Molecular Ecology (2005) **14**, 1641–1653 doi: 10.1111/j.1365-294X.2005.02537.x

Inferring contemporary levels of gene flow and demographic history in a local population of the leaf beetle *Gonioctena olivacea* from mitochondrial DNA sequence variation

PATRICK MARDULYN and MICHEL C. MILINKOVITCH
Laboratory of Evolutionary Genetics, Université Libre de Bruxelles, Rue Jeener et Brachet 12, 6041 Gosselies, Belgium

Abstract

We have studied mitochondrial DNA variation in a local population of the leaf beetle species *Gonioctena olivacea*, to check whether its apparent low dispersal behaviour affects its pattern of genetic variation at a small geographical scale. We have sampled 10 populations of *G. olivacea* within a rectangle of 5 × 2 km in the Belgian Ardennes, as well as five populations located approximately along a straight line of 30 km and separated by distances of 3–12 km. For each sampled individual (8–19 per population), a fragment of the mtDNA control region was polymerase chain reaction-amplified and sequenced. Sequence data were analysed to test whether significant genetic differentiation could be detected among populations separated by such relatively short distances. The reconstructed genealogy of the mitochondrial haplotypes was also used to investigate the demographic history of these populations. Computer simulations of the evolution of populations were conducted to assess the minimum amount of gene flow that is necessary to explain the observed pattern of variation in the samples. Results show that migration among populations included in the rectangle of 5 × 2 km is substantial, and probably involves the occurrence of dispersal flights. This appears difficult to reconcile with the results of a previous ecological field study that concluded that most of this species dispersal occurs by walking. While sufficient migration to homogenize genetic diversity occurs among populations separated by distances of a few hundred metres to a few kilometres, distances greater than 5 km results in contrast in strong differentiation among populations, suggesting that migration is drastically reduced on such distances. Finally, the results of coalescent simulations suggest that the star-like genealogy inferred from the mtDNA sequence data is fully compatible with a past demographic expansion. However, a metapopulation structure alone (without the need to invoke a population expansion event) cannot be dismissed as the cause of this star shape.

Keywords: Chrysomelidae, control region, demographic history, gene flow, *Gonioctena olivacea*, mitochondrial DNA, population differentiation

Received 16 October 2004; revision received 17 January 2005; accepted 2 February 2005

Introduction

Like most chrysomelid beetles, insects from the genus *Gonioctena* are specialist herbivores. That is, their diet is restricted to a few host-plant species. Both larvae and adults feed on the same host plant. Although adults are capable of flying, they are rarely seen doing so (sometimes to reach the host canopy; Richards & Waloff 1961; Mason & Lawson 1982), and their power of dispersal hence appears limited. Several allozyme studies have revealed high level of differentiation among leaf beetle populations, including species from *Gonioctena* (McCauley *et al*. 1988; Rank 1992; Knoll *et al*. 1996; Knoll & Rowell-Rahier 1998), which is probably explained by the combination of this limited dispersal behaviour with the patchy distribution of their host plant. More recently, a phylogeographical study of a regional population (Vosges Mountains) of *Gonioctena*

Correspondence: Patrick Mardulyn, Fax: 32-71-378950; E-mail: pmarduly@ulb.ac.be

pallida showed, from the analysis of mitochondrial DNA (mtDNA) variation, that the presence of several effective barriers prevents recurrent gene flow within the studied area (Mardulyn 2001).

In the present study, we focus on the pattern of mtDNA variation in a local population of another species, *Gonioctena olivacea* (Forster). As is the case for most leaf beetles, populations of this species are structured, as their distribution in the field follows the patchy distribution of their host plant. Although *G. olivacea* is associated with a few Fabaceae plant species (belonging for example to the genera *Cytisus* and *Genista*), in the studied area (Belgian Ardennes), it feeds on a single plant species, the broom (*Cytisus scoparius*). This shrub grows easily on disturbed soils and it appears often after a portion of spruce plantation has been cleared. Broom patches seldom survive more than 20 years in this environment, either because new spruces are planted before, or because native tree species colonize the habitat, that then becomes unsuitable for broom. Populations of beetles go extinct when the host plant disappears. A metapopulation model (Hanski & Simberloff 1997), in which several local populations (or demes) are connected by migration events and are subject to extinction and recolonization through dispersal (population turnover), is therefore likely to be appropriate to describe the population dynamic of this species.

The goal of this study was to check whether the apparent low dispersal behaviour of *G. olivacea* affects its pattern of genetic variation at a small geographical scale. Indeed, as part of a 5-year extensive ecological field study of one *G. olivacea* population associated with a 0.8 ha large patch of the host plant *Cytisus scoparius*, Richards & Waloff (1961) detected almost no flight by means of traps (2 years experiment) and observed very little dispersal of individuals on distances of only 100 feet (1 year experiment). They concluded that walking on small distances seems to be the main means of dispersal for these insects. If this is true, one would expect to observe significant genetic differentiation among groups of leaf beetles associated with patches of host plant separated by relatively short distances. To test this hypothesis, we have sampled 10 populations of *G. olivacea* (here, a population is defined as a group of beetles associated to a geographically well-delimited patch of host plants) within a rectangle of 5×2 km in the Belgian Ardennes, as well as five populations located approximately along a straight line of 30 km and separated by distances of 3–12 km. For each sampled individual (8–19 per population), a fragment of the mtDNA control region (CR) previously found to be highly variable for this species (Mardulyn *et al*. 2003) was amplified and sequenced using polymerase chain reaction (PCR). Sequence data were analysed to test whether significant genetic differentiation could be detected among populations separated by such relatively short distances, and to estimate the level of migration of individuals among host plant patches. Because the studied populations of *G. olivacea* are unlikely to have reached migration–drift equilibrium, it was not possible to directly infer gene flow from the level of genetic differentiation observed among them. Instead, we have conducted simulations of the evolution of leaf beetle populations in order to determine the minimum level of gene flow required to explain the genetic differentiation among populations, as well as the genetic diversity within populations, observed in the field. Finally, the reconstructed genealogy of the mitochondrial haplotypes was used to investigate the demographic history of these populations. In particular, because a metapopulation model is likely to be appropriate to describe the dynamic of the *Gonioctena olivacea* populations, we used coalescent simulations to assess whether the shape of this genealogy can be explained by a metapopulation structure alone, or if a demographic expansion needs to be invoked.

Materials and methods

Insect collection

One hundred sixteen individuals from 10 populations were sampled in August 1999 from a 5×2 km rectangle in the Belgian Ardennes. In August 2003, 75 individuals from five additional populations, separated by larger distances, were sampled in the vicinity of the samples from 1999. For the purpose of the present study, a population is defined as a group of beetles associated with a patch of broom, the only host plant species of *Gonioctena olivacea* found in this part of Europe. Due to the patchy distribution of the host plant, delimiting leaf beetles populations was relatively straightforward, although patches varied in size from ±100 to ±10 000 m^2. In general, plants separated by less than 10 m are considered belonging to a single patch. Special care was taken to sample each population as homogeneously as possible. Figure 1 shows the relative position of each sampled population, and sample details are given in Table 1.

