J Mol Evol (2003) 56:38–45
DOI: 10.1007/s00239-002-2378-1

JOURNAL OF MOLECULAR EVOLUTION

Structure and Evolution of the Mitochondrial Control Region of Leaf Beetles (Coleoptera: Chrysomelidae): A Hierarchical Analysis of Nucleotide Sequence Variation

Patrick Mardulyn,[1,2] Arnaud Termonia,[1,2] Michel C. Milinkovitch[1]

[1] Unit of Evolutionary Genetics, Free University of Brussels, Rue Jeener et Brachet 12, 6041 Gosselies, Belgium
[2] Laboratoire de Biologie Animale et Cellulaire, CP 160/12, Free University of Brussels, av. F.D. Roosevelt 50, 1050 Brussels, Belgium

Received: 31 October 2001 / Accepted: 19 July 2002

Abstract. To assess the levels of variation at different evolutionary scales in the mitochondrial (mt) control region of leaf beetles, we sequenced and compared the full mt control region in two genera (*Chrysomela* and *Gonioctena*), in two species within a genus (*Gonioctena olivacea* and *G. pallida*), in individuals from distant populations of these species in Europe, and in individuals from populations separated by moderate (10- to 100-km) to short (< 5-km) distances. In all individuals, a highly repetitive section consisting of the tandem repetition of 12 to 17 imperfect copies of a 107- to 159-bp-long core sequence was observed. This repetitive fragment accounts for roughly 50% of the full control-region length. The sequence variability among repeated elements within the control region of a given individual depends on the species considered: the variability within any *G. olivacea* individual is much higher than that within *G. pallida* individuals. Comparisons of the repeated elements, in a phylogenetic framework, within and among individuals of *G. olivacea* and *G. pallida* suggests that the repetitive section of the control region experienced recurrent duplications/deletions, leading to some degree of concerted evolution. Comparisons between *Chrysomela* and *Gonioctena* control regions revealed virtually no significant sequence similarity, except for two long stretches of A's and several [T(T)A(A)] repeats, all found in the control region of other insect orders. Our analyses allowed us to identify portions of the control region with enough variation for population genetic or phylogeographic studies.

Key words: Mitochondrial DNA — Control region — Repeated elements — Chrysomelidae — Sequence variation — Concerted evolution

Correspondence to: Patrick Mardulyn; *email:* pmarduly@ulb.ac.be

Introduction

The noncoding portion of the mitochondrial (mt) genome in animals is called the control region because it is believed to control the transcription and replication of the mtDNA molecule. In vertebrates, the control region has been shown to contain the promoters for transcription initiation and the origin of heavy-strand DNA replication (e.g., Shadel and Clayton 1997). In insects, this region is usually called the "A + T-rich region" following Fauron and Wolstenholme (1976), because it exhibits a frequency of deoxyadenylate and deoxythymidylate residues higher than in the rest of the mt molecule. It was found that the origin of replication lies within the control region in *Drosophila* (Goddard and Wolstenholme 1978), and promoters for transcription have possibly been identified in this segment as well (Lewis et al. 1994).

The control region varies greatly in size among insects: from ±350 bp in Lepidoptera (Taylor et al. 1993) to 13 kb in bark weevils (Coleoptera) (Boyce et al. 1989). Large differences in size can also be found among closely related taxa, as sizes from 1 to 5 kb were reported among *Drosophila* species (Lewis et al. 1994), as well as within species. This length polymorphism in the control region often results from the presence of a variable number of tandemly repeated elements (Rand 1993; Zhang and Hewitt 1997; Lunt et al. 1998).

Because the control region is known generally to experience high rates of evolution in both nucleotide substitution and insertion/deletion events (e.g., Moritz et al. 1987), although this is not always true for insects (Zhang and Hewitt 1997), it is increasingly used as a polymorphic marker for population genetic studies or for phylogenetic analyses involving closely related taxa (e.g., McMillan and Palumbi 1997; Rüber et al. 2001). However, to our knowledge, this marker has never been used for coleopterans, an insect order for which no sequence of the control region is available in international sequence databases. We present here a hierarchical study of sequence variation in leaf beetles: we sequenced and compared the full mt control region in two genera, in two species within a genus, in individuals from distant populations of these species in Europe, and in individuals from populations separated by moderate (10- to 100-km) to short (<5-km) distances. The aim of this exploratory work is to assess the levels of variation at different evolutionary scales in the mt control region of leaf beetles and infer its mode and tempo of evolution. We also compare the leaf beetle control-region organization to those observed in other orders of insects.

Materials and Methods

Insect Collection

Sequencing of the mt control region from 24 individuals (see Table 1 for collection sites) allowed the comparison of these sequences between two leaf beetle genera (*Gonioctena* and *Chrysomela,* both belonging to the tribe Chrysomelini, subfamily Chrysomelinae), between two species within the genus *Gonioctena* (*G. pallida* and *G. olivacea*), among geographically distant populations of these three species in Europe, among five populations of *G. pallida* across the Vosges mountains and the Black Forest, and among five populations of *G. olivacea* inside a 5 × 2-km area in the Belgian Ardennes. A population is defined here 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* for *G. pallida*, *Sarothamnus scoparius* for *G. olivacea*, and *Salix spp.* or *Betula spp.* for *Chrysomela lapponica*).

