (Bonnin *et al*. 1996b). In the absence of perturbation, rare migration/recombination events can structure the population into several independent recombinant lines, such as the populations of a metapopulation, allowing a high level of genetic diversity to be maintained.

In contrast, in Aude 1, which is not structured, other mechanisms must be called upon to explain the high level of diversity observed. It is likely that this subpopulation is not spatially structured because it is ploughed every year. Although seed dispersal is likely to occur mainly by animals (as pods are spiny) ploughing may facilitate the subsequent dispersal of genotypes in this population. Half of the heterozygous individuals were found in this subpopulation (15/30), which showed also the lowest proportion of linkage disequilibria (one locus being in complete linkage equilibrium with other loci). Moreover, in Aude 1, several multilocus genotypes could be easily created by recombination between two randomly chosen genotypes. For example, the multilocus genotypes 7, 12, 9, 28, 29, 37, 39, 23 and 24 can be considered as recombinant lines issued from an outcrossing event between the genotype 7 and the genotype 24 followed by several generations of self-fertilization. Outcrossing events due to pollen migration into Aude 1 were also suggested by combining the results of assignment tests and network analysis. These results suggest a higher level of recombination in subpopulation 1. Different ecological factors such as plant density, flower number, and pollinator movements are known to influence the outcrossing rate, especially for plants pollinated by animals (Barret *et al*. 1994; Karron *et al*. 1995; Franceschinelli & Bawa 2000). In the population studied here, the density of plants was higher in Aude 1 (1.3 individuals/$m^2$) than in Aude 2 and Aude 3 (0.7 individuals/$m^2$), which may facilitate pollen dispersal events. Unfortunately we do not have any information about the ecology of this process in *Medicago truncatula*. In such a selfing species, pollen dispersal is a rare event in comparison to outcrossing species, but this event has important consequences as it generates large amounts of genetic variability *via* recombinant lines. Moreover, if recombining individuals have a higher fitness, outcrossing rate might be enhanced in the following generation if it is somehow heritable. An alternative explanation involves seed dormancy. Hardseedness is known in commercial lines of *M. truncatula*. Higher seed dormancy may have been selected for by disturbance in subpopulation 1. By storing several generations, the seed bank would be more efficient to maintain diversity by reintroducing genotypes that were successful under past selective regimes (Cabin 1996; McCue & Holtsford 1998; Mahy *et al*. 1999). A temporal study using the same genetic markers for different years between 1990 and 1996 is in progress. It will show if some genotypes are more successful than others whatever the year, or if they appear randomly between years.

This study is the first published report of outcrossing rates in *M. truncatula*, using highly polymorphic markers. These estimates are nevertheless based on $F_{IS}$ values, i.e. it is assumed that the subpopulations are at equilibrium for a fixed selfing rate, and both biparental inbreeding and selfing occur. Temporal variations in outcrossing rates can be estimated by looking at multilocus individual heterozygosity patterns as recently suggested by Enjalbert & David (2000). However, as for selfing rates based on $F_{IS}$ values, selfing rates based on this method are likely to be especially biased in subdivided populations (Enjalbert & David, personal communication). In the future, it will be necessary to use maternal progenies and likelihood methods (Ritland 1983; Ritland & El-Kassaby 1985) to avoid these potential biaises. The fine scale spatial structure analysis presented here suggests that various mechanisms, such as seed bank and outcrossing events (Aude 1) or small neighbourhood sizes (Aude 2), may contribute jointly to the maintainance of diversity at a very fine spatial scale in a selfing species. The subpopulation 3 provides an intermediate situation between Aude 1 and Aude 2, being more disturbed than Aude 2, but exhibiting a higher spatial structure than Aude 1. This process associated to immigration from other subpopulations could explain the level of

diversity observed in Aude 3. To our knowledge, no study at such spatial scale has been previously conducted in a selfing species. More studies of this type are necessary to better understand the evolutionary potential of inbreeding populations as the conservation of rare species and genetic resources management are of major interests. *M. truncatula* is now the main model organism for nitrogen-fixing species (Cook 1999), there is thus a strong interest in basic data on its reproductive biology. If plant and nitrogen fixing *Bradyrhizobium* sp. co-evolve, any attempts to determine the spatial genetic structure of bacteria should rely on a good knowledge of the host.

## Acknowledgements

The authors thank Renaud Vitalis and Olivier Hardy for helpful comments on the manuscript, Sylvie Roussel, Marie-Christine Quillet, Denis Tauzin, and Marin Vabre for sampling assistance and Frédérique Viard for wisely advicing the use of microsatellites markers. This is publication ISEM 2000–117 of the Institut des Sciences de l'Evolution, Montpellier.

## References

Awadalla P, Ritland K (1997) Microsatellite variation and evolution in the Mimulus guttatus species complex with contrasting mating systems. *Molecular Biology and Evolution*, **14**, 1023–1034.

Baquerizo E, Desplanque B, Santoni S (2001) Isolation and characterization of microsatellites in *Medicago truncatula* (Leguminosae). *Molecular Ecology Notes*, in press.

Barret SCH, Harder LD, Cole WW (1994) Effects of flower number and position on self-fertilization in experimental populations of *Eichornia paniculata*. *Functional Ecology*, **8**, 526–535.

Barton NH, Whitlock MC (1997) The evolution of metapopulations. In: *Metapopulation Biology* (eds Hanski IA, Gilpin ME), pp. 183–210, Academic Press, San Diego.