DNA sequencing

Genomic DNA was extracted from ethanol-preserved insects using the DNeasy Tissue Kit from QIAGEN. Whole specimens where each ground in the QIAGEN ATL buffer, and incubated for 3 h with proteinase K at 55 °C. The remaining steps were performed as described in the manufacturer's protocol. For each of the 191 specimens, the entire CR (± 3.2 kb) was amplified using the Expand Long Template PCR System Kit (Roche) following the manufacturer's protocol, with an annealing temperature of 55 °C, and an extension step of 6 min at 66 °C. The primers used were located on either side of the CR, in the small subunit (12s) ribosomal RNA gene (*SR-J-14766*; 5′-CATTATTTGTATAACCGCAACTGCTGGCAC-3′; (Mardulyn *et al*. 2003), and the cytochrome oxidase 1 gene

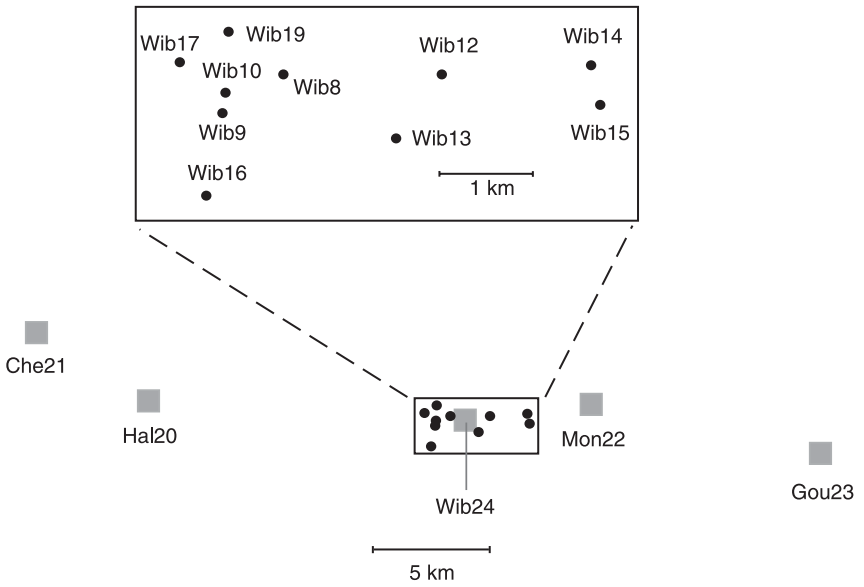

Fig. 1 Relative position of sampled populations of *Gonioctena olivacea* in the Ardennes. Black dots indicate position of populations sampled in 1999, while grey squares indicate position of populations sampled in 2003. Geographical coordinates are given in Table 1.

Table 1 Sampled populations with their geographical coordinates, number of individual sampled, year of sampling and haplotypes identified in each sampling localities

Population	Geographical coordinate	Sample size	Year of sampling	Haplotypes (no. individuals) CR2	Haplotypes (no. individuals) CR1
Wib8	50°10′51″ N, 5°42′17″ E	13	1999	1(7), 18(3), 19, 23(2)	4(5), 10, 12, 16, 20, 21, 22, 23, 24
Wib9	50°10′39″ N, 5°41′43″ E	10	1999	1(7), 2, 15, 18	4(6), 25, 27, 28, 33
Wib10	50°10′44″ N, 5°41′44″ E	8	1999	1(6), 18(2)	4(3), 23, 37, 38, 40, 41
Wib12	50°10′50″ N, 5°43′44″ E	11	1999	1(7), 6, 13, 23(2)	4(3), 20, 21, 81, 83, 84(2), 86, 87
Wib13	50°10′31″ N, 5°43′20″ E	14	1999	1(8), 3, 4, 18, 23, 34(2)	4(5), 20, 33, 58, 59, 60, 61(2), 63, 69
Wib14	50°10′53″ N, 5°45′05″ E	11	1999	1(4), 5, 14, 18, 22, 28(3)	4(4), 70, 73(3), 75, 77, 80
Wib15	50°10′41″ N, 5°45′11″ E	13	1999	1(6), 18(3), 23, 30, 32(2)	4(6), 20, 33(2), 43, 49, 52, 54
Wib16	50°10′15″ N, 5°41′35″ E	13	1999	1(7), 7(2), 8, 10, 23, 26	4(6), 21, 106(2), 108, 111, 112, 114
Wib17	50°10′54″ N, 5°41′20″ E	12	1999	1(6), 21, 23(3), 29, 33	4(5), 20(3), 92, 99, 100, 102
Wib19	50°11′02″ N, 5°41′45″ E	11	1999	1(8), 18, 19, 32	4(7), 1, 2, 3, 10
Hal20	50°11′09″ N, 5°31′08″ E	9	2003	23(9)	
Che21	50°12′27″ N, 5°27′02″ E	13	2003	23(13)	
Mon22	50°11′05″ N, 5°47′25″ E	17	2003	1, 18(2), 24, 25(13)	
Gou23	50°10′05″ N, 5°55′50″ E	19	2003	1(6), 23, 18(4), 20(2), 27, 31, 32(4)	
Wib24	50°10′44″ N, 5°42′48″ E	17	2003	1(9), 9, 11, 12(2), 17, 18(3)	

(5′-AGCAATATTTGCAGATAGWGGGGGATAAAC-3′; this study). The entire repetitive section of the CR (15 imperfect copies of a 106 to 142 bp-long core sequence) was then sequenced for all individuals collected in 1999 using the PCR product as a template and three internal primers (5′-TAAATCTCCAAGTTGAATCG-3′; 5′-AATTTAAAATCCCGAATATTC-3′; 5′-TTGTTTTATAGGGATTTATTG-3′). This region had been found to be highly polymorphic in a previous hierarchical analysis of nucleotide sequence variation in the CR of *G. olivacea* (Mardulyn *et al*. 2003). In that study, we were able to unambiguously assess orthology among repeats making this region potentially useful for studying patterns of sequence variation at very low geographical scales. Only a portion of the repetitive section of the CR was sequenced for all 2003 collected individuals, using a single internal primer (5′-TAAATCTCCAAGTTGAATCG-3′). Finally, a portion of about 1250 nucleotides of the cytochrome c oxidase I (*COI*) gene was PCR-amplified (one step of 30 s at 94 °C followed by 35 cycles of 60 s at 94 °C, 60 s at 52 °C, and 2 min at 72 °C, and a final extension of 2 min at 72 °C) and sequenced using the insect universal primers *C1-J-1751* (modified version: 5′-GGAGCTCCTGATATAGCWTTYCC-3′) and *TL2-N-3014* (Simon *et al*. 1994), for 58 individuals. This was carried out to compare the rate of evolution of the CR repetitive section to that of the widely used *COI* mitochondrial gene. Sequencing reactions were performed using

the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Sequencing products were separated by electrophoresis on an ABI 377 or AB 3730 automated DNA sequencer.

Data analyses

Sequences were edited and aligned with the software SE-AL version 2.0a11 (Rambaut 1996). A haplotype network was constructed using the program TCS version 1.13 (Clement *et al*. 2000), for two data sets: the first including the entire repetitive section of the CR for the individuals sampled in 1999, and the second including only a portion of the repetitive section of all individuals sampled in 1999 and 2003. Differentiation among populations was assessed by an analysis of molecular variance (AMOVA; Excoffier *et al*. 1992) as well as by calculating pairwise F_{ST} between populations (conventional F_{ST} from haplotype frequencies as well as F_{ST} calculated from *p* distances), using the program ARLEQUIN (Schneider *et al*. 2000). Statistical significance of F_{ST} values was evaluated, as implemented in this program, by their comparison to a null distribution of pairwise F_{ST} values generated under the hypothesis of no difference between populations (by permuting haplotypes between populations). To evaluate genetic variation within populations, gene diversity (*sensu* Nei 1987) (i.e., expected heterozygosity for diploid data), corrected for sample size, was calculated within each sample. A Mantel test was performed in ARLEQUIN on the data from the samples collected in 2003 to test for a possible correlation between pairwise geographical distances and F_{ST} values.