DNA Sequencing

We sequenced all or part (i.e., excluding the repetitive region) of the mt control region for all individuals listed in Table 1. Genomic DNA was extracted from ethanol-preserved insects. Whole specimens were individually ground in an SDS homogenization buffer and incubated overnight with proteinase K (2 mg/ml) at 40°C, followed by three phenol/chloroform extractions, ethanol precipitation, and resuspension in TE buffer (10 m*M* Tris, 1 m*M* EDTA). Because no sequences of the control region and of the adjacent NADH dehydrogenase subunit 2 (ND2) were available in GenBank for Coleoptera, a large fragment of 6–8 kb, including the control region, the ND2 gene, and portions of the cytochrome oxidase I (COI) and of the small-subunit rRNA (12S) genes, was PCR-amplified from two individuals per species. This amplification was performed with 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 12 min at 68°C. The primers used are modified versions of universal primers described by Simon et al. (1994) and are located in the COI and 12s genes: C1-N-2414 (5′GTGCTAATCATCTAAAAATTTTAATTCCTG3′) (modified reverse complement of Cl-J-2441) and modified (longer version) SR-J-14233 (5′CCAGTAAGAGYGACGGGCGATGTGT3′). These primers were chosen because they were previously used successfully for the amplification of shorter fragments in *Gonioctena* leaf beetles (unpublished data). Shorter versions of these primers were used to sequence the flanking regions (+/−700 nucleotides) of the long PCR products (ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kit FS; Applied Biosystems). Sequencing products were separated by electrophoresis on an ABI 377 automated DNA sequencer. New primers were designed on the sequenced flanking regions and used to sequence the long PCR fragment further. This primer-walking strategy was continued until reaching the 3′ and 5′ boundaries of the control region. We thereafter designed new primers [and named them according to Simon and co-workers' (1994) nomenclature] for the amplification of the complete mt control region (3.3–5.5 kb): SR-J-14766 (5′CATTATTTGTATAACCGCAACTGCTGGCAC3′) in 12S and TM-N-204 (5′TAACCTTYATAAATGGGGTATG3′) in the methionine transfer RNA (tRNA). This was done 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 amplified products were purified and ligated into a pBluescript II SK(+) vector (Stratagene) and transferred to *E. coli* DH5α competent cells. Plasmids from positive clones were isolated and subjected to in vitro transposition reactions (TGS; Finnzymes) that randomly insert one (and only one) transposon per plasmid. The transposon contains priming sites and a gene conferring antibiotic resistance. The resulting products were again transferred to *E. coli* DH5α competent cells. Clones containing a randomly inserted transposon were selected through antibiotic resistance and sequenced using two primers annealing to the transposon. Each clone sequenced with both primers (sequencing in opposite directions) yielded approximately 1400 bp of sequence. Contigs were assembled manually.

Control-region sequences were obtained with this transposon insertion strategy for three individuals per species. On the basis of the alignments among these three sequences we designed, for each species, a set of 8–12 primers allowing the sequencing (on both strands) of the entire (or nearly so, see below) control region initially PCR-amplified with SR-J-14766 and TM-N-204. In some instances, the within-species variability of control region sequences forced us to design population-specific primers. The control region of all studied species includes an area of about 1500–2500 bp (i.e., roughly 50% of the full control region) consisting of >10 imperfect, tandemly arranged, copies of a >100-bp motif (see Results). For two species, *G. pallida* and *C. lapponica*, the repeated elements were too similar in sequence for the design of specific sequencing primers. This problematic region was therefore sequenced in these two species using the cloning–transposon insertion strategy described above. At least three clones (up to five) per individual were sequenced to avoid sequencing errors due to the cloning procedure.

Table 1. Sampling localities of individuals collected for which the control region was sequenced

Species	Locality	Date	Sample size	Abbreviation
Gonioctena olivacea	Wibrin, Belgium (B)[a]	August 1999	9	GoliB1–GoliB9
Gonioctena olivacea	Hald, Denmark (DK)	July 1992	1	GoliDK
Gonioctena olivacea	Sines, Portugal	May 1993	1	GoliPT
Gonioctena pallida	Tshiertshen, Swiss Alps	June 1992	1	GpalS1
Gonioctena pallida	Les Hautères, Swiss Alps	June 1993	1	GpalS2[b]
Gonioctena pallida	Oberried, Black Forest, Germany	May 1999	2	GpalDO1[b]–GpalDO2[b]
Gonioctena pallida	Wieden, Black Forest, Germany	May 1999	2	GpalDW1[b]–GpalDW2
Gonioctena pallida	Ste-Odile, Vosges, France	May 1999	2	GpalFS1[b]–GpalFS2[b]
Gonioctena pallida	Orbey, Vosges, France	May 1999	2	GpalFO1–GpalFO2[b]
Gonioctena pallida	Giromagny, Vosges, France	May 1999	1	GpalFG2[b]
Chrysomela lapponica	Queyras, France	August 1995	1	ClapF
Chrysomela lapponica	Finland	August 1995	1	ClapFi

[a] Individuals sampled in five populations located in an area of 5×2 km.
[b] Repetitive region not sequenced for this sample.

