Molecular Ecology (2001) **10**, 1751–1763

Phylogeography of the Vosges mountains populations of *Gonioctena pallida* (Coleoptera: Chrysomelidae): a nested clade analysis of mitochondrial DNA haplotypes

P. MARDULYN
Unit of Evolutionary Genetics, Free University of Brussels, rue Jeener et Brachet 12, 6041 Gosselies, Belgium; Laboratoire de Biologie Animale et Cellulaire, Free University of Brussels, Brussels, Belgium

Abstract

The pattern of genetic variation in the leaf beetle *Gonioctena pallida* was investigated inside the Vosges mountains using a highly variable 363 bp DNA fragment of the mitochondrial control region. Sequencing of 242 individuals, sampled in a geographical area of 100 × 40 km, identified 61 haplotypes whose genealogy was inferred. The resulting haplotype network exhibits four star-like phylogenies, two of which may be indicative of a population having recently expanded in size from a small number of founders. Nested clade analysis suggested multiple past expansion events, but also isolation by distance and possibly past fragmentation events, as the causes of the detected geographical associations of haplotypes. These results indicate the existence of effective barriers to gene flow inside the investigated area. Because the oldest demographic events inferred in the nested clade analysis were identified as expansion events, we hypothesize that a small population of founders have expanded not only in size, but also in geographical range from the south towards the north and east of the Vosges.

Keywords: Chrysomelidae, *Gonioctena pallida*, haplotype network, mitochondrial control region, nested clade analysis, phylogeography

Received 9 November 2000; revision received 19 February 2001; accepted 19 February 2001

Introduction

Most chrysomelid beetles are specialist herbivores, i.e. their diet is restricted to a few host-plant species. Both larvae and adults feed on the same host plant and their power of dispersal appears limited. Indeed, although adults are capable of flying, they are rarely seen doing so (sometimes to reach the host canopy; Richards & Waloff 1961; Mason & Lawson 1982). Walking over small distances seems to be the main means of dispersal for these insects (Richards & Waloff 1961; Mason & Lawson 1980; Knoll *et al*. 1996). Yet, given the paucity of field studies directly assessing the movement of individual leaf beetles, one cannot reject the possibility, at least in some species, of a dispersal flight occurring soon after emergence from hibernation (as reported for *Gonioctena fornicata* by Gradojevic 1953, cited in Richards & Waloff 1961).

Correspondence: Patrick Mardulyn. Unit of Evolutionary Genetics, Free University of Brussels, rue Jeener et Brachet 12, 6041 Gosselies, Belgium. Fax: 32 71 378 950; E-mail: pmarduly@ulb.ac.be

The combination of a limited dispersal behaviour in chrysomelid beetles and the patchy distribution of their host plant is assumed to be responsible for the high levels of differentiation among populations reported in several allozyme studies. For example, studying patterns of allozyme variation in the montane willow leaf beetle *Chrysomela aeneicollis*, Rank (1992) found significant genetic differentiation at different geographical levels, i.e. among beetle groups sampled in localities 0.3–40 km apart, and even among beetle groups from different trees within a given locality. Similar results were found with allozyme markers by McCauley *et al*. (1988) investigating the spatial variation of another willow leaf beetle, *Plagiodera versicolora*. High genetic differentiation was also observed in different *Oreina* and *Gonioctena* species among localities separated by distances from a few km to several hundred km (Knoll *et al*. 1996; Knoll & Rowell-Rahier 1998). In particular, fixation index (F_{ST}) values ranging from 0.1 to 0.3 were often reported among populations from different mountain ranges in Europe (Vosges, Black Forest, and different parts of the Alps).

If it is assumed that measures of genetic differentiation are indicative of levels of gene flow among populations

(Slatkin 1985), then the high F_{ST} values reported for chrysomelid beetles (among the highest reported for insects; Knoll *et al*. 1996) can be interpreted as indicators of low rates of migration. In particular, mountains appear to be effective barriers to leaf beetle migration, as their presence is often correlated with a dramatic increase in population differentiation (Knoll *et al*. 1996). However, translation of population structure estimators (such as F_{ST}) into migration rate or gene flow should be interpreted with caution (Bossart & Pashley Prowell 1998; Templeton 1998; Avise 2000). Indeed, these estimates are based on theoretical models (e.g. the island model of migration; Wright 1931) whose a priori assumptions are not necessarily met. An important assumption that is likely to be violated is that observed genetic differentiation is due to recurrent, but restricted, gene flow among populations. Indeed, similar distributions of genetic diversity can be produced through historical demographic events (population fragmentation or range expansion) even with a lack of recurrent gene flow (Templeton *et al*. 1995; Templeton 1998). For example, multiple populations produced through fragmentation of a single parental population can exhibit high to moderate levels of similarity (even if sister populations have been fully isolated since the fragmentation event) due to sharing of ancestral character states. If an island model of migration is accepted, these similarities can be inappropriately interpreted as evidence of gene flow between subpopulations, even though no gene flow actually occurs.

Analysis of DNA sequences, in contrast to allozymes or microsatellite markers, allows the estimation of evolutionary relationships among the different alleles of one locus. This historical information can then be combined with the observed geographical distribution of haplotypes to investigate the demographic factors that have produced the observed patterns (phylogeographic analysis, Avise 2000). This approach allows one to discriminate between recurrent gene flow and historical events (population fragmentation or population expansion) to explain the current distribution of genetic diversity in a population or species (Templeton 1998).

In this study, we investigate the geographical distribution of mitochondrial DNA (mtDNA) haplotypes of *Gonioctena pallida* across the Vosges mountains (12 populations sampled within an area of 100×40 km). Two populations were also sampled in the nearby Black Forest mountains, separated from the Vosges by the Rhine valley, a 30–40 km-wide lowland zone dominated by fields and urban areas. This leaf beetle species is known to feed on different willow species (*Salix*) and on hazel (*Corylus avellana*) (Axelsson *et al*. 1974; Mardulyn *et al*. 1997; this study). It is distributed over Northern Europe, but is also present in Middle Europe where it is restricted to mountain habitats at elevations of 500–2000 m (Kippenberg 1966; Cantonnet 1968; this study). Some of the host-plant species are present in the Rhine valley, but *G. pallida* seems absent from this region, presumably because of its low elevation.

The objectives of our analyses were threefold. First, we wanted to test whether there is sufficient genetic variation in the mtDNA of *G. pallida* to conduct a phylogeographic study at the regional level. Second, we tested for the presence of geographical associations of haplotypes as expected from the high levels of allozymic differentiation among populations. Finally, we attempted to investigate the demographic events and identify the potential barriers to gene flow that might have shaped the distribution of mtDNA haplotypes in the sampled area. For these purposes, we sequenced a fragment of the mtDNA control region which was found to be highly variable at different population levels for this species of leaf beetle (P. Mardulyn *et al*., unpublished data) and analysed the data using nested clade analysis (Templeton *et al*. 1995; Templeton 1998). Using phylogenetic information, this approach specifically tests for geographical associations of haplotypes, and can discriminate between gene flow and population history as the actual causes of geographical distribution of haplotypes.

Materials and methods

Insect collection

A total of 14 populations of *Gonioctena pallida* was sampled during May 1999 in the Vosges mountains (12 populations) and part of the Black Forest (two populations). Sampling locations are shown in Fig. 1 and described in Table 1. The Vosges and Black Forest mountain ranges are separated from each other by a 30–40-km-wide lowland area, in which the river Rhine flows, dominated by urban areas and fields. Sampling in the Vosges mountains had to be restricted to high elevation areas of the southeast portion of the range, i.e. no population of *G. pallida* was found in lower elevation areas corresponding to the north and southwest portions of the Vosges. Here, a population is defined as a group of beetles associated with a geographically well-delimited patch of host plants (*Salix caprea* and/or *Salix aurita* and/or *Corylus avellana*). A maximum of one specimen per tree (larva or adult) was collected and special care was taken to sample each population as homogeneously as possible. Host-plant patches occupied 5000–40 000 m^2.