Average pairwise sequence divergence was calculated, with standard deviation, for the *COI* and CR sequences, for all 58 individuals for which both fragments were sequenced, using MEGA version 2.1 (Kumar *et al*. 2001). This allowed us to compare the rate of evolution of both fragments. Estimations of confidence intervals to these values were inferred by bootstrapping.

Simulations

Evolution of genetic diversity within, and genetic differentiation among, populations. In order to estimate gene flow among the populations sampled in 1999 (within the 5 × 2 km area), we have simulated the evolution of populations using the program ESP version 1.16 (Bohonak *et al*. 2001; Bohonak & Roderick 2003). This program implements a Monte Carlo-based algorithm to simulate the evolution of populations following an island model of migration and generates null distributions of different statistics via parametric bootstrapping.

The following parameters were fixed in all simulations: number of populations ($P = 50$), effective size of each population ($n = 1000$), number of generations ($g = 25$), length of the DNA sequence ($b = 1000$), number of loci ($l = 1$), and mutation rate per base pair per generation (10^{-8}). The choice of the population effective size is based on the census size estimate (10 000 individuals) of a population of *G. olivacea* studied by Richards & Waloff (1961) and covering an area of 0.8 ha, which is similar to the size of some of the host plant patches sampled in this study. The effective population size (N_e) was simply considered to be 1/10th of the census size, which is close to what is inferred in other animal species, including some insects (Frankham 1995). The choice of the number of generations is based on the ecology of the host plant: *Cytisus scoparius* is a shrub that grows to a height of about 2 metres and has a lifespan of 10–20 years. In the studied area, patches of broom seldom survive more than 20 years (see Introduction). Hence, the number of generations has been limited to 25 generations in our simulations (assuming one generation per year). The mutation rate has virtually no effect on the simulation as the number of generations is too low for allowing the occurrence of a significant number of mutations in the DNA fragment considered. At the beginning of each simulation, all populations were founded from a single source population that consisted in one haplotype with a frequency of 40%, one haplotype with a frequency of 5%, two haplotypes with a frequency of 2%, 11 haplotypes with a frequency of 1%, and 200 haplotypes with a frequency of 0.2%. The frequencies of the high frequency (= 1%) haplotypes were chosen to be similar to those found in the overall beetles sample (116 individuals). Using ESP, we empirically chose the remaining frequencies (low frequency haplotypes) such that if we sample this source population (100 replicates) in the same way we sampled the populations in the field (i.e., sampling of 10 populations for a total of 116 collected individuals), we would measure an average gene diversity (as defined previously) equal to that estimated in our real sample.

Several simulations of population evolution were performed (100 or 1000 replicates each) with different values of the proportion of migrants *m* (ranging from 0.1% to 10%) per population and per generation, and with different numbers of founding individuals per population. Two types of populations were included in each simulation: newly founded populations and 'old' populations. Newly founded populations were started using ($m \times N_e$) individuals from the source population. Old populations were created as replicates of the entire source population. Some simulations were performed with 25 of each type of populations, while others were performed with one newly founded population and 49 'old' populations. The rationale was to monitor the newly founded populations to find out whether they would show, after 25 generations and different levels of migration (*m*), similar patterns of variation than the populations sampled in the field. Two statistics were calculated to compare the simulated and real populations: F_{ST} and gene diversity (the latter corrected for

sample size). These statistics were calculated based on a sample of 8–14 individuals as in our real samples. Statistical significance was assessed (i) for F_{ST} values, by checking whether values from the real data set fall within the 95% limit of the generated null distribution and (ii) for gene diversity values, by computing an unpaired *t*-test (corrected for unequal variances, as implemented in Microsoft Excel) between the values from our 10 real samples with those of the 100 simulated samples (because these values are corrected for sample size, it was possible to compare them across samples).

Coalescent simulations. We tested whether the star-like genealogy inferred from the CR data set is compatible with a simple metapopulation model or if a population size expansion is required to explain the observed pattern. For this reason, we computed coalescent simulations of the evolution of populations using the program SIMCOAL version 2.0 (Laval & Excoffier 2004). Coalescent simulations consist in reconstructing (backward in time) the genealogy of a sample of genes drawn from one or more demes in a population. Under the assumption of selective neutrality, the probability of two given genes to coalesce at one generation is directly related to the history and demography of the populations and is independent from the mutation process. Mutations are added to the reconstructed genealogy in a second step, using a specific model of DNA sequence evolution.

We approximated the metapopulation model as follows: each simulation featured 49 populations ($N_e = 1000$) that exchange migrants following a two-dimensional stepping-stone model (populations arranged on a 7×7 grid; migration rate equal to 0.01 among neighbouring populations, and equal to 0 among non-neighbouring populations). At each generation, looking forward in time, the effective sizes of two *a priori* chosen populations (out of 49) are reduced abruptly, and gradually recover during the following seven generations (exponential growth option) their original size of 1000. These bottleneck events were defined in the input file (i.e., prior to the simulation) for approximately 5 000 generations, that is well beyond the number of generations necessary to reach the coalescence of the last two lineages. Simulations were performed 200 times for each of eight different bottleneck intensities. The following additional parameter values (estimated from the real data set) were used: sequence length of 1100, transitions/transversion (ti/tv) rate of 0.909, 116 sequences sampled evenly from all populations (the current version of SIMCOAL 2.0 does not allow defining a sample size of zero, and it was therefore not possible to reproduce exactly the sampling scheme conducted in our study). The mutation rate varied between 2.2×10^{-7} and 0.0024 and was adjusted in each simulation to allow the mean number (across all replicates) of mutations to be roughly equal to the number of mutations observed in the real data set. The resulting data sets were analysed using ARLEQUIN to calculate Tajima's *D* statistic (Tajima 1989), i.e. a standardized measure of the difference between the number of segregating sites and the average pairwise difference between sequences, that takes significant negative values in the case of a star-like genealogy (e.g. Slatkin & Hudson 1991). Tajima's *D* is easily and quickly computed by ARLEQUIN on a high number of data sets and was therefore preferred to a visual inspection of each inferred genealogy.

The model used in the coalescent simulations is not exactly a metapopulation model. In a classic metapopulation model, some populations (or demes) go extinct, and new populations are founded elsewhere from a small number of individuals. In the studied *G. olivacea* populations, host-plant patches disappear after less than 25 years, and new patches are colonized by migrating individuals. In our simulations, we have approximated this process by considering that each population goes through a bottleneck every 25 generations. That is, we substituted the extinction–colonization process by linking the extinction of one population to the founding of a new population with founder individuals coming from the extinct population. It is then closer to a propagule-pool metapopulation model (Pannell & Charlesworth 2000), in which new demes are founded by individuals coming from a single source population.

The results of these simulations were also compared to other coalescent simulations, performed also with SIMCOAL 2.0, featuring a geographical (and size) expansion. In this case, we have used the same parameters as previously mentioned, but with no historical bottlenecks implemented. In contrast, a geographical expansion (going forward in time) was created from a single population: each 10 generations (at generation 1000, 1010, and 1020), 9, then 15, and finally 24 new populations were created. The same stepping-stone model as described previously was used to describe migration of individuals. Tajima's *D* values were computed as explained. It was therefore possible to compare the effect of a constant size metapopulation to the effect of a metapopulation subject to a size expansion, on the structure of the inferred genealogy.