Because the cloning–transposon insertion strategy is labor-intensive, the repetitive region was sequenced for only 3 *G. pallida* individuals of 11 (GpalS1, GpalDW2, and GpalFO1) and for the 2 *C. lapponica* individuals.

Data Analyses

The edges of the newly determined control regions were identified by aligning the flanking 12S rRNA and isoleucine tRNA sequences with their homologues from *Drosophila* (Diptera), *Locusta* (Orthoptera), and *Triatoma* (Hemiptera), retrieved from GenBank (accession numbers AJ400907, NC001712, and NC002609). Control-region sequences were edited with SEQPUP version 0.6 (Gilbert 1996) and aligned using Clustal X (Thompson et al. 1997) with default parameter settings (gap opening/extension penalties of 15.00/6.66). The stability of these alignments was explored further using SOAP version 1.1 (Löytynoja and Milinkovitch 2001). This program automatically generates several Clustal alignments, using a specified range of parameter settings, and allows easy comparison of the resulting alignments to detect unstable aligned positions. The proportion of identical sites among sequences and nucleotide frequencies were calculated using PAUP* (Swofford 2001). Haplotype networks were inferred with the parsimony criterion using the method described by Templeton et al. (1992) and implemented in TCS V. 1.13 (Clement et al. 2000). Phylogenetic networks are more convenient to represent phylogenetic relationships among closely related sequences than strictly bifurcating trees, as the former allow the display of all equally parsimonious hypotheses (i.e., ambiguous relationships) on a single figure. In addition to the Clustal alignments, comparisons of sequences were performed by searching blocks of similarity across the entire control region using the program Macaw version 2.0.5 (Schuler et al. 1991; available by FTP at ncbi.nlm.nih.gov in the directory pub/Macaw). This program locates regions of local similarity across sequences and evaluates their statistical significance (Karlin and Altschul 1990). These searches were performed within and among leaf beetle species, but also between leaf beetles and other insect orders.

Investigation of the sequence variability across the control region in pairwise comparisons of individuals at several hierarchical levels was performed with a 40-nucleotide-long sliding window, moving one base at a time across the pairwise alignment, each time returning a similarity value (i.e., the proportion of identical sites). When comparing highly divergent sequences, it was difficult to distinguish between true homology and chance similarity. We therefore generated, with MacClade 3.05 (Maddison and Maddison 1992), 100 pairs of random sequences of size and base composition identical to the two original sequences being compared. Using the same sliding window of 40 nucleotides, maximum similarity values were recorded for each of the 100 pairwise comparisons of random sequences. In pairwise comparisons of original experimental sequences, any similarity value higher than the maximum value recorded in pairwise comparisons of random sequences was considered significant at the 1% level. The evolution of the repetitive region (see below) was explored by inferring the phylogenetic relationships among the repeated elements within and among individuals. This was done (1) among repeats from all nine Belgian samples of *G. olivacea* (GoliB1–GoliB9), (2) among repeats from the three *G. pallida* samples for which we have sequenced the repetitive region (GpalS1, GpalDW2, GpalFO1), and (3) among repeats of three distantly related *G. olivacea* individuals (GoliB2, GoliDK, GoliPT). Phylogenetic networks of repeated elements were inferred with the parsimony criterion using the TCS program (see above).

Results and Discussion

All sequences are deposited in GenBank under accession numbers AY117367–AY117382, and AF530028–AF530043. The length of the control region is 3172–3296, 3417–3701, and 4640–5276 bp for *G. olivacea*, *G. pallida*, and *C. lapponica*, respectively. Mt control-region length differences among individuals of each species is essentially due to variation in a highly repetitive section of 12 to 17 imperfect copies of a 107- to 159-bp-long core sequence. The presence of one or several such repetitive segments has already been reported for many animal species for which the sequence variation of the control region has been investigated. These repeated elements represent 51–53, 49–53, and 51–57% of the control region in *G. olivacea*, *G. pallida*, and *C. lapponica*, respectively. Examples of these repeated elements for all three species are available from the authors upon request or can be found at the following URL: http://www.ulb.ac.be/sciences/ueg/html_files/publications.html. The sequence variability among the elements within the repetitive section of a given individual depends on the species considered: the tandemly repeated elements within each *G. olivacea* control-re-

gion sequence are much more variable than those within *G. pallida* or *C. lapponica* sequences. Using Macaw 2.0.5, we were not able to find nonrandom similarity among the repetitive sections from different species. All sequences reported here were obtained through procedures requiring PCR amplification from genomic DNA. Hence, all ambiguities (after direct sequencing of PCR products or in contigs of sequences obtained from different clones of the same PCR product) were reported with the appropriate symbol. These ambiguities could correspond to heteroplasmy and/or to cloning artifacts. Identification of the presence/absence of heteroplasmy in the investigated leaf beetles is clearly beyond the scope of the present paper.