DNA sequencing

Genomic DNA was extracted from ethanol-preserved insects. Whole specimens were each ground in an SDS homogenization buffer, incubated overnight with proteinase K (2 mg/mL) at 40 °C, followed by three phenol–chloroform extractions, ethanol precipitation and resuspension in 10 mM Tris, 1 mM EDTA buffer (TE). For each of the 242 specimens, the entire mtDNA control region ($\approx$ 4 kb) was amplified

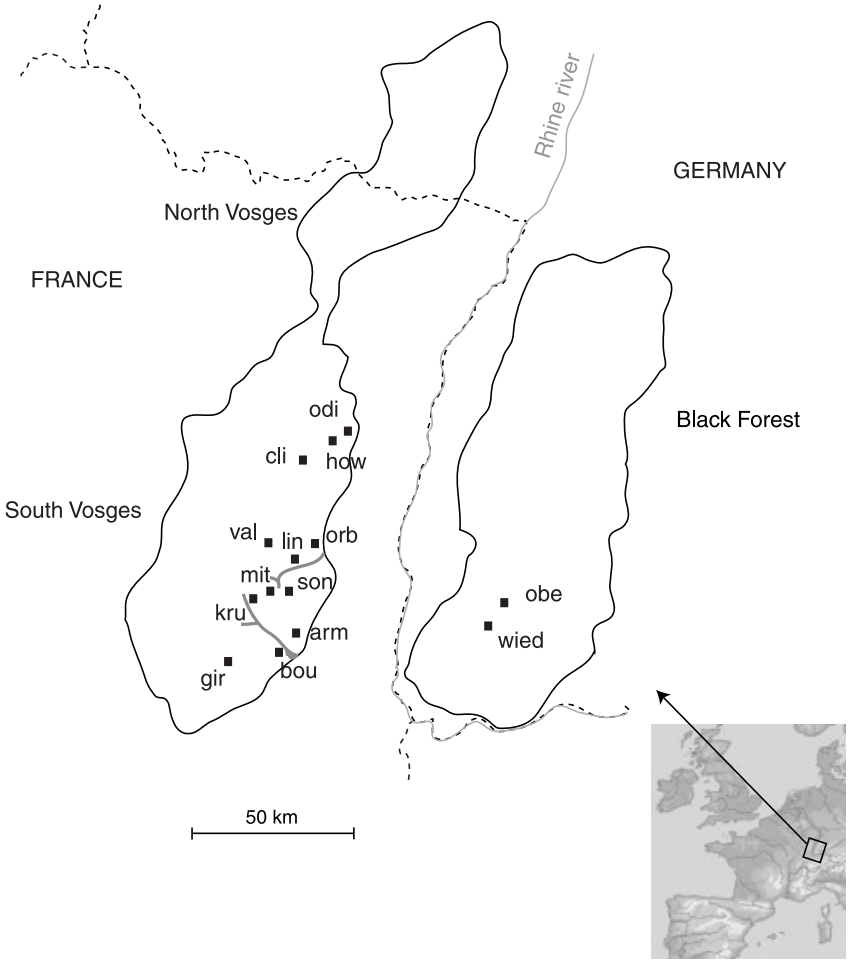

Fig. 1 Map showing the sampling localities in the Vosges (France) and Black Forest (Germany). Both mountain ranges are separated by a lowland area in which runs the river Rhine (grey line). A dashed line represents the border separating France from Germany. Sampling in the Vosges mountains was concentrated in the southeast portion of the range, characterized by higher elevations, where *Gonioctena pallida* is abundant. Abbreviations used for the localities names are those from Table 1. Two important valleys mentioned in the text (see discussion) are depicted (thick grey lines) inside the Vosges (other major valleys are also present in the Vosges but are not represented here).

Table 1 Sampling localities, geographical coordinates (decimal degrees), number of individuals collected and haplotypes identified in each sampling locality

Localities	Abbr.	Mountain range	Geographical coordinates	Sample size	Haplotypes (no. individuals)
Orbey	orb	Vosges	48.11–7.17	17	1(4), 20, 26(2), 34, 37, 45, 46, 50(6)
Ste-Odile	odi	Vosges	48.42–7.36	15	1(2), 6(2), 19(6), 20(2), 29(3)
Giromagny	gir	Vosges	47.80–6.83	18	1(7), 7, 8, 18, 24, 37, 40, 42(2), 50, 55, 56
le Vieil Armand	arm	Vosges	47.85–7.15	15	1(2), 3, 26(2), 32, 34, 35, 37(3), 43, 50(3)
le Climont	cli	Vosges	48.34–7.19	18	1(5), 5, 6, 19, 20(2), 25, 41, 44, 49, 50(4)
Kruth	kru	Vosges	47.95–6.96	19	1(8), 3(2), 9(2), 34, 37, 39, 43, 45, 47, 50
Bourbach	bou	Vosges	47.80–7.04	19	1(3), 3(2), 4, 5, 14, 24, 26, 34(2), 42(4), 50(2), 54
le Hohwald	how	Vosges	48.40–7.32	13	1(6), 5, 18(3), 33, 50(2)
Linge	lin	Vosges	48.07–7.13	17	1(9), 4(4), 35, 42, 50, 51
Mittlach	mit	Vosges	47.98–7.02	18	1(7), 10, 21, 26, 34, 36, 37(2), 45, 50, 57, 58
Sondernach	son	Vosges	47.98–7.06	17	1(4), 2(2), 6, 34, 35, 37, 38, 52, 59(3), 60(2)
le Valtin	val	Vosges	48.11–7.03	18	1(8), 6, 18, 24, 35(3), 50(2), 51, 53
Oberried	obe	Black Forest	47.93–7.98	19	4, 11, 12, 14, 16, 22, 26(10), 28, 30, 31
Wieden	wied	Black Forest	47.81–7.91	19	1(2), 13, 14(3), 15, 17, 23, 26(4), 27, 48, 50(3), 61

using the Expand Long Template PCR System kit (Roche) following the manufacturer's protocol and an annealing temperature of 55 °C. The primers used were designed in this study by reference to known sequences of different *Gonioctena* species (P. Mardulyn *et al.*, unpublished data) and are located either side of the control region, in the small subunit (12*S*) ribosomal RNA (5′-CATTATTTGTATAA-CCGCAACTGCTGGCAC-3′), and the methionine transfer

RNA (tRNA) (5′-TAACCTTYATAAATGGGGTATG-3′). These primers are named SR-J-14766 and TM-N-204, respectively, according to Simon *et al*. (1994). Alignment of the mtDNA control region from a restricted number of individuals identified a fragment ≈ 400 bp long as exhibiting the highest level of variability (P. Mardulyn *et al*., unpublished data). This fragment was then sequenced from each of the 242 PCR products using an internal primer (5′-AAATAAATATTCACAAAACCCC-3′) and the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit FS (Perkin–Elmer). It was not possible to name this new primer according to Simon *et al*. (1994) because the fragment sequenced displayed no region of homology with the control region of the *Drosophila* yakuba mtDNA reference sequence. Sequencing products were separated by electrophoresis on an ABI 377 automated DNA sequencer.