Results

Sequence variation

Sequences of the entire repetitive section of the CR were obtained for 116 individuals collected in 1999, as well as a portion of this repetitive section for an additional 75 individuals collected in 2003. The aligned sequences were pruned at both the 5′- and 3′-ends to ensure that no trailing gaps were present in the analysed data set. In addition, a large portion of ambiguously aligned sequences, which included many gaps, was deleted from the 1999 data set.

Two data sets were eventually generated: one containing the entire repetitive section for all 116 individuals from 1999 (CR1 data set), resulting, after removal of the ambiguously aligned positions, in a final alignment of 1875 nucleotides, and one including all 191 individuals from 1999 and 2003 but for a portion of the repetitive section (CR2 data set), resulting in a final alignment of 649 nucleotides. Eleven gaps are included in the CR1 data set, varying in length from 1 to 212 nucleotides. These gaps were recoded as multistate characters at the end of the data matrix. The length of the sequences in this data set, excluding gaps, varied from 1250 to 1564 nucleotides. Average base frequencies were 52.8%, 11.9%, 4.9% and 30.4% for A, C, G and T, respectively. Four gaps are included in the CR2 data set, varying in length from 1 to 18 nucleotides. The length of the sequences in this data set, excluding gaps, varied from 627 to 648 nucleotides. Average base frequencies were 52.5%, 11.8%, 5.2%, and 30.5% for A, C, G and T, respectively. We also obtained COI sequences for 58 of the 75 individuals collected in 2003. After pruning the sequences at both ends to avoid the presence of trailing gaps, the COI data set resulted in an alignment of 1176 nucleotides (no gaps). All sequences were deposited in GenBank under accession numbers AY904804–AY904891, and alignments are available at <www.ulb.ac.be/sciences/ueg/olivacea.html>. Gaps were considered as missing data in all analyses. Numbers of variable sites were 57, 32 and 6 in the CR1, CR2 and COI data sets, respectively. Merging identical sequences resulted in 48, 34 and 6 unique haplotypes for the CR1, CR2 and COI data sets, respectively. Haplotype networks were generated for both CR data sets using TCS and are shown in Figs 2 and 3. The average COI and CR pairwise sequence divergences (absolute number of differences; MEGA 2.1) among 58 individuals are 0.306 ± 0.100 and 1.990 ± 0.880, respectively. Given that the COI fragment is roughly twice the size of the CR fragment, the repetitive section of the CR evolves in average 12 times more rapidly than the COI.

Patterns of variation within and among populations

The AMOVA conducted on the CR1 data set yielded a non-significant global F_{ST} value of 0.0027. Pairwise F_{ST} values

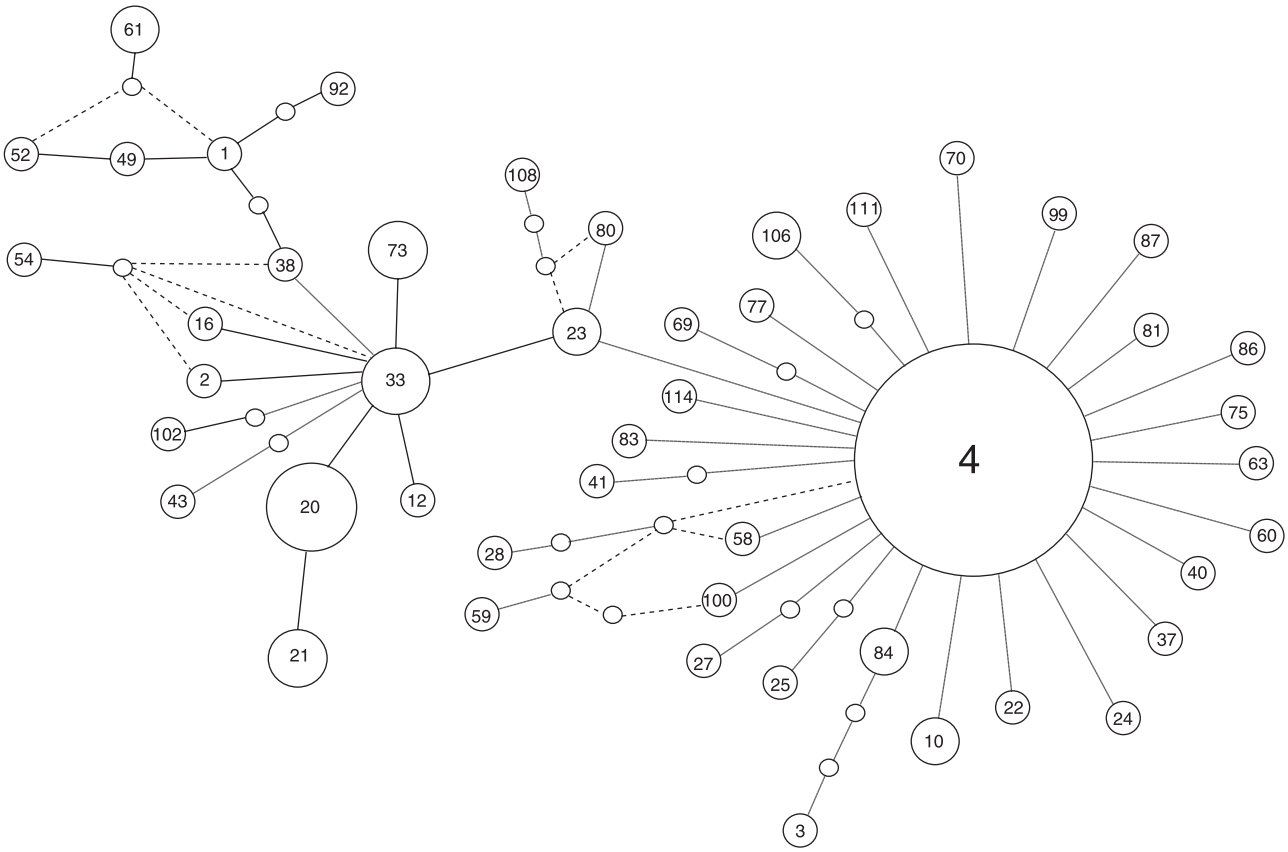

Fig. 2 Haplotype network for the mitochondrial DNA sequences (CR1) of *Gonioctena olivacea* sampled in 1999 (see black dots on Fig. 1), constructed with TCS version 1.13 (Clement *et al*. 2000). Each haplotype is represented by a circle, the surface area of which is proportional to the number of individuals bearing this particular haplotype in the sample. Each haplotype is identified by a number. Empty circles indicate intermediate haplotypes that are not present in the sample. Ambiguities (haplotypes that have alternative most parsimonious connections to the rest of the network) are represented by broken lines.