Different publications have recently reported that, in a wide variety of animal species, large stretches of mitochondrial DNA (mtDNA) have been transferred to the nucleus (e.g., Collura and Stewart 1995; Bensasson et al. 2001). Although we cannot be certain that the DNA we have sequenced for this study is not of nuclear origin, we believe that this is unlikely. Indeed, the PCR-amplified product on which all further primer design was based in our study is 6–8 kb in length. Amplifying a large portion of the mtDNA molecule is one advised method (see http://www.pseudogen.net/) to avoid amplification of nuclear copies of mtDNA or, at least, to increase the proportion of mt relative to nuclear amplifications. If the proportion of nuclear copies in the PCR product is sufficiently small, direct sequencing of this product will yield mitochondrial sequence. Only when sequencing the repetitive region of *G. pallida* and *C. lapponica* was a cloning strategy used (see Materials and Methods). In this case, however, we have cloned a PCR product of the entire control region (3400–5200 nucleotides). At least three clones per fragment were sequenced and compared, and a consensus sequence was built.

A + T account for 83, 82, and 75% of the nucleotides in the *G. olivacea*, *G. pallida*, and *G. lapponica* mt control regions, respectively. This represents the lowest control-region A + T content of all insects studied thus far (Zhang and Hewitt 1997).

The stability of Clustal alignments was tested using SOAP version 1.1. All alignments of intraspecific sequences, when excluding the repetitive section, are very stable, i.e., virtually no alignment differences are observed when using different alignment parameters (weighted matrix, gap penalties from 12 to 20 by steps of 2, and extension penalties from 4 to 10 by steps of 2). This is not surprising given the relatively low level of divergence observed at the intraspecific level. On the other hand, alignment of the repetitive section is more problematic, as it is difficult to establish orthology among repeats from different individuals (see below). SOAP analyses identified that the alignments among sequences (even outside the repetitive region) from different species or genera are highly unstable. Still, we have checked that all conclusions drawn below are unaffected by the chosen alignment parameters.

Control-region sequence variations at different hierarchical levels are summarized in Fig. 1. Only one example of pairwise comparison is shown for each hierarchical level, as the same pattern is observed for all comparisons at one given level. The proportion of identical sites found in Clustal pairwise alignments of sequences from different genera (*Gonioctena* and *Chrysomela*) using a 40-bp sliding window is mostly nonsignificantly different, i.e., below 65% (Fig. 1A), from what can be expected between random sequences. Searching for blocks of similarity using Macaw 2.0.5, we found very few significant shared patterns: two long stretches of A's (5′ATTAAAAAAAAA3′ and 5′AWAAAAAAAAAAAAACT3′, similarly located in all three species) and different $[T(T)A(A)]_{6–22}$ repeats. Comparable patterns are also found in the control region of species from other insect orders (Zhang and Hewitt 1997). For example, in *Drosophila* species, two conserved thymidylate stretches (19–25 and 13–17 nucleotides long) have been observed, one of which is adjacent to a conserved block of 300 bp that has been proposed to be involved in mtDNA replication (Lewis et al. 1994). Whether these patterns are functionally constrained in leaf beetles or correspond to chance similarity is unclear. No blocks of similarity other than the ones described above were detected in this study when comparing leaf beetle sequences with control-region sequences from other insect species.

The pairwise Clustal alignment of the full control region of the two *Gonioctena* species investigated here yields an overall uncorrected similarity of 71% and exhibits many blocks with significant similarity, mainly in the first half of the control region (Fig. IB).

In intraspecific comparisons, the proportion of identical sites found in pairwise alignments of sequences using a 40-bp sliding window is always above the estimated level of random similarity and is mostly above 90% (Figs. 1C–F). Within *G. olivacea*, the mt control regions of the three distantly located populations in Europe (Belgium–Denmark–Portugal) have an overall similarity of 95–96%, with much of the variation located in the repeated elements region (Fig. 1E; sites 1180 to 2912). This pattern of variation across the control region is confirmed when comparing sequences from populations at a lower geographic scale: variable sites within populations sampled in the Belgian Ardennes (within an area of 5×2 km) are almost exclusively restricted to the repetitive section of the control region (Fig. IF). Parsimony network analyses (Fig. 2) also indicate that the number of mutations between pairs of haplotypes