Data analyses

Sequences were edited using SEQPUP version 0.6 (Gilbert 1996) and could be aligned unambiguously. All analyses were performed on the 363 nucleotide-long portion of the alignment for which none of the sequences presented missing data. Individuals sharing the same haplotype were identified using the program COLLAPSE (v. 1.0). PARSPROB (v. 1.1) was then used to determine whether the parsimony criterion was valid for the data we wished to analyse. This program implements a Bayesian approach (Templeton *et al*. 1992) to calculate the probability that the differences between haplotypes are parsimonious (i.e. with no hidden mutations). We then used the parsimony criterion to infer a haplotype network with the method described in Templeton *et al*. (1992) and implemented in TCS ALPHA, v. 1.01 (Clement *et al*. 2000). COLLAPSE, PARSPROB (both written by David Posada), and TCS software are available at http://bioag.byu.edu/zoology/crandall_lab/programs.htm.

Ambiguities in the haplotype network were resolved following the two criteria suggested by Crandall & Templeton (1993): (i) rare haplotypes are more likely to be found at the tip, and common haplotypes at interior nodes, of a cladogram; and (ii) a singleton (i.e. a haplotype represented by a single individual) is more likely to be connected to haplotypes from the same population than to haplotypes from different populations. These criteria are derived directly from predictions of the coalescent theory that were inferred by different authors (e.g. Donnelly & Tavaré 1986). Crandall & Templeton (1993) further tested these predictions empirically by using 29 data sets displaying different levels of genetic variation in *Drosophila* at different loci. Based on their results, they specifically suggest using the criteria outlined above to resolve ambiguities in a cladogram.

Finally, a nested clade analysis (Templeton *et al*. 1995; Templeton 1998) was performed. The haplotype network was first converted manually into a nested series of clades, using the rules defined in Templeton *et al*. (1987) and Templeton & Sing (1993). A first test for geographical association was performed by treating sample locations as categorical variables: an exact permutational contingency test was conducted for each clade and a chi-squared statistic was calculated from the contingency tables (clades vs. geographical locations). A more refined test, using information on geographical distances, was then used. Two types of distances were calculated: (i) the 'clade distance', D_c, which measures how geographically widespread are the haplotypes within a given clade; and (ii) the 'nested clade distance', D_n, which measures how far the haplotypes within that clade are from the haplotypes of its evolutionarily closest sister clades (Templeton *et al*. 1995). The distributions of these two distances under the null hypothesis of no geographical associations within each clade (no correlation between the position of a haplotype in a clade and its geographical position) is determined by recalculating these distances after each of 10 000 permutations of the nested clades vs. geographical locations. This procedure allows us to test statistically for nonrandom geographical associations of haplotypes and to identify historical demographic events (population fragmentation or population expansion) or recurrent restricted gene flow among populations as the cause of the geographical association (Templeton *et al*. 1995; Templeton 1998). The rationale of the method can be described briefly as follows: inside a given tested clade, nested tip subclades are considered younger than the nested interior subclades to which they are connected (a prediction supported by neutral coalescent theory). Statistical comparison of the clade and nested clade distances calculated for the tip and interior subclades inside a given tested clade is therefore performed to search for patterns characteristic of range expansion or range fragmentation (defined as two nonrecurrent historical events, implying no recurrent gene flow), or of isolation by distance due to restricted gene flow. A detailed description of the distance patterns expected under these three alternative demographic hypotheses can be found in Templeton *et al*. (1995) and Templeton (1998). The statistical analyses applied to the constructed nested design were done with the program GEODIS 2.0 (Posada *et al*. 2000). The interpretation of the observed distance patterns was done using a revised version of the inference key published by Templeton *et al*. (1985) and Templeton (1998) (available at http://bioag.byu.edu/zoology/crandall_lab/geodis.htm).

Results

Sequence variation

Sequences of a short but highly variable fragment of the mtDNA control region were obtained for 242 individuals sampled from 14 populations. The aligned sequences were

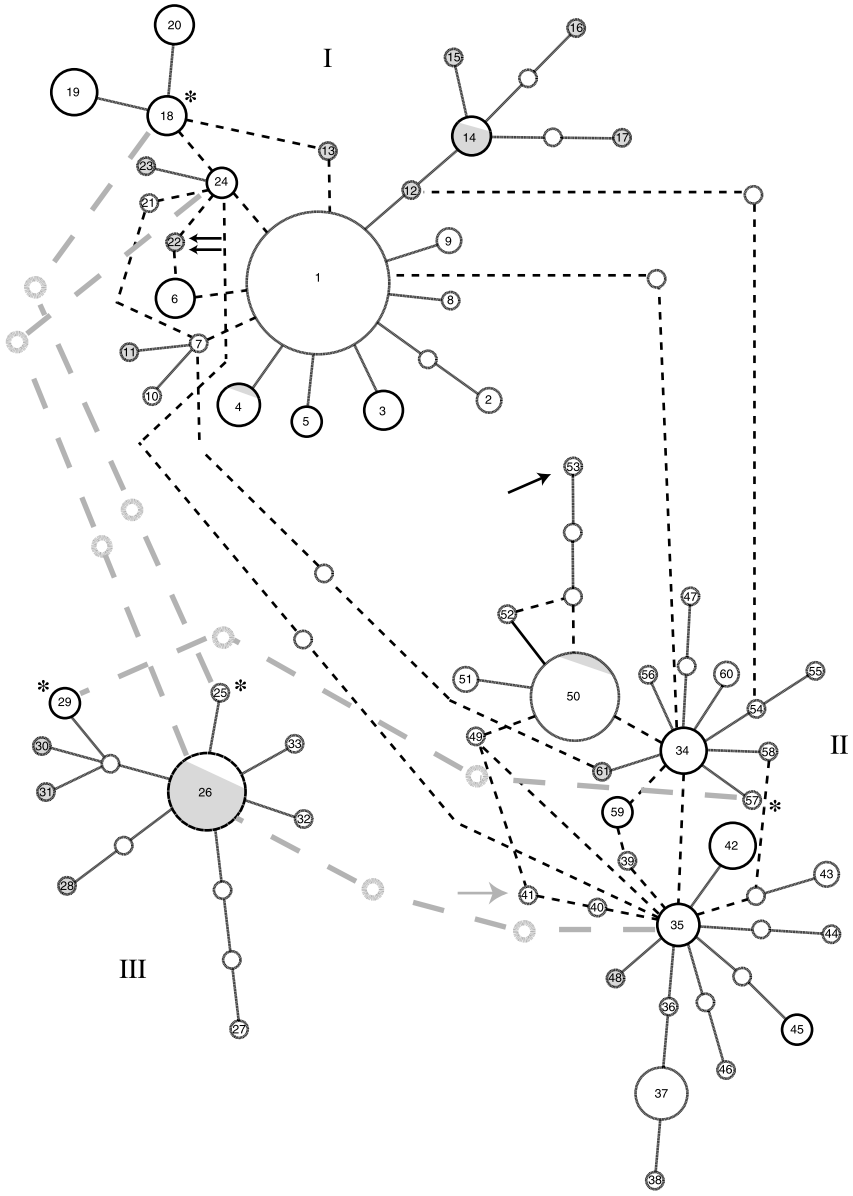

Fig. 2 Haplotype network for the mitochondrial DNA haplotypes of *Gonioctena pallida* inferred using the criterion of parsimony with the program TCS ALPHA 1.01 from Clement *et al*. (2000). Each line in the network represents a single mutational change. A haplotype is represented by a circle, the surface 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, but are necessary to link all observed haplotypes to the network. A grey area inside a given circle represents the proportion of individuals bearing the corresponding haplotype that were sampled in the Black Forest. Internal ambiguities (haplotypes interconnected forming a closed loop that can be broken at any places) and external ambiguities (tip haplotypes that have alternative most parsimonious connections to the rest of the network) are represented by dashed lines in the network. Three subnetworks can be identified (I, II and III) that are connected to one another by two or three mutational steps. Subnetwork III can be connected to subnetworks I or II by three mutational steps (grey dashed lines). Arrows and asterisks identify haplotypes that are referred to in the text.

pruned at both 5′- and 3′-ends to ensure that no missing data were present in the analysed data set. This resulted in a final alignment of 363 characters per taxon. Merging identical sequences yielded 61 unique haplotypes. The numbers of samples and haplotypes from each population are shown in Table 1. All sequences are deposited in GenBank under Accession nos AF346227–AF346287. Of a total of 363 sites, 47 were variable, two of which were due to the presence of indels. The resulting gaps (length of one and two nucleotides) in the data set were treated as missing data and recoded as present/absent characters at the end of the data matrix. Among the 47 variable sites, 19 are parsimony informative, including one indel site.