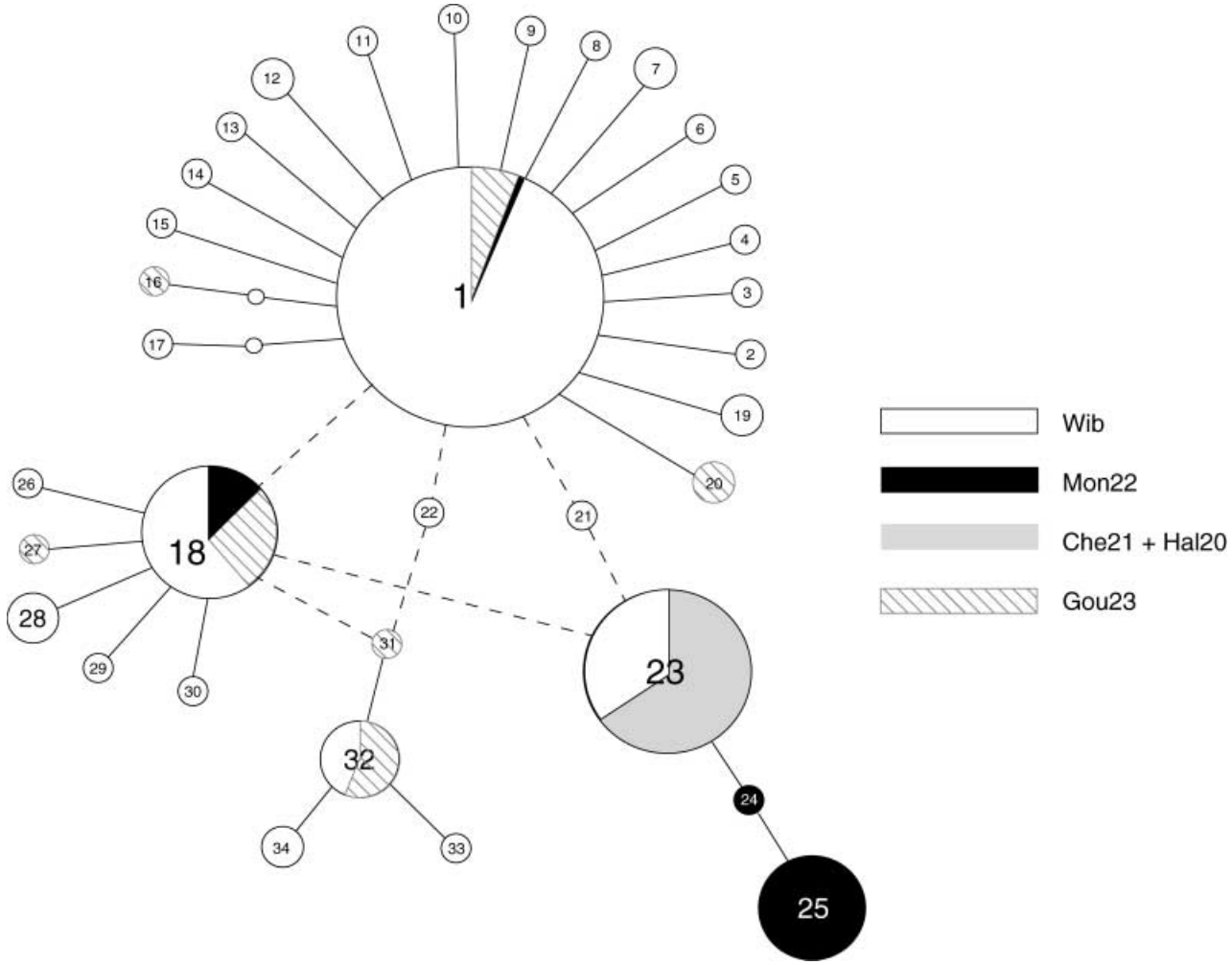

Fig. 3 Haplotype network for the mitochondrial DNA sequences (CR2) of all *Gonioctena olivacea* sampled in 1999 and 2003, constructed with TCS version 1.13 (Clement *et al*. 2000). Each haplotype is represented by a circle, the surface area of which is proportional to the number of individuals bearing this particular haplotype in the sample. Inside each circle, a specific shade of grey or bars indicate the different proportions of the individuals bearing this haplotype that were found in different populations. Empty circles indicate intermediate haplotypes that are not present in the sample. Ambiguities (haplotypes that have alternative most parsimonious connections to the rest of the network) are represented by broken lines.

calculated for the same data set, from both haplotype frequencies and *p* distances (Table 2), are small and generally not significant, confirming the result of the AMOVA. This suggests the absence of differentiation among populations sampled in 1999 in the rectangular area of 5 × 2 km. In contrast, an AMOVA conducted on the 2003 samples alone resulted in a highly significant global F_{ST} of 0.594. Most pairwise F_{ST} values calculated among these samples (Table 3) are large and significant. Clearly, genetic differentiation among populations becomes highly significant when populations are separated by geographical distances > 5 km. However, a Mantel test failed to detect any significant correlation between genetic and geographical distances calculated between populations sampled in 2003.

Gene diversity measured within each population (Table 4) is relatively high in general. All gene diversity values are above 0.6 for the CR1 data set, and most are above 0.4 for the CR2 data set. Two populations sampled in 2003, Hal20 and Che21, contain a single haplotype and therefore display a gene diversity value of 0.

Estimating gene flow among populations

Although a simple mathematical relationship between F_{ST} and N_m has been found by Wright (1931), it assumes, among others, that the studied populations are at migration–drift equilibrium (Slatkin 1985; Whitlock & McCauley 1999). Because *Gonioctena olivacea* populations are unlikely to

Table 2 Pairwise F_{ST} between populations calculated from the CR1 data set

	Wib19	Wib8	Wib9	Wib10	Wib15	Wib13	Wib14	Wib12	Wib17	Wib16
Wib19	—	−0.007	−0.019	−0.034	−0.001	−0.022	0.001	−0.023	0.030	−0.003
Wib8	0.002	—	0.030	−0.006	0.002	−0.004	0.012	−0.009	−0.021	0.020
Wib9	−0.039	−0.003	—	−0.013	0.010	−0.007	−0.001	−0.001	0.083	−0.009
Wib10	0.015	−0.030	−0.002	—	−0.023	−0.030	−0.025	−0.006	0.032	−0.014
Wib15	−0.002	−0.021	−0.034	−0.018	—	−0.001	0.005	0.008	−0.009	*0.028
Wib13	0.027	−0.015	0.012	−0.023	−0.002	—	0.010	−0.004	−0.009	0.013
Wib14	0.054	0.007	0.038	0.000	0.020	0.014	—	0.006	0.012	0.012
Wib12	0.065	−0.020	0.045	−0.014	0.008	0.000	0.021	—	0.027	−0.006
Wib17	0.032	−0.021	0.018	−0.003	−0.013	0.011	0.034	−0.001	—	*0.058
Wib16	−0.017	−0.010	−0.024	−0.004	−0.007	0.011	0.035	0.030	0.020	—

above diagonal: *p* distance F_{ST}; below diagonal: conventional F_{ST}.
*statistically significant at the 0.05 level.

Table 3 Pairwise F_{ST} calculated between populations sampled in 2003

	Hal20	Che21	Mon22	Gou23	Wib24
Hal20	—	0	*0.736	*0.490	*0.572
Che21	0	—	*0.767	*0.532	*0.614
Mon22	*0.624	*0.663	—	*0.339	*0.407
Gou23	*0.502	*0.544	*0.584	—	0.030
Wib24	*0.712	*0.745	*0.702	*0.125	—

below diagonal: *p* distance F_{ST}; above diagonal: conventional F_{ST}.
*statistically significant at the 0.001 level.