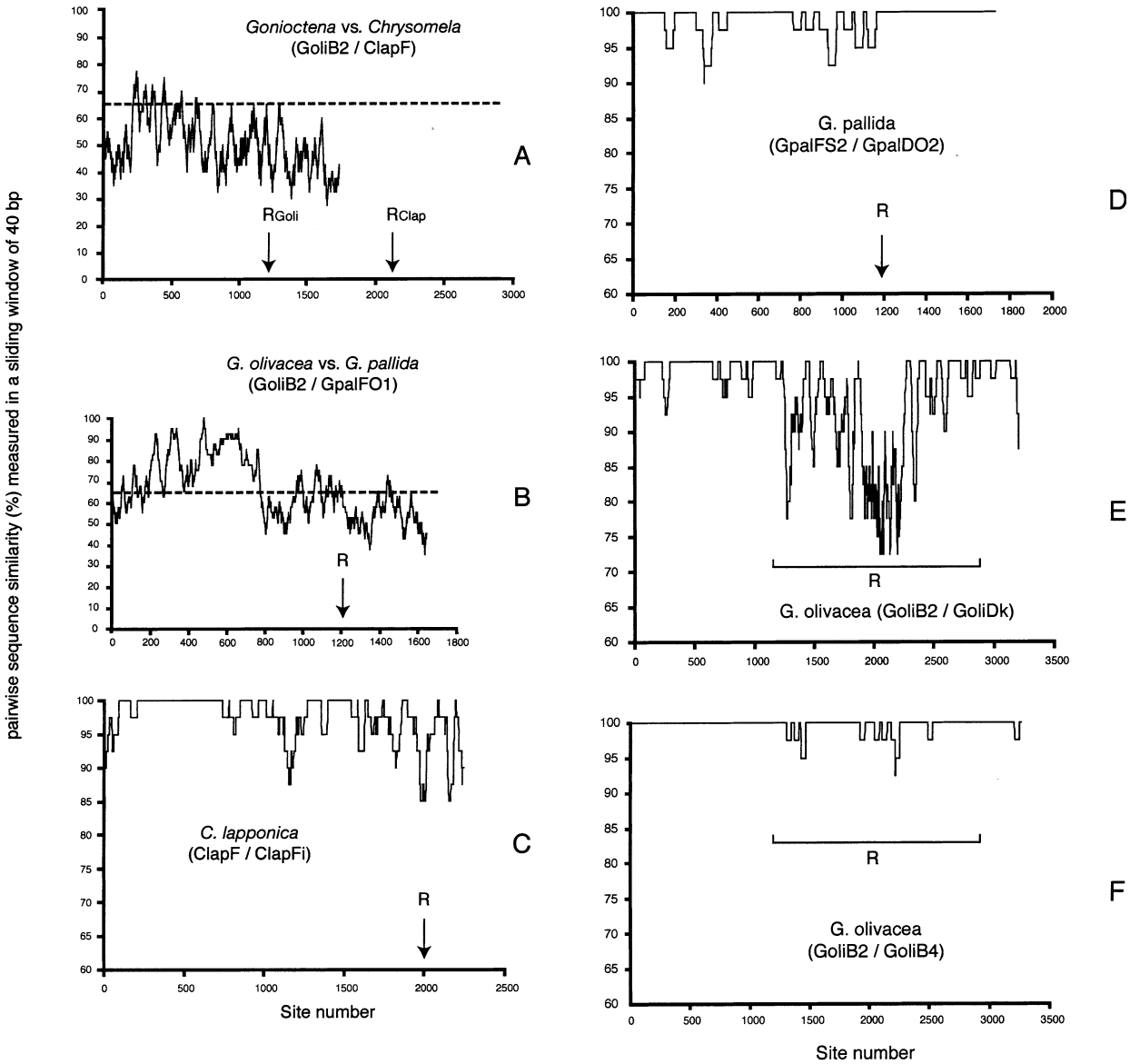

Fig. 1. Pairwise sequence similarity of mt control-region sequences at different taxonomic/population levels. Uncorrected pairwise sequence similarity is measured in a sliding window of 40 bp. The scale on the horizontal axis shows the nucleotide position of the first site inside the sliding window. Comparisons are made between (**A**) GoliB2 and ClapF, after removal of the repeated elements, (**B**) GoliB2 and GpalFO1, after removal of the repeated elements, (**C**) ClapF and ClapFi, after removal of repeated elements and indel sites, (**D**) GpalFS2 and GpalDO2, after removal of repeated elements and indel sites, (**E**) GoliB2 and GoliDK, after removal of indel sites but including repeated elements, (**F**) GoliB2 and GoliB4, after removal of indel sites but including repeated elements. Sequences are presented from the methionine tRNA end (*left*) to the 12S end (*right*). The *horizontal dashed line* in A and B indicates the similarity value assessed as significantly different from random ($p < 0.01$; see Materials and Methods). *Arrows* indicate the location of the repetitive region when it was excluded (A–D); *brackets* indicate the range spanned by the repetitive region when it was included (E and F).

is much higher when including the repeated elements. The high variability of the repetitive section of the mt control region may not be the result of point mutations only. Indeed, when comparing individuals from distant populations of *G. olivacea* (e.g., GoliDK and GoliB2; see examples available at http://www.ulb.ac.be/sciences/ueg/html_files/publications.html), more variation is often observed between repeats of two individuals than among repeats within an individual. For example, there are 22 differences (including point mutations and indels) between repeats 8 from GoliDK and GoliB2, while, within GoliDK, repeat 8 is identical to repeat 5 (which is not the case in GoliB2). These observations suggest that repeats within the control region of a given individual might have been generated through recurrent and recent (post-dating the common ancestor of the individuals compared) duplication/deletion events. Such a mechanism could lead to the partial homogenization of the repeats within individuals (i.e., concerted evolution) and explain the much higher conservation of repeat sequences within than among *G. pallida* individuals.