Haplotype network and nested design

Evaluation of the limits of validity of parsimony with the program PARSPROB suggests that a most parsimonious (MP) relationship of eight steps or fewer has a probability of > 95% of being true. The inference of the haplotype network with the TCS program resulted in the network presented in Fig. 2. Four star-like patterns (i.e. a centrally located common ancestral haplotype to which recently derived haplotypes are each connected) are linked to one another by one, two or three mutational steps, and form three subnetworks (I, II, III) whose evolutionary interrelationships are ambiguous. Several ambiguous relationships are also present inside two of the three subnetworks (I and II). These ambiguities result

from the presence of more than one most parsimonious connection of a haplotype to the rest of the network. We used the criteria suggested by Crandall & Templeton (1993) (see Materials and methods) to resolve these ambiguities. For example, haplotype 53 (cf. black arrow in Fig. 2) is three steps away both from haplotype 50 and haplotype 52. Because the frequency of haplotype 50 is much greater than that of haplotype 52, haplotype 53 is more likely to be connected to haplotype 50. Haplotype 41 (a singleton) (cf. grey arrow in Fig. 2) is one step away from haplotypes 49 and 40. Because haplotype 41 is found in the same locality as 49 (which is located far away from the locality in which 40 is found), haplotype 41 was connected to haplotype 49.

The relationships among subnetworks were resolved as follows. Subnetworks I (which includes haplotype 1) and II (which includes haplotypes 34, 35 and 50) are linked by two mutational steps. The connection between haplotypes 1 and 34 is more probable than the three other equally parsimonious possibilities because haplotypes 1 and 34 are likely to be ancestral haplotypes: they are high-frequency interior haplotypes connected to many low-frequency haplotypes. Subnetwork III (containing haplotype 26) is three steps away from both subnetwork I and subnetwork II. The most likely connection between subnetworks I and III (involving haplotypes 24 and 26, respectively) is only marginally less likely than the most likely connection between subnetworks II and III (involving haplotypes 35 and 26, respectively), as the frequency of haplotype 35 is only slightly higher than that of haplotype 24. We therefore considered the two connections alternatively when performing the nested clade analysis.

When resolving ambiguities, i.e. when choosing the most likely connections among all equally parsimonious connections, we may be left with connections that are no longer supported by a > 95% probability of being true. Although Crandall & Templeton (1993) do not present a method to calculate the probability of different alternative resolutions of ambiguities in a network, their results suggest a high negative correlation between the probability of a haplotype being at the tip of a cladogram and its frequency of occurrence in the sample. Moreover, their study suggests that haplotypes with a greater frequency of occurrence tend to have a greater number of mutational connections. When different resolutions of a given ambiguity did not have clearly contrasted probabilities (as suggested by a large difference in haplotype frequencies), we successively tested these alternative resolutions on the results of the nested clade analysis. For example, haplotype 41 (grey arrow) was alternatively connected to haplotypes 40 and 49; haplotype 22 (double arrow) was alternatively connected to haplotypes 24 and 6; and haplotype 26 (central haplotype of subnetwork III) was connected to haplotypes 35 and 24. However, we did not test the connection between haplotypes 29 and 57, or between 25 and 18 (all indicated by asterisks in Fig. 2), as these connections seem much less likely than their alternatives. Figure 3 shows the haplotype network, and its nested design inferred following the rules of Templeton *et al*. (1987) and Templeton & Sing (1993), with the preferred resolution of ambiguities [other resolutions were tested as well, as explained above: 1-24-22 (instead of 1-6-22), 1-7-21 (instead of 1-24-21), 26-24 (instead of 35-26), 34-59-39 (instead of 35-39 and 34-59), 35-49 (instead of 50-49) and 35-40-41 (instead of 50-49-41)].

Table 2 Nested contingency analysis of geographical associations

Clade	Chi-square statistic	Probability
1-1	106.60	0.080
1-5	3.50	0.525
1-6	6.00	1.000
1-7	10.00	1.000
1-8	21.37	0.002*
1-10	39.78	0.064
1-11	6.00	0.667
1-13	2.00	1.000
1-14	52.57	0.207
1-16	49.33	0.008*
1-17	7.74	0.721
1-22	5.00	0.102
1-23	52.90	0.013*
2-1	65.55	0.014*
2-2	2.91	0.719
2-3	16.31	0.017*
2-4	16.37	0.673
2-5	15.44	0.405
2-6	75.34	0.036*
2-7	24.10	0.407
3-1	88.05	0.000*
3-2	23.20	0.006*
3-3	39.87	0.000*
Total cladogram	53.63	0.001*

*Significant at the 0.05 level.

Nested clade analysis

Table 2 shows the results of the nested contingency analysis of geographical associations. Figure 4 presents the results of the nested clade analysis of geographical distances performed using GEODIS 2.0. Table 3 presents the interpretations of these statistical results using the latest key from Templeton (cf. Materials and methods). For each clade that has resulted in the inference of a specific demographic event in Table 3, Fig. 5 shows the geographical distribution of the nested interior clade(s) vs. the geographical distribution of some of the nested tip clade(s) (those that show a significantly small or large geographical range) on a map. We attempt in this figure to highlight the phylogeographic patterns uncovered by the nested clade analysis.