Table 4 Gene diversity calculated for each population from the CR1 and CR2 data sets

	CR2 data	CR1 data
Wib8	0.68	0.872
Wib9	0.533	0.667
Wib10	0.429	0.893
Wib12	0.6	0.927
Wib13	0.681	0.879
Wib14	0.836	0.836
Wib15	0.756	0.795
Wib16	0.718	0.718
Wib17	0.727	0.803
Wib19	0.491	0.618
Hal20	0	—
Che21	0	—
Mon22	0.419	—
Gou23	0.836	—
Wib24	0.706	—

be more than 20 years old (see Introduction), they did not have the time to reach such an equilibrium. Moreover, even for populations at equilibrium, F_{ST} is a nonlinear function of N_m and variation of F_{ST} is too small when N_m is large (> 2) to get even a rough estimate of the level of gene flow among populations (Whitlock & McCauley 1999).

Instead, we have performed simulations of the evolution of leaf beetle populations, assuming an island model of migration (but not mutation–drift equilibrium), to roughly determine the minimum amount of gene flow that is necessary to explain the pattern of variation observed in our year 1999 samples. In other words, we have determined what minimum value of m is required to obtain simulated samples displaying similar values of (i) differentiation (measured as pairwise F_{ST}) and (ii) within-population gene diversity than those found in the real samples. Indeed, if our assumptions about the lifespan of a patch of broom is correct, it seems that a substantial amount of gene flow would be required to explain the observed high level of variation within samples (Table 4), after colonization of a new patch, while keeping a low level of differentiation among patches (Table 2).

Two types of simulations were performed. In one case, 25 of the 50 populations were founded at the beginning of the simulation (see Materials and methods), while the other half were already in the state expected at the end of the simulation (i.e. showing the same pattern of genetic variation as found in the real samples). In the second case, only one population was founded, while the 49 others were in the state expected at the end of the simulation. The newly founded populations were sampled every five generations to evaluate the evolution of their gene diversity. The average F_{ST} among all pairs of samples were measured for the first type of simulations. Results of our simulations show that, if the leaf beetle populations evolve according to the implemented island model of migration, a proportion of migrants $m > 1\%$ is required to obtain the level of variation observed in our field samples. For a population effective size (N_e) of 1000, this represents 10 effective migrant individuals. Clearly, the within-population gene diversity values observed in the field require m values higher than,

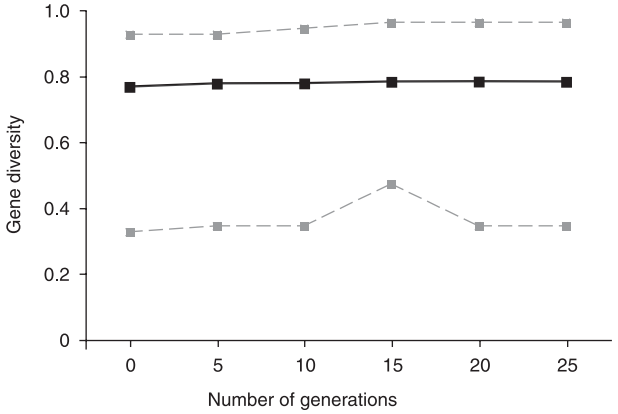

Fig. 4 Average gene diversity (black squares) of one newly created population (10 colonizing individuals) connected to 49 'old' populations by a migration parameter $m = 1\%$, as calculated in our simulations (100 replicates), between 0 and 25 generations. These values were calculated on the basis of a random sample of 11 individuals. Grey squares represent the 95% confidence interval for these values.

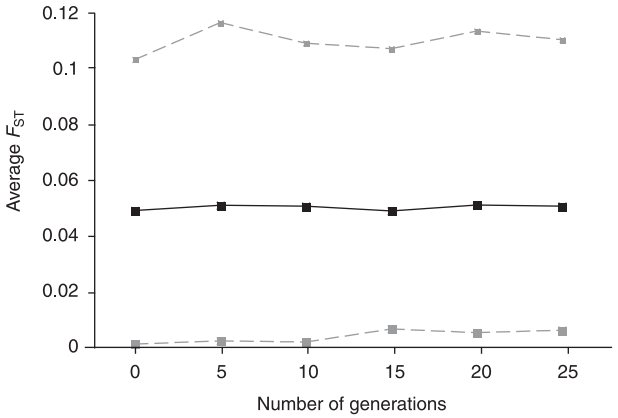

Fig. 5 Average pairwise F_{ST} (black squares) calculated for 50 populations (25 newly created populations with 10 colonizing haplotypes, 25 'old' populations) all connected by a migration parameter $m = 1\%$, for which the evolution was simulated (100 replicates), between 0 and 25 generations. These values were calculated on the basis of a random sample of 11 individuals. Grey squares represent the 95% confidence interval for these values.

but compatible with, those ($\geq 0.4\%$) required to explain F_{ST} field values. Importantly, the gene diversity value observed at the beginning of the simulation (when the population is founded) is close to the value found after 25 generations (see Fig. 4), i.e. the number of founders, rather than the number of migrants at each generation (as far as the m values tested here are concerned), is the primary determinant of the gene diversity of a population. Similarly, average F_{ST} values remain stable throughout the simulation (Fig. 5): once the population is founded (i.e. when the effective number of individuals has reached its maximum value of 1000), the impact of new migrants on genetic variation is much lower than when the population is being founded. In our simulations, the number of founders equals the number of migrants at each generation. As the field study of Richards & Waloff (1961) found no evidence of flight dispersal in *G. olivacea* populations, it is possible that migration among patches occurs sporadically (i.e. not at each generation), but that when it occurs, it involves an important number of individuals. This hypothesis would require a substantial amount of founders to explain the pattern of genetic variation observed among and within patches of broom (see Discussion).

In our simulations, an island model of migration is assumed, meaning all pairs of populations exchange the same proportion of migrants. In the real world, however, pairs of populations are more likely to exchange a higher proportion of migrants if they are geographically closer. The actual value of m is therefore likely to be different from the value inferred in our analyses. However, assuming that the real population is better described using a stepping-stone model of migration, the island-model based estimate is likely to be an underestimate of the actual migration rate (Slatkin 1985). Indeed, if the migration rate necessary to explain the pattern of genetic variation in the field (high gene diversity within populations) among geographically distant populations (i.e. separated by 3–4 km) is > 1%, this migration rate is likely to be even higher among pairs of populations separated by shorter distances (< 1 km).