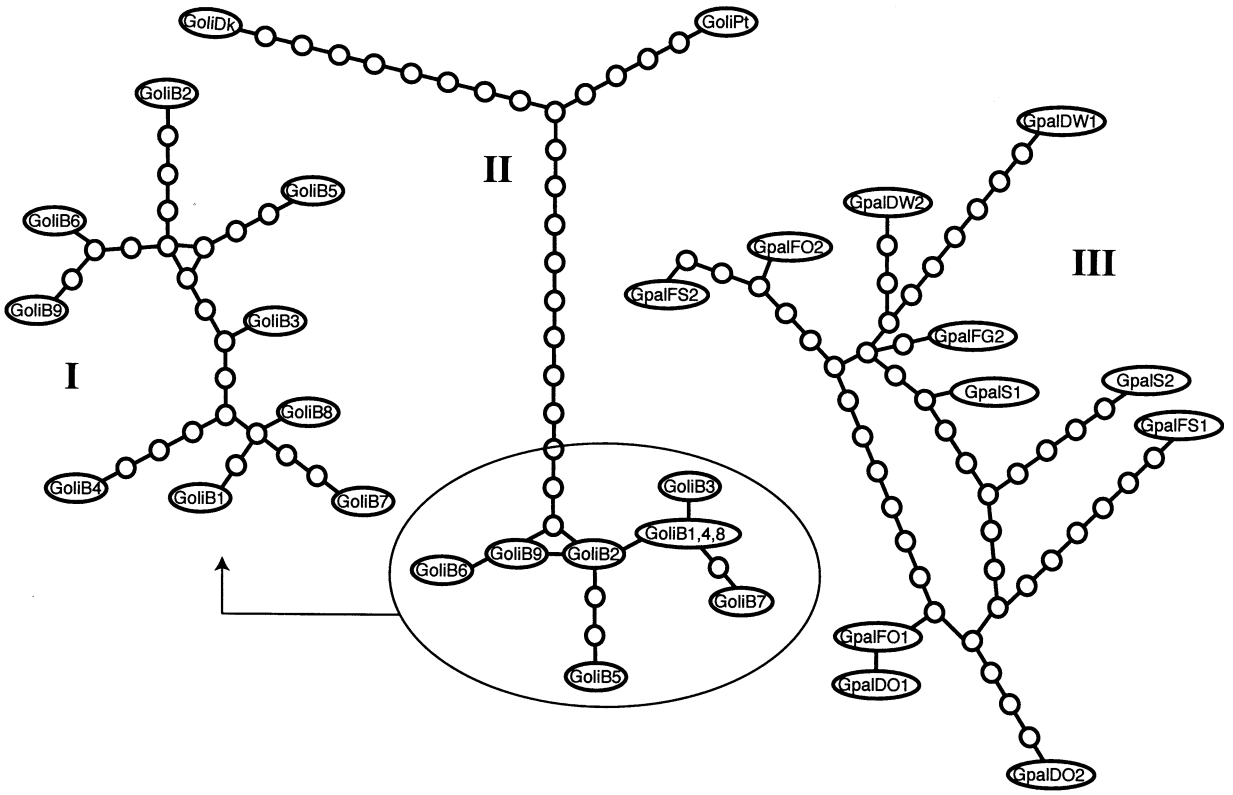

Fig. 2. Networks showing the most parsimonious relationships among control-region sequences inferred with the TCS program of Clement et al. (2000). (**I**) *Gonioctena olivacea* individuals collected within five populations inside a 5 × 24-km area. (**II**) all *Gonioctena olivacea* individuals collected in Europe; network inferred after removing the repeated elements, (**III**) All *Gonioctena pallida* individuals collected in the Vosges mountains, the Black Forest mountains, and the Alps; network inferred after removing the repeated elements. In network II, a *circle* is drawn around the haplotypes that are displayed in network I.

Also, diagnostic single-nucleotide substitutions are often present in more than one repeat, as if they had propagated across different repeats within an individual. These substitutions could be homoplasious and result from the presence of mutational hot spots in the sequence or could simply reflect recurrent duplications/deletions of repeats as suggested above. The second explanation is more likely, as those single-nucleotide mutations are often covariating with each other.

We tested the occurrence of concerted evolution further by inferring the phylogenetic relationships among repeated elements within and among individuals (see Materials and Methods). In the absence of duplication/deletion events, we would expect that each repeat in one individual clusters with one (and only one) closely related repeat from each of the other individuals, i.e., duplication events would have occurred before the divergence of all investigated individuals from their most recent common ancestor. On the other hand, if concerted evolution does occur, we would expect to observe some clustering of repeats from the same individual. The inference of phylogenetic relationships among repeated elements from all nine Belgian samples of *G. olivacea* led to no evidence of concerted evolution. Indeed, in the resulting phylogenetic network (which can be seen at http://www.ulb.ac.be/sciences/ueg/html_files/publications. html), each repeat in one individual can be grouped with the corresponding orthologous repeats (here, two repeats are said to be orthologous if they have been generated by vertical descent from a common ancestral genome rather than by duplication within that genome) in all other individuals.