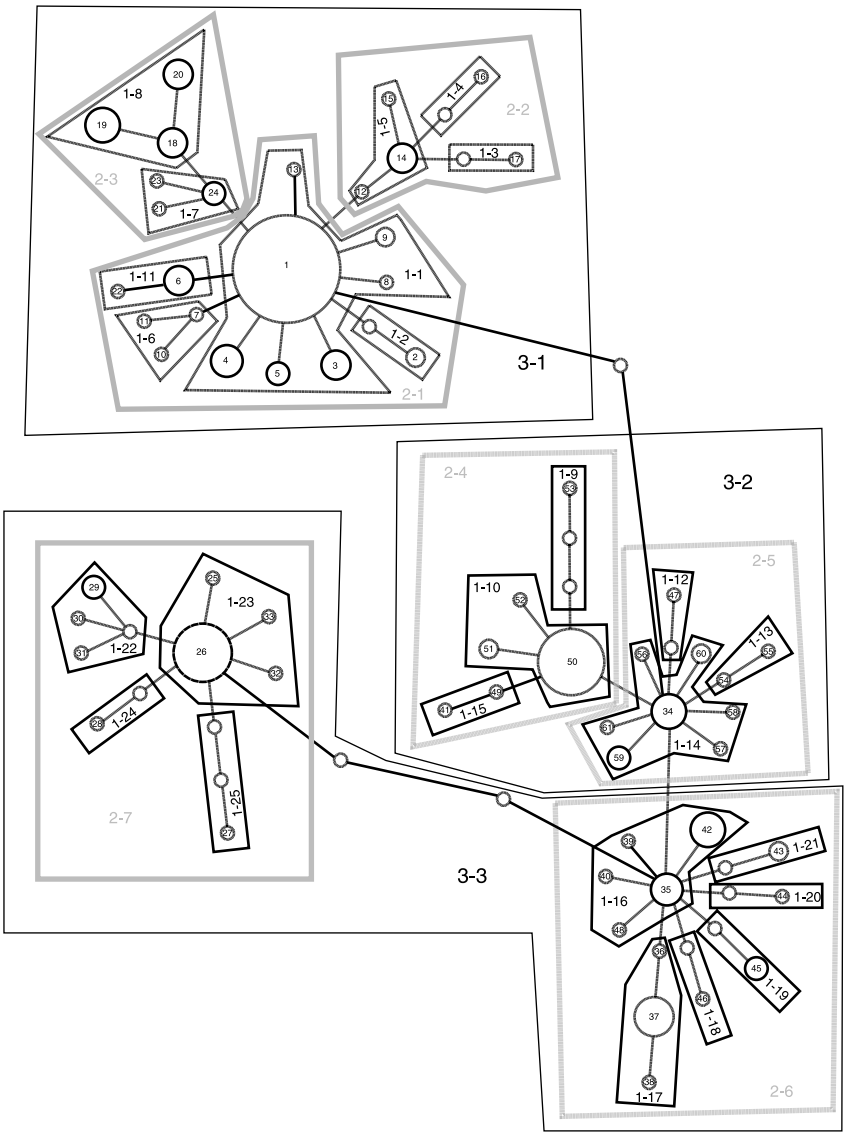

Fig. 3 The preferred resolution of the ambiguities for the haplotype network shown in Fig. 2, with the associated nested design used for the statistical analyses and inferred using the rules defined by Templeton *et al*. (1987) and Templeton & Sing (1993).

Table 3 Interpretation of the results of Fig. 4 using the key of inference of Templeton *et al*. (1995)

Clade	Chain of inference	Demographic event inferred
1-8	1-2-3-4-NO	restricted gene flow with isolation by distance
1-14	1-2-3-4-NO	restricted gene flow with isolation by distance
1-16	1-2-3-4-NO	restricted gene flow with isolation by distance
	1-2-3-4-9-NO	or past fragmentation
1-23	1-2-11-12-NO	contiguous range expansion
2-1	1-2-11-12-NO	contiguous range expansion
2-3	1-2-3-4-NO	restricted gene flow with isolation by distance
	1-2-3-4-9-NO	or past fragmentation
2-6	1-2-3-4-NO	restricted gene flow with isolation by distance
2-7	1-2-11-17-NO	inconclusive outcome
3-1	1-2-11-12-13-14-YES	long distance colonization
3-2	1-2-11-12-NO	contiguous range expansion
3-3	1-2-11-12-NO	contiguous range expansion
Total	1-2-11-12-NO	contiguous range expansion

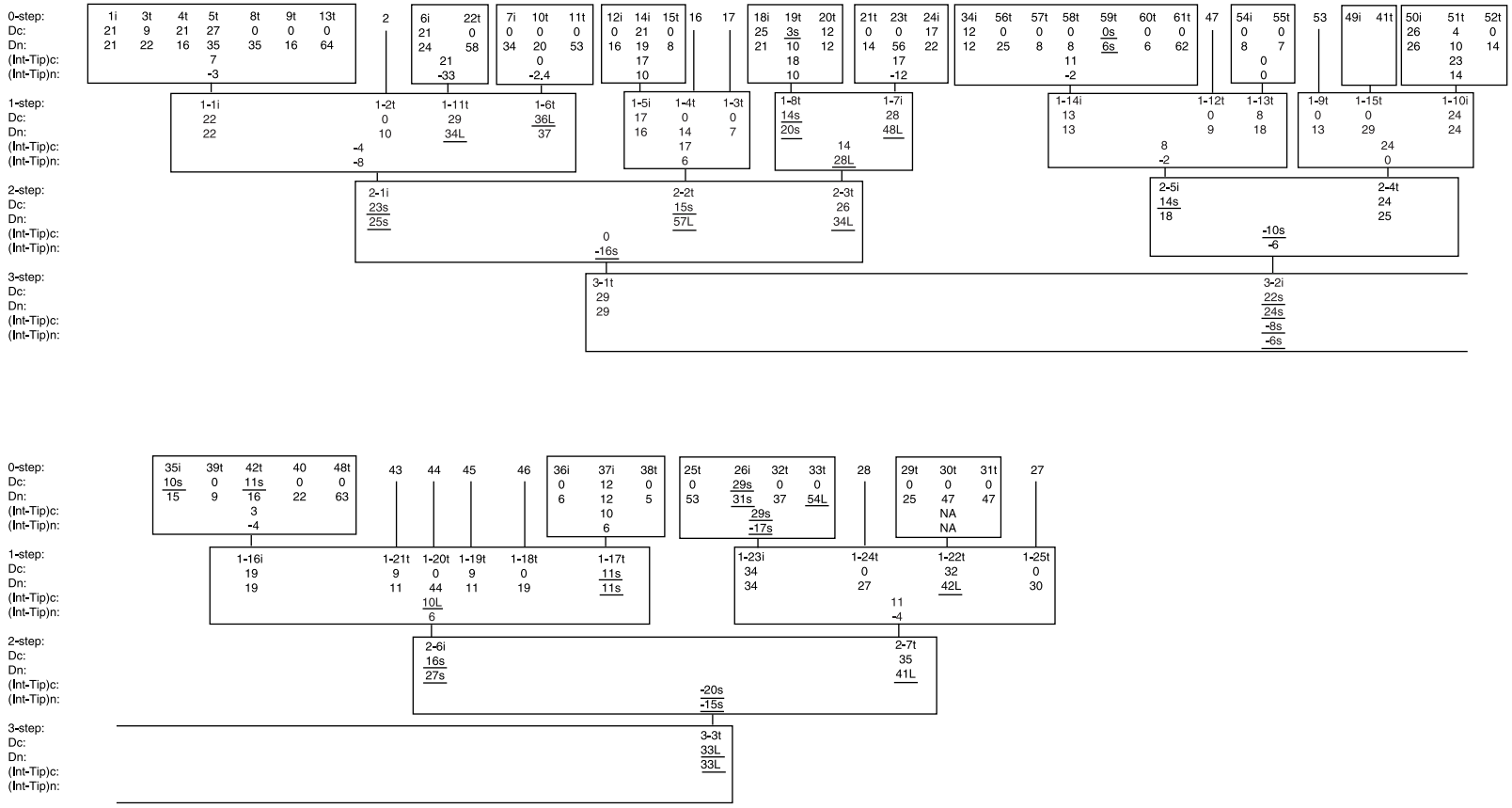

Fig. 4 Results of the nested clade analysis of geographical distances for the mitochondrial DNA haplotypes of *Gonioctena pallida*. Haplotypes numbers (first line) are those shown on Figs 2 and 3. An 'i' or a 't' next to the haplotype number stands for 'interior' or 'tip', respectively. Haplotypes nested in a one-step clade are grouped in boxes as in the nested design of Fig. 3. Higher level clades are represented as one moves down the figure. In each box, the clade distance (D_c) and nested clade distance (D_n) calculated for each clade within the nested group is shown, as well as the average difference in distances between interior clades and tip clades for D_c and D_n [(Int–Tip)$_c$ and (Int–Tip)$_n$, respectively]. Significantly large or small distance values (at the 5% level) are underlined and characterized with an 'L' or an 's', respectively.