Inferring demographic history

The genealogies presented in Figs 2 and 3 include a characteristic star-like pattern, in which a common haplotype, presumably ancestral, lies at the centre and is connected by independent mutation steps to many much rarer haplotypes. Such a star phylogeny is generally believed to be typical of a population that has recently expanded in size from one or a small number of founders (Slatkin & Hudson 1991; Avise 2000). Similarly, Tajima's D statistic, which was shown to become negative under a sudden population expansion model (Slatkin & Hudson 1991; Aris-Brosou & Excoffier 1996), calculated for the CR1 data set, resulted in a statistically significant value of –2.44 (in ARLEQUIN, the statistical significance of this value is inferred by comparison to values generated through coalescence simulations of a random sample under the hypothesis of selective neutrality and population equilibrium). These observations suggest that the studied population of *G. olivacea* has recently experienced an important size expansion. However, most theoretical works that have studied the impact of population expansion on the shape of genealogies or on different summary statistics have dealt with the simple case of the evolution of a single nonstructured population. The case of leaf beetle populations is quite different. As explained previously (see Introduction), a metapopulation model is

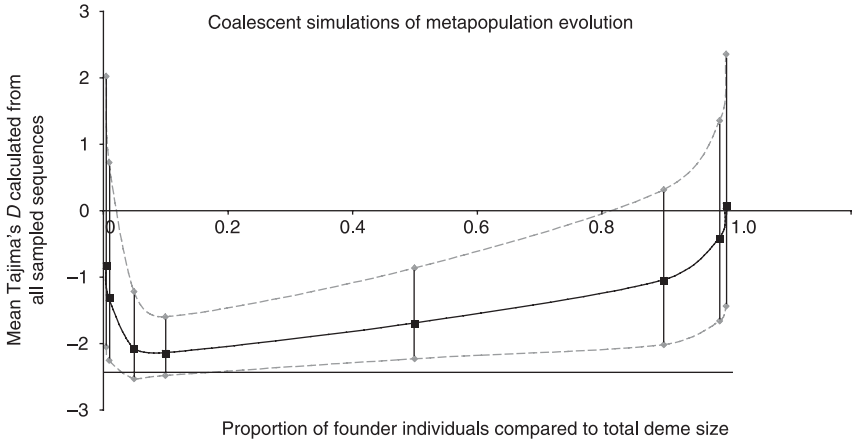

Fig. 6 Results of the simulation of the evolution of populations following a simplified metapopulation model. Mean Tajima's *D* values (averaged over 200 simulations) are shown along with 95% confidence interval, for different number of founding individuals in newly created populations (given as a proportion of founder individuals compared to the total deme size of 1000). The grey straight line drawn in the bottom indicates the position of Tajima's *D* value calculated from the real data set (−2.44).

likely to be more appropriate to describe the population dynamic of this species. Although spruce plantation appeared relatively recently in the Belgian Ardennes, compared to coalescence times, exploitation of the forest by humans is much more ancient and results in open areas that are easily colonized by broom. It then seems reasonable to assume that the metapopulation dynamic of *G. olivacea* as we see it today was similar in historical times. Therefore, to verify whether a metapopulation structure could mimic signs of population expansion (e.g. star-like genealogies), we performed coalescent simulations of the evolution of populations following an approximation of the metapopulation model. For each resulting simulated data set, Tajima's *D* statistic was computed and the generated distribution was compared to the value obtained from the real data.

Coalescent simulations of a subdivided population subject to a geographical expansion resulted in low values of Tajima's *D* (mean *D* of −2.36; range: −2.78 to −1.75), completely compatible with the value calculated from the CR1 data set (−2.44).

Figure 6 shows the mean Tajima's *D* values for the simulations performed under our simplified metapopulation model for different proportions of founder individuals, as well as the 95% limit of the generated null distributions (i.e. for each simulation, we show the mean Tajima's *D* calculated for 200 replicates and the range of values found excluding the five smaller and five larger values). Clearly, negative values of Tajima's *D* are expected under the metapopulation model used for the simulations. However, the real data value is included in the 95% limit of the generated null distributions only for a restricted portion of the *x* axis (values 0.05–0.1, corresponding to bottlenecks of 50 and 100 individuals, respectively). Nonetheless, such proportions of founders are compatible with the relatively high contemporary migration rates inferred above. Hence, we cannot reject the possibility that the star-like genealogy reconstructed from the CR1 data set is due to a metapopulation structure alone.

Discussion

The repetitive section of the mitochondrial control region was found to be roughly 12 times more polymorphic than COI, confirming the utility of the former for population genetic studies of *Gonioctena olivacea*. Using this DNA fragment, we observe extensive beetle mtDNA variation within a small area of 5 × 2 km. While gene diversity is high within each sampled population, genetic differentiation among sampled populations is extremely low. Performing computer simulations of the evolution of populations, we have estimated that the proportion of migrants necessary to explain this observed pattern of genetic variation must be above 1%. This result is difficult to reconcile with the results of the field study of Richards & Waloff (1961) concluding that most of this species dispersal occurs by walking. The host plant patches sampled in 1999 are separated by geographical distances of 400 m to 4 km. While all patches present in the 5 × 2 km studied area were obviously not sampled, each of the sampled broom patch was distant from at least several hundred metres from any other patch. Moreover, while sporadic broom plants are sometimes present between pairs of sampled populations (e.g. between Wib19, Wib10, and Wib9), and could serve as a migration bridge for beetles, other sampled populations (e.g. Wib16 or Wib17) are completely isolated from other patches because they are surrounded by spruce plantations and/or meadows, in which the host plant is completely absent. In those cases, it seems extremely unlikely that migration might have occurred only by walking. Also, if we assume that migration has occurred only among geographically close populations, we would expect to see some correlation between geographical distances and genetic differentiation. In contrast, F_{ST} measured among all pairs of populations are small and nonsignificant, even between populations separated by distances of 4 km.

If we assume that the rate of migration is more or less constant across generations, it seems unavoidable to

conclude that more than 1% of the individuals in each population are replaced by new migrants at each generation. Such a substantial migration rate could easily be explained by the occurrence of dispersal flights in these populations, possibly restricted to a small period of the season. For example, Gradojevic (1953, in Richards & Waloff 1961) reported the occurrence of a dispersal flight soon after emergence from hibernation for another species of this genus, *Gonioctena fornicata*. As noted earlier however, migration may not occur at each generation. Dispersal could occur only in specific circumstances, for example when a broom patch disappears and the beetles associated to it need to find another source of food. In these circumstances, dispersal would involve a high number of individuals and it would explain the colonization of new host plant patches by a relatively high number of individuals (> 10 effective migrants). For this hypothesis to hold, it is important that, when specific circumstances for triggering a dispersal flight are not met, no migration occurs. Indeed, if migration occurs between two massive dispersal events, but is low (i.e. < 10 effective migrants per generation), the number of individuals involved during this period in the founding of a new population will be too low to explain the observed pattern of variation in the data. As shown by our simulations, new migrants coming in after the foundation of the new population will not be able to re-establish the expected pattern of variation for this population.

While the pattern of genetic variation found within the 5×2 km studied area in 1999 suggests the absence of significant population structure, thereby providing evidence to a relatively high rate of migration among populations, a strong pattern of differentiation is observed among the more geographically distant populations sampled in 2003 (see Fig. 3). These populations are separated by distances of 5–30 km and are characterized by a global F_{ST} of 0.594 (see Results). Clearly, distances > 5 km dramatically reduces the dispersal ability of this species. However, at larger scales, no correlation is statistically detected between geographical distances and genetic differentiation. Indeed, populations Wib24 and Mon22 are separated by a distance of 5 km and by a large F_{ST} of 0.7, whereas Wib24 and Gou23 are separated by almost 15 km and a much lower F_{ST} of 0.125. Also, more than 50% of the individuals in the sample Mon22 bear a unique haplotype that is not found anywhere else, including in the Wib24 sample, located only 5 km away. No obvious barrier to migration other than geographical distance appears to prevent gene flow from or to Mon22. This suggests that no migration occurs among those populations. In sum, sufficient migration to homogenize genetic diversity occurs among populations when separated by distances of a few hundred metres to a few kilometres, but a distance greater than 5 km results in migration being too low to prevent strong differentiation to appear among populations.