In contrast, evidence for concerted evolution can be found at a higher geographic level, i.e., when inferring phylogenetic relationships among repeated elements from three *G. pallida* samples in Europe or among three distantly related *G. olivacea* individuals. In the resulting networks, we can indeed find several examples of clusters of repeated elements from the same individual. One of the most conspicuous examples is shown in Fig. 3 (arrow), where one block of identical repeats (*i*) includes five repeats from individual GpalFO1 (black rectangles), (*ii*) is directly connected to four or five other GpalFO1 repeats, (*iii*) includes only one repeated element from GpalS1 (white rectangles) and one from GpalDW2 (gray rectangle), and (*iv*) is directly connected to only two GpalS1 repeats and to one GpalDW2 repeat. Another obvious example can be found in the phylogenetic network of the three distantly related *G. olivacea*

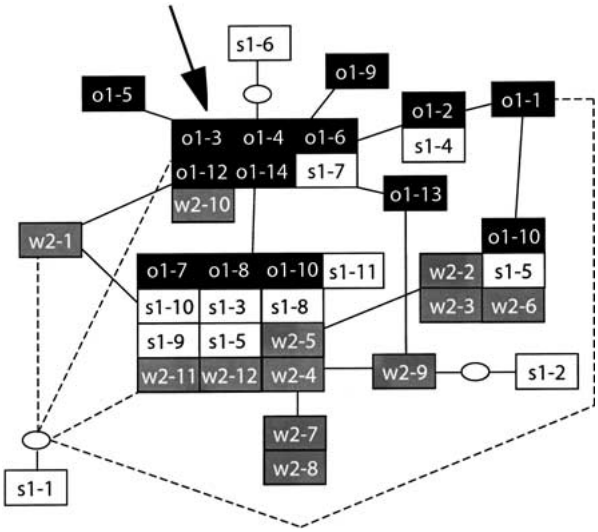

Fig. 3. Network showing the most parsimonious relationships among repeated elements from three *G. pallida* individuals (s1 = GpalS1, o1 = GpalFO1, w2 = GpalDW2, from the Alps, Vosges, and Black Forest, respectively), inferred using the statistical parsimony approach with the program TCS 1.13 from Clement et al. (2000). Each repeat is represented by a *rectangle* and is named after the individual it comes from, followed by a number identifying the position of the repeat in the sequence (e.g., w2-5 refers to repeat number 5 in individual GpalDW2). Repeat elements not separated by a branch represent identical sequences. Each branch represents a single mutational step, i.e., one nucleotide substitution or one insertion/deletion event. Open ovals represent missing (i.e., unobserved) intermediate sequences. The *dashed lines* represent four alternative equally parsimonious ways to connect s1-1 to the rest of the network. The *arrow* identifies a group of repeats that are referred to in the text.

individuals (available at http://www.ulb.ac.be/sciences/ueg/html_files/publications.html), where four repeated elements from GoliPT can be isolated from the rest of the network. This strongly suggests the occurrence of duplication/deletion of repeated elements in the control region of *Gonioctena* leaf beetles, leading to the homogenization of the repetitive segment.

The occurrence of duplication and deletion events of repeated elements in the mitochondrial control region has already been encountered in many animal species. The molecular mechanism(s) responsible for the evolution of the repeated elements is(are) currently not fully understood, although slipped-strand mispairing during replication—i.e., frequent misalignment in the repeat region, prior to elongation, causing length heteroplasmy (Levinson and Gutman 1987; Buroker et al. 1990)—is believed to be involved (Lunt et al. 1998). Homologous recombination has also been cited as a possible mechanism (e.g., Rand and Harrisson 1989), for which empirical evidence was recently found (e.g., Lunt and Hyman 1997; Awadalla et al. 1999; Ladoukakis and Zouros 2001).

The occurrence of concerted evolution in the leaf beetle control region makes it particularly difficult to differentiate orthologous from paralogous repeats among distantly related individuals. The repeated elements characterized here will be of particular interest for population genetic and phylogenetic analyses of *G. olivacea* involving closely related individuals because, at this taxonomic level, not only can orthology among repeats be unambiguously assessed, but also repeats exhibit useful levels of variation, even among individuals collected inside areas as restricted as 5×2 km. However, the usefulness of the repetitive region depends on the species considered. Indeed, as already mentioned, the level of repeat sequence variation strongly varies among the three species investigated here.

P distances (i.e., the proportion of sites that are different between two aligned sequences) among control-region sequences of *G. pallida* individuals from different locations (the Alps, the Vosges mountains, and the Black Forest) range from 0.1 to 0.9%. The repeated elements, available for only three individuals (see Materials and Methods), were not included in this calculation, as it is difficult to establish orthology among repeats for *G. pallida* (because repeats are too similar across the control region of a single individual). Interestingly, the haplotype network in Fig. 2 (III) indicates no clear separation among the individuals from the Vosges, the Black Forest, and the Alps, despite large low-elevation areas (i.e., unsuitable habitats for *G. pallida*) separating these mountain ranges. Indeed, the two alpine individuals (GpalS1 and GpalS2), for which the sampling sites are located several hundreds of kilometers apart, are not genetically more distant from each other or from the Vosges and Black Forest individuals (all sampled within a geographical area of 100×40 km) than the latter individuals are from each other.