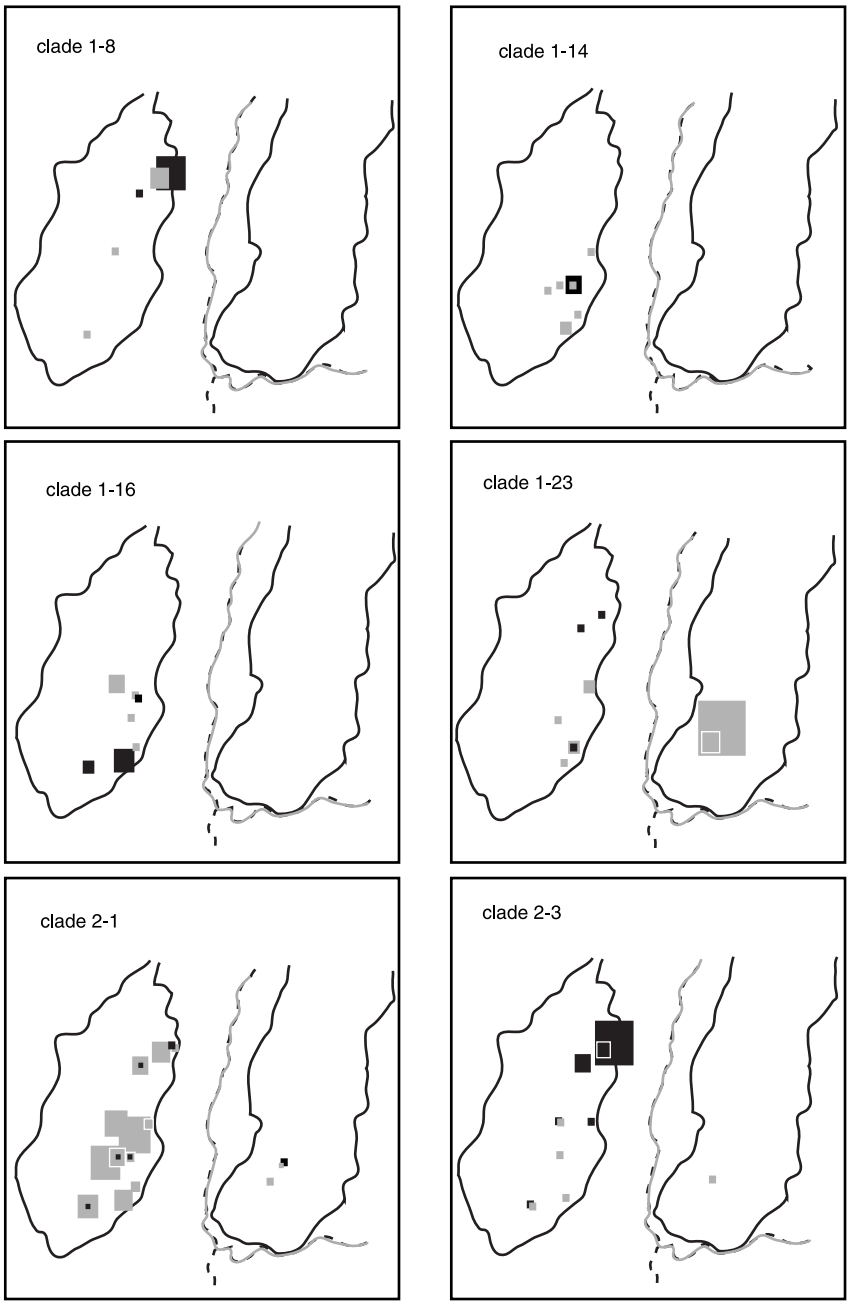

Fig. 5 For each clade for which a demographic event was inferred in Table 3, location of the different individuals belonging to some of the nested clades (those with significantly small or large D_c values) is shown on a map. A square represents individuals from one or more nested clades that were sampled in the corresponding locality. Grey squares represent individuals belonging to interior clades, while black or white squares (with black borders) represent individuals belonging to tip clades. The size of a square within a map is proportional to the number of individuals it represents (it is *not* proportional to the size of the sampled area). The size of the squares within a map are adjusted to avoid their extensive overlap, and square sizes can therefore not be compared among different maps. When the surfaces of two squares from the same colour overlap completely, a white border is drawn around the smaller square to allow visualization of the smaller square. When two squares of different colours and of identical size are located in the same sampling site, they are moved away from one another slightly so that both can be seen. The geographical distribution of the following nested clades is depicted: clades 0-18 and 0-19 nested in clade 1-8, clades 0-34 and 0-59 nested in clade 1-14, clades 0-35 and 0-42 nested in clade 1-16, all clades nested in clade 1-23, clades 1-1, 1-6 and 1-11 nested in clade 2-1, all clades nested in clade 2-3, clades 1-16 and 1-17 nested in clade 2-6, all clades nested in clade 3-1, all clades nested in clade 3-2, all clades nested in clade 3-3, clades 3-2 and 3-3 nested in the entire cladogram. The last map of the figure presents a summary of the expansion events suggested by the nested clade analysis.

On two occasions (clades 1-16 and 2-3), it was difficult to choose between past fragmentation and restricted gene flow with isolation by distance. In both cases (Fig. 5), an overlap between the geographical ranges of the individuals bearing haplotypes from interior and tip clades occurs, yet this overlap is due to one (for clade 1-16) or two (for clade 2-3) individual(s) from one tip clade. This pattern of overlap may result from restricted gene flow causing isolation by distance. However, because the overlap is due only to one or two individuals, a past range fragmentation event can also be invoked, followed by a more recent expansion event that has brought both isolated populations back into contact. If gene flow has been sufficiently restricted, the genetic pattern revealing the past fragmentation event persists, causing both interior and tip clades to overlap only partially in their geographical range (Templeton *et al*. 1995).

Testing the alternative resolutions of ambiguities in the haplotype network produced very similar results. The pattern inferred with clade 2-1 resulted in an inconclusive outcome (instead of a contiguous range expansion) in one alternative network, and the pattern inferred with the total cladogram was interpreted as a restricted gene flow with