Coalescent theory shows that the shape of gene genealogies can be influenced by the demographic history of populations (e.g. Hudson 1990). It is therefore possible to infer demographic history parameters from observed genealogies. Indeed, Slatkin & Hudson (1991) have clearly shown that a single population subject to a sudden size expansion is expected to be characterized by a star-shape genealogy, and a highly negative Tajima's *D*. More recent theoretical works have studied the effect of subdivided population expansion on the shape of genealogies, and in particular on Tajima's *D*. For example, Ray *et al*. (2003) showed that Tajima's *D* is significantly negative in an expanding subdivided population (2-dimensional stepping-stone model), as long as N_m is > 50. However, Ray *et al*. (2003) computed Tajima's *D* values for each deme separately while, here, we computed this statistics for the whole population. Similarly, Pannell (2003) performed coalescent simulations following two metapopulation models and studied the impact of different rates of migration and deme extinction on, among others, the shape of genealogies. Positive mean *D* values were found in his study, when calculated for sequences sampled from different demes, when extinction rates are greater than zero. However, no results (i.e. mean *D* values calculated for a batch of sequences sampled from different demes) are shown for a model in which both the migration and extinction rate are simultaneously greater than zero. In contrast, in our simulations of a simplified metapopulation model, we have assumed that both the extinction rate and the migration rate are relatively high (0.04 and 0.01, respectively), and Tajima's *D* was calculated for sequences sampled across different demes.

Our coalescent simulations of a subdivided population (2-dimensional stepping-stone model without deme extinction) show that a geographical range and size expansion is compatible with the star-like genealogy inferred from the data. Indeed, the *D* value calculated for the real data is very close to the mean *D* obtained from the simulations under this hypothesis. A geographical range expansion having occurred at some point of the history of the studied populations is therefore probable to have generated the star-like pattern observed in the genealogy shown in Fig. 2. However, one cannot reject the possibility that the metapopulation structure (likely to represent population dynamics of the studied *G. olivacea* populations) is responsible for the same pattern, without the need to invoke a population expansion event. Indeed, although our simulations of the evolution of populations following a simplified metapopulation model have resulted in a mean *D* quite different from the value obtained from the CR1 data set, more than 2.5% of individual Tajima's *D* values from the set of values obtained through simulations (with 50 or 100 effective individuals founding each newly created population) are more negative than the value calculated from the real data. Because these numbers of effective founders are compatible

with the relatively high rate of migration inferred from the data, a metapopulation model cannot be dismissed as the cause of the shape of the genealogy in Fig. 2. A subdivided population in which each subpopulation is repeatedly subject to bottlenecks will also produce negative Tajima's *D*. This clearly demonstrates that one should be cautious about directly inferring a size expansion event from a star shape genealogy. Finally, note that a star-shape genealogy can also be indicative of a selective sweep, resulting from a rapid increase in frequency of an advantageous mutation located anywhere on the mtDNA molecule (e.g. Excoffier 1990). The typing of nuclear loci will be necessary to distinguish this hypothesis from the two other strictly demographic explanations.

Acknowledgements

Interpretation of the results presented here has greatly benefited from discussions with I. Cassens. We thank two anonymous reviewers for useful suggestions on a previous version of this manuscript. This research was supported by the FNRS (Belgian National Fund for Scientific Research) and the Van Buuren Fund. P. M. is 'Research Associate' at the FNRS.

References

Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution*, **13**, 494–504.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachussetts.

Bohonak AJ, Davies N, Villablanca FX, Roderick GK (2001) Invasion genetics of New World medflies: testing alternative colonization scenarios. *Biological Invasions*, **3**, 103–111.

Bohonak AJ, Roderick GK (2003) ESP: *Evolution in simulated populations*. Computer program available at http://www.bio.sdsu.edu/pub/andy/ESP.html

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Excoffier L (1990) Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *Journal of Molecular Evolution*, **30**, 125–139.

Excoffier L, Smouse PE, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.

Gradojevic Z (1953) Ecologie et développement de *Phytodecta fornicata* Brüggm (Col., Chrysomelidae). *Travaux de l'Académie Serbe des Sciences de Belgrade*, **31**, 1–50.

Hanski I, Simberloff D (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. In: *Metapopulation Biology: Ecology, Genetics, and Evolution* (eds Hanski I, Gilpin ME), pp. 5–26. Academic Press, San Diego, California.

Hudson RR (1990) Gene genealogies and the coalescent process. In: *Oxford Surveys in Evolutionary Biology* (eds Futuyma DG, Antonovics J), pp. 1–44. Oxford University Press, Oxford.

Knoll S, Rowell-Rahier M (1998) Distribution of genetic variance and isolation by distance in two leaf beetle species: *Oreina cacaliae* and *Oreina speciosissima*. *Heredity*, **81**, 412–421.

Knoll S, Rowell-Rahier M, Mardulyn P, Pasteels JM (1996) Spatial genetic structure of leaf beetle species with special emphasis on alpine populations. In: *Chrysomelidae Biology*, vol. 1: *The Classification, Phylogeny and Genetics* (eds Cox PHA, Ja ML), pp. 379–388. SPB Academic Publishing, Amsterdam.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244–1245.

Laval G, Excoffier L (2004) SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. *Bioinformatics*, **20**, 2485–2487.

Mardulyn P (2001) Phylogeographiy in the Vosges Mountains populations of *Gonioctena pallida* (Coleoptra: Chrysomelidae): a nested clade analysis of mitochondrial DNA haplotypes. *Molecular Ecology*, **10**, 1751–1763.

Mardulyn P, Termonia A, Milinkovitch M (2003) Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera: Chrysomelidae): a hierarchical analysis of nucleotide sequence variation. *Journal of Molecular Evolution*, **56**, 38–45.

Mason ML, Lawson FA (1982) Biology of the American aspen beetle [Coleoptera: Chrysomelidae: *Gonioctena americana* (Schaeffer)] in the medicine Bow National Forest, Wyoming. *Journal of the Kansas Entomological Society*, **55**, 779–788.

McCauley DE, Wade MJ, Breden FJ, Wohltman M (1988) Spatial and temporal variation in group relatedness: evidence from the imported willow leaf beetle. *Evolution*, **42**, 184–192.

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

Pannell JR (2003) Coalescence in a metapopulation with recurrent local extinction and recolonization. *Evolution*, **57**, 949–961.

Pannell JR, Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **355**, 1851–1864.

Rambaut A (1996) SE-AL: sequence alignment editor. Available at http://evolve.zoo.ox.ac.uk/.

Rank NE (1992) A hierarchical analysis of genetic differentiation in a montane leaf beetle *Chrysomela aeneicollis* (Coleoptera: Chrysomelidae). *Evolution*, **46**, 1097–1111.

Ray N, Currat M, Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.

Richards OW, Waloff N (1961) A study of a natural population of *Phytodecta olivacea* (Forster) (Coleoptera, Chrysomelidae). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **244**, 205–257.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: *a software for population genetics data analysis, version 2.000*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematic*, **16**, 393–430.

Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Whitlock MC, McCauley DE (1999) Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4N_m + 1)$. *Heredity*, **82**, 117–125.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

PM is Research Associate of the Belgian National Fund for Scientific Research (FNRS) and Assistant Professor at the Université Libre de Bruxelles. He works in the Laboratory of Evolutionary Genetics where he conducts research in phylogeography and population genetics, using mainly leaf beetles as model organisms. MCM is professor at the Université Libre de Bruxelles (ULB), he heads the Laboratory of Evolutionary Genetics at the Institute of Molecular Biology and Medicine, and works on various aspects of conservation genetics, phylogenetics, and phyloinformatics.