The pairwise alignment between the control-region sequences of the two *Chrysomela lapponica* individuals yields an uncorrected sequence similarity of 96.9%. When excluding the repeated elements region, for which it is difficult to establish homology among repeats, this value increases to 97.6%. Figure 1C shows that the differences between the two sequences are not homogeneously distributed, as there are many more differences located at the end of the sequence than there are in the first 800 bp of the control region.

Although it has been shown previously in some insect species that the substitution rate is lower in the control region than in the third positions of some protein coding genes (Caccone et al. 1996; Zhang and Hewitt 1997), the rates of evolution inferred in this study are highly variable across the control region and some specific portions of this mtDNA fragment reach levels of variation clearly sufficient for population genetic or phylogeographic studies. For example, one of us characterized a portion of about 400 nucleotides (Fig. 2D; sites 720–1120) as effective for conducting a phylogeographic study of *G. pallida*

populations within the Vosges mountains (Mardulyn 2001). When focusing on this 400-nucleotide region, the sequence divergence among *G. pallida* individuals increase to 2%.

Acknowledgments. Individuals of *Chrysomela lapponica* and of *Gonioctena olivacea* from Portugal were collected by Jacques Pasteels, and the *Gonioctena olivacea* individual from Denmark by J. Nielsen. We are grateful to two anonymous reviewers for providing many useful comments on an early version of the manuscript. This research was supported by the FNRS (Belgian 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 a Postdoctoral Researcher at the FNRS.

References

Awadalla P, Eyre-Walker A, Maynard Smith J (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. Science 286:2524–2525

Bensasson D, Zhang D-X, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: Evolution's misplaced witnesses. TREE 16:314–321

Boyce TM, Zwick, ME, Aquadro CF (1989) Mitochondrial DNA in the bark weevils: Size, structure and heteroplasmy. Genetics 123:825–836

Buroker NE, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics 124:157–163

Caccone A, Garcia BA, Powell JR (1996) Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. Insect Mol Biol 5:51–59

Clement M, Posada D, Crandall KA (2000) TCS: A computer program to estimate gene genealogies. Mol Ecol 9:1657–1659

Collura RV, Stewart C-B (1995) Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. Nature 378:485–489

Fauron CM-R, Wolstenholme DR (1976) Structural heterogeneity of mitochondrial DNA molecules within the genus *Drosophila*. Proc Natl Acad Sci USA 73:3623–3627

Gilbert DG (1996) SeqPup version 0.6. Published electronically on the Internet; available via anonymous ftp to ftp.bio.indiana.edu

Goddard JM, Wolstenholme DR (1978) Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. Proc Natl Acad Sci USA 75:3886–3890

Karlin S, Altschul SF (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc Natl Acad Sci USA 87:2264–2268

Ladoukakis ED, Zouros E (2001) Direct evidence for homologous recombination in mussel (Mytilus galloprovincialis) mitochondrial DNA. Mol Biol Evol 18:1168–1175

Levinson G, Gutman GA (1987) Slipped-strand mispairing: A major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994) Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. Mol Biol Evol 11:523–538

Löytynoja A, Milinkovitch MC (2001) SOAP, cleaning multiple alignments from unstable blocks. Bioinformatics 17:573–574

Lunt DH, Hyman BC (1997) Animal mitochondrial DNA recombination. Nature 387:247

Lunt DH, Whipple LE, Hyman BC (1998) Mitochondrial DNA variable number of tandem repeats (VNTRs): utility and problems in molecular ecology. Mol Ecol 7:1441–1455

Maddison WP, Maddison DR (1992) MacClade: Analysis of phylogeny and character evolution (version 3.05). Sinauer Associates, Sunderland, MA

Mardulyn P (2001) Phylogeography of the vosges mountains populations of *Gonioctena pallida* (Coleoptera: Chrysomelidae): a nested clade analysis of mitochondrial DNA haplotypes. Mol Ecol 10:1751–1763

McMillan WO, Palumbi SR (1997) Rapid rate of control region evolution in Pacific butterflyfishes (Chaetodontidae). J Mol Evol 45:473–484

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. Annu Rev Ecol Syst 18:269–292

Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome size variation. J Mol Evol 37:281–295

Rand DM, Harrison RG (1989) Molecular population genetics of mtDNA size variation in crickets. Genetics 121:551–569

Rüber L, Meyer A, Sturmbauers C, Verheyen E (2001) Population structure in two sympatric species of the lake Tanganyika cichlid tribe Eretmodini: Evidence for introgression. Mol Ecol 10:1207–1225

Schuler GD, Altschul SF, Lipman DJ (1991) A workbench for multiple alignment construction and analysis. Proteins 9:180–190

Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem 66:409–433

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. Ann Entomol Soc Am 87:651–701

Swofford DL (2001) PAUP*: Phylogenetic analysis using parsimony (and other methods), (version 4.0b8). Sinauer Associates, Sunderland, MA

Taylor MFJ, McKechnie SW, Pierce N, Kreitman M (1993) The lepidopteran mitochondrial control region: structure and evolution. Mol Biol Evol 10:1259–1272

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Zhang D-X, Hewitt GM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 25:99–120