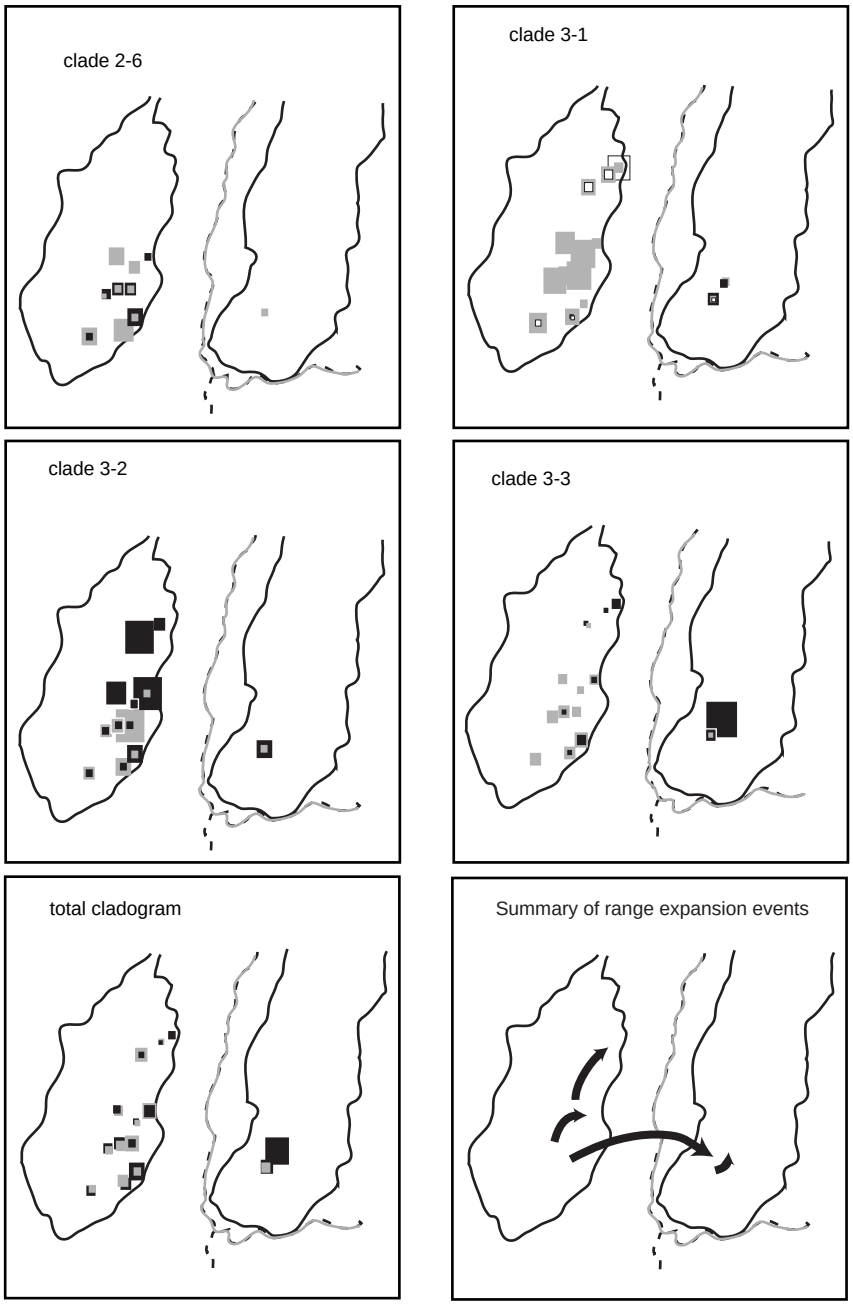

Fig. 5 *Continued*

isolation by distance (instead of a contiguous range expansion) in another alternative network (when subnetwork III is connected to subnetwork I). Otherwise, all other inferred demographic patterns were identical to those inferred with the network presented in Fig. 3.

Discussion

Extensive mtDNA variation was observed in *Gonioctena pallida* populations of the Vosges and the Black Forest. In the 242 individuals surveyed, 61 haplotypes were identified for a small 363 bp fragment of the mtDNA control region which appears to be sufficiently variable for phylogeographic studies of leaf beetles at the regional level, i.e. in the 100×40 km mountainous area surveyed.

Four star-like phylogenies were observed in the haplotype network (Fig. 2). A star-like pattern in which one very common haplotype, presumably ancestral, lies at the centre of a network and is connected by independent mutation steps to many much rarer haplotypes, is usually regarded as indicative of a population having recently expanded in size from one or a small number of founders (Avise 2000; see examples in Ball *et al*. 1988 and Merilä *et al*. 1997). Subnetworks I and III conform rather well to this description.

Subnetwork II includes two central haplotypes (34 and 35) also each connected to many other haplotypes. In those cases, however, the central haplotypes are only slightly more common than some of the tip haplotypes. Some of the peripheral haplotypes are even more frequent than the central haplotypes (i.e. haplotypes 50, 37, 42). If we assume that the great number of connections associated with central haplotypes 34 and 35 indicates that these haplotypes were more frequent at some point in the past [as already mentioned, Crandall & Templeton (1993) have shown that haplotypes with greater frequency tend to have a greater number of mutational connections], then the observed pattern for subnetwork II may reflect an older expansion event, as the frequency of the central haplotypes may have decreased from the presumed expansion event to the present. More theoretical and empirical work is needed, however, to firmly link the patterns observed in the network with the proposed scenario. Nevertheless, the four lineages defined by each star phylogeny are broadly sympatric and the observed pattern of genetic variation is compatible with a more or less recent expansion of the *G. pallida* population from at least two of the four suggested ancestral haplotypes.

The null hypothesis of no geographical association was rejected for a large number of clades (12 of 24 clades tested, see Fig. 4), confirming the strong genetic differentiation found by allozyme studies of leaf beetle populations. Three reasons may explain why the null hypothesis was not rejected in 12 clades (Templeton *et al*. 1995): (i) the populations are virtually panmictic (high level of gene flow) and have not experienced fragmentation or expansion events, (ii) sampling is inadequate (sample sizes are too small, or the geographical location of the samples is inadequate), or (iii) the genetic variation present in the sample is not sufficient. The first explanation can be rejected safely as strong geographical associations were observed at different clade levels.

Using nested clade analysis, we were able to suggest specific demographic events as the causes of the detected geographical associations. First, three cases of restricted gene flow leading to isolation by distance (clades 1-8, 1-14 and 2-6) were inferred (tip clade has a significantly smaller D_c than the interior clade, see Fig. 5). Secondly, six range expansion events, including one long-distance colonization from the Vosges to the Black Forest (clade 3-1; see the geographical distribution of the nested tip clade 2-2 represented by black boxes in Fig. 5) and five contiguous range expansions (restricted range for the nested interior clade vs. the nested tip clades). These contiguous range expansion events appear to have been directed from the middle of the Vosges to the north (clades 1-23, 3-2 and 3-3), and toward the northeast inside the Black Forest (clades 2-1 and 3-3, total cladogram).

Thirdly, two clades display a pattern which cannot be interpreted unambiguously. Indeed, although clade 1-16 may define a former split between the southern and middle ranges of the studied area, and clade 2-3 a former split between the north and the rest of the Vosges, each suggesting a past fragmentation event, restricted gene flow with isolation by distance cannot be excluded as an explanation for the geographical distribution observed.

The pattern of contiguous range expansion inferred for the total cladogram (geographical distribution of interior clade 3-2 vs. tip clade 3-3 represented in Fig. 5) is very similar to that inferred for clade 3-3 (interior clade 2-6 vs. tip clade 2-7, Fig. 5). It is likely that the strong geographical association pattern detected for clade 3-3 and suggesting a contiguous range expansion event is the same detected at the next higher level of nesting.

Given the many demographic events highlighted by the nested clade analysis implying that gene flow is restricted or absent (range expansion and range fragmentation are defined as two nonrecurrent historical events, implying no recurrent gene flow; see Materials and methods) effective barriers to gene flow are clearly present inside the studied area. This is not surprising given the low power of dispersal suspected for leaf beetles and the patchy distribution of the host plant that results in important habitat discontinuities. Although the distribution of the host plant may be sufficient to restrict gene flow among populations, habitat discontinuity is further reinforced by the rapid elevation changes encountered in mountainous habitats. For *G. pallida*, valleys (low elevation) are inappropriate environments which may therefore prevent gene flow and explain some of the demographic patterns inferred in this study. One obvious example is the large habitat gap formed by the lowland area separating the Vosges mountains from the Black Forest. Indeed, only one long-distance colonization pattern was detected between both mountain ranges. Because this historical event implies, by definition, a single colonization event, with no recurrent gene flow (even restricted) occurring between both mountain ranges, the lowland area separating the Vosges from the Black Forest appears to be a very effective barrier.