Berg EE, Hamrick JL (1995) Fine-scale genetic structure of a turkey oak forest. *Evolution*, **49**, 110–120.

Bonnin I, Huguet T, Gherardi M, Prosperi JM, Olivieri I (1996a) High level of polymorphism and spatial structure in a selfing plant species, *Medicago truncatula* (Leguminosae), shown using RAPD markers. *American Journal of Botany*, **83**, 843–855.

Bonnin I, Prosperi JM, Olivieri I (1996b) Genetic markers and quantitative genetic variation in *Medicago truncatula* (Leguminosae): a comparative analysis of population structure. *Genetics*, **143**, 1795–1805.

Brown AHD, Schoen DJ (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences of the USA*, **88**, 4494–4497.

Cabin RJ (1996) Genetic comparisons of seed bank and seedling populations of a perennial desert mustard, *Lesquerella fendleri*. *Evolution*, **50**, 1830–1841.

Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

Chaulet E, Prosperi JM (1994) Genetic diversity of a collection of *Medicago truncatula* Gaertn. from Algeria. In: *Proceedings of the Genetic Resources Section Meeting of Eucarpia*, pp. 255–257. Eucarpia, Clermont Ferrand, France.

Cook DR (1999) *Medicago truncatula*, a model in the making! *Current Opinion in Plant Biology*, **2**, 301–304.

Doligez A, Joly H (1997) Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana. *Heredity*, **79**, 72–82.

Enjalbert J, David JL (2000) Inferring recent outcrossing rates using multilocus individual heterozygosity (MIH): application to evolving wheat populations. *Genetics*, **156** (4), 1973–1982.

Franceschinelli EV, Bawa KS (2000) The effect of ecological factors on the mating system of a South American shrub species (*Helicteres brevispira*). *Heredity*, **84**, 116–123.

Fréville H, Justy F, Olivieri I (2001) Comparative allozyme and microsatellite population structure in a narrow endemic plant species Centaurea corymbosa Pourret (Asteraceae). *Molecular Ecology*, **10**, 879–889.

Gehring JL, Delph LF (1999) Fine-scale genetic structure and clinal variation in *Silene acaulis* despite high gene flow. *Heredity*, **82**, 628–637.

Goodman SJ (1997) Rst calculation: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their difference. *Molecular Ecology*, **6**, 881–885.

Goudet J (1995) Fstat (Version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity*, **86**, 485–486.

Hamrick JL, Godt MJW (1990) Allozyme diversity in plant species. In: *Plant Population Genetics, Breeding and Genetic Resources* (eds Brown AHD, Clegg MT, Kahter AL, Weir BS), pp. 43–62. Sinauer Associates Inc., Sunderland, Massachussets.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London*, **351**, 1291–1298.

Hardy O, Vekemans X (1999) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, **83**, 145–154.

Heywood JS (1991) Spatial analysis of genetic variation in plant populations. *Annual Review of Ecology and Systematics*, **22**, 335–355.

Jarne P, Lagoda JL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424–429.

Karron JD, Thumser NN, Tucker R, Hessenauer AJ (1995) The influence of population density on outcrossing rates in *Mimulus ringens*. *Heredity*, **75**, 175–180.

Kuittinen H, Mattila A, Savolainen O (1997) Genetic variation at marker loci and in quantitative traits in natural populations of *Arabidopsis thaliana*. *Heredity*, **79**, 144–152.

Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany*, **82**, 1420–1425.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65–95.

Mahy G, Vekemans X, Jacquemart A (1999) Patterns of allozymic variation within *Calluna vulgaris* populations at seed bank and adult stages. *Heredity*, **82**, 432–440.

McCue KA, Holtsford TP (1998) Seed bank influences on genetic diversity in the rare annual *Clarkia springvillensis* (Onagraceae). *American Journal of Botany*, **85**, 30–36.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.

Nordborg M, Donnelly P (1997) The coalescent process with selfing. *Genetics*, **146**, 1185–1195.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics*, **147**, 1943–1957.

Peet (1974) The measurement of species diversity. *Annual Review of Ecology and Systematics*, **5**, 285–307.

Pielou EC (1969) *An Introduction to Mathematical Ecology*. New York, Wiley-Interscience.

Raymond M, Rousset F (1995) GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Ritland K (1983) Estimation of mating systems. In: *Isozymes in Plant Genetics and Breeding* (eds Tanksley SD, Orton TJ), pp. 289–302. Elsevier Science Publishers, Amsterdam.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, **67**, 175–185.

Ritland K, El-Kassaby YA (1985) The nature of inbreeding in a seed orchad of Douglas fir as shown by an efficient multilocus model. *Theoretical and Applied Genetics*, **71**, 375–384.

Rousset F (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin, Version 2000: a software for population genetics data analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites. *Molecular Ecology*, **7**, 317–328.

Sweigart A, Karoly K, Joness A, Willis JH (1999) The distribution of individual inbreeding coefficients and pairwise relatedness in a population of *Mimulus guttatus*. *Heredity*, **83**, 625–632.

Waser PM, Strobeck C (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*. 43–44.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Williams CF, Waser NM (1999) Spatial genetic structure of *Delphinium nuttallianum* populations: inferences about gene flow. *Heredity*, **83**, 541–550.

---

This work is part of a general study on the evolutionary potential of inbreeding populations. All authors are interested in microevolutionary processes that govern genetic variation in subdivided populations and life history evolution.

---