Other nonrecurrent historical patterns were recorded within both mountain ranges. Two of these patterns could be explained by the topology of the Vosges. First, a long valley running from the east of the Vosges (from the town of Colmar) toward the southwest separates the 'val', 'lin' and 'orb' localities from the other southern localities (Fig. 1) and may represent an effective barrier to migration for *G. pallida*. This valley may therefore explain the range expansion pattern revealed by clade 3-2 (Fig. 5). Secondly, another long valley in the Vosges separates the two southernmost localities from the rest of the Vosges (Fig. 1) and may explain the past fragmentation event possibly detected by analysis of clade 1-16. Other nonrecurrent historical events detected in the nested clade analysis cannot be easily explained by the topology of the studied area. For

example, no obvious geographical barrier can be observed on a map between the two sampling sites in the Black Forest, or between the 'cli' locality and the two other localities of the northern Vosges, although contiguous range expansion events in those areas were inferred. Of course, important discontinuities in host-plant distribution could also act as an important barrier to gene flow. Unfortunately, we lack accurate information on the host-plant's geographical distribution inside the studied area, as well as its dynamic in the relevant period. Moreover, the absence of a molecular clock calibration for the DNA fragment used in this study precludes the possibility of dating the inferred historical demographic events. For these reasons, it is unfortunately difficult to discuss the causes (e.g. climatic) of gene flow restriction in more detail.

Nevertheless, our results indicate that the movements of *G. pallida*, a species restricted to high elevation areas in central Europe, are highly restricted inside the mountainous regional area investigated. No recurrent gene flow occurs among mountain ranges in *G. pallida* (e.g. between the Vosges and the Black Forest) and it is likely that this pattern is relevant for other leaf beetle species restricted to mountain habitats. Moreover, we demonstrated that effective barriers to gene flow are also present within a mountain range as small as the Vosges. Although patterns of isolation by distance were identified in a few cases inside a limited area, several expansion events, implying no recurrent gene flow, were suggested by the nested clade analysis. As a consequence, extreme caution must be taken when estimating gene flow levels among leaf beetle populations using genetic differentiation patterns inferred with molecular data such as DNA, allozyme or microsatellite data. Indeed, similarity among populations is likely to be due to shared ancestry rather than to gene flow occurring among them. Before attempting to estimate gene flow among leaf beetle populations, using any kind of molecular data, it is recommended that the actual presence of gene flow be tested first using a method such as nested clade analysis.

More than half of the demographic events revealed by our analyses are range expansions. Moreover, most of the evidence for range expansion events was observed at the highest nesting clade levels (Table 3). Indeed, statistical analysis of each of the three-step clades and of the total cladogram resulted in the inference of a range expansion event. Because the age of a clade is the age of the oldest clade nested within it, it tends to increase on average when going from one clade level to the next higher (Templeton *et al*. 1995). Thus, the expansion events inferred for each of the four highest level clades (3-1, 3-2, 3-3 and total cladogram) are the most ancient demographic events recorded in the DNA fragment sequenced. This suggests that the small population of founders at the origin of the actual population of the Vosges mountains has not only expanded in size, as hinted by the star-like patterns observed in the haplotype network, but has also expanded in geographical range from the south to the north as well as towards the east (i.e. into the Black Forest mountains). More studies of this type are needed to determine if the phylogeographic pattern observed for *G. pallida* is representative of leaf beetle species in general.

It should be possible to deduce the history of colonization and fragmentation events in *G. pallida* among European mountain ranges (Black Forest, Alps, Massif Central, Pyrenees) by sampling populations in these areas and surveying their patterns of DNA variation. Understanding the demographic events that result in the current distribution of genetic variation from the lower to the higher population level, ultimately reaching the interface between population and species, should allow us to gain insight into the mechanisms of speciation for *G. pallida* in particular and leaf beetles in general. Note that the patterns inferred in this study were based solely on mtDNA, which is inherited through maternal lines only (e.g. Dawid & Blackler 1972; Hutchison *et al*. 1974). As the migration behaviour of females can be different from that of males, nuclear loci should also be investigated in future studies.

Acknowledgements

I thank Keith Crandall and Alan Templeton for answering some of my questions on the nested clade analysis, and John Avise for answering some of my questions on the interpretation of haplotype networks. Michel Milinkovitch, Jacques Pasteels and three anonymous reviewers provided many helpful comments on an earlier version of the manuscript. This research was supported by the Belgian FNRS (National Fund for Scientific Research), the Free University of Brussels (ULB), the Van Buuren Fund, the 'Communauté Française de Belgique' (ARC 98/03-223), and the Defay Fund. P. M is 'Postdoctoral Researcher' at the FNRS.

References

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

Axelsson B, Bossata E, Lohm U, Persson T, Tenow O (1974) Energy flow through a larval population of *Phytodecta pallidus* L. (Col., Chrysomelidae) on *Corylus avellana* L. *Zoon*, **2**, 49–55.

Ball RM Jr, James FC, Freeman S, Bermingham E, Avise JC (1988) Phylogeographic population structure of red-winged blackbirds assessed by mitochondrial DNA. *Proceedings of the National Academy of Sciences of the USA*, **85**, 1558–1562.

Bossart JL, Pashley Prowell D (1998) Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution*, **13**, 202–206.

Cantonnet F (1968) Révision des espèces françaises du genre *Phytodecta* et description d'une espèce nouvelle. *Entomologiste*, **24**, 38–49.

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

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

Dawid IB, Blackler AW (1972) Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. *Developmental Biology*, **29**, 152–161.

Donnelly P, Tavaré S (1986) The ages of alleles and a coalescent. *Advance Applied Probability*, **18**, 1–19.

Gilbert DG (1996) SEQPUP, *Version 0.6*. Available via anonymous ftp to ftp.bio.indiana.edu

Hutchison CA III, Newbold JE, Potter SS, Edgell MH (1974) Maternal inheritance of mammalian mitochondrial DNA. *Nature*, **251**, 536–538.

Kippenberg H (1966) Familie: Chrysomelidae. Band 9. In: *Die Käfer Mitteleuropas* (eds Freude H, Harde KW, Lohse GA), pp. 1–299. Goecke & Evers, Krefeld.

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 Jolivet PHA, Cox ML), pp. 379–388. SPB Academic Publishing, Amsterdam.

Mardulyn P, Milinkovitch MC, Pasteels JM (1997) Phylogenetic analyses of DNA and allozyme data suggest that *Gonioctena* leaf beetles (Coleoptera; Chrysomelidae) experienced convergent evolution in their history of host-plant family shifts. *Systematic Biology*, **46**, 722–747.

Mason ML, Lawson FA (1980) Dispersal of the pre-overwintering adult stage of the American aspen beetle (Coleoptera: Chrysomelidae: *Gonioctena americana* Schaeffer) using Hg-203. *Journal of the Kansas Entomological Society*, **53**, 595–596.

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.

Merilä J, Björkland M, Baker AJ (1997) Historical demography and present day population structure of the greenfinch, *Carduelis chloris*—an analysis of mtDNA control-region sequences. *Evolution*, **51**, 946–956.

Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.

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

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

Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (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 Systematics*, **16**, 393–430.

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–351.

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

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.

Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.

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

Patrick Mardulyn is a Postdoctoral Researcher at the Belgian National Fund for Scientific Research (FNRS). His current study investigates the phylogeography of leaf beetles, using molecular markers, to understand the mechanisms leading to genetic differentiation and to investigate the interface between populations and species.
