Incongruent nuclear and mitochondrial phylogeographic patterns in the *Timarcha goettingensis* species complex (Coleoptera, Chrysomelidae)

J. GÓMEZ-ZURITA*† & A. P. VOGLER*†

*Department of Entomology, The Natural History Museum, London, UK
†Department of Biological Sciences, Imperial College at Silwood Park, Ascot, Berkshire, UK

Keywords:

cytochrome oxidase II;
gene tree–species tree problem;
hybrid zones;
incongruence;
nested geographical clade analysis;
nuclear rDNA ITS2;
phylogeography.

Abstract

Phylogeographic analyses have mostly been based on single-gene genealogies but it is unclear how conclusions from such studies depend on the choice of gene markers. We conducted a nested geographical clade analysis [A.R. Templeton, E. Routman & C.A. Phillips (1995) *Genetics* **140**: 767–782] based on nuclear rDNA internal transcribed spacer region 2 (ITS2) sequences in the *Timarcha goettingensis* species complex (Coleoptera, Chrysomelidae), and compared the inferences with an updated version of previously published results using mitochondrial cytochrome oxidase II (COII) sequences. Inferences from ITS2 suggest that patterns of marker distribution are mostly explained by restricted gene flow with isolation by distance. In contrast, COII revealed a history of geographical structure resulting from episodic population contiguous-range expansions. Both markers also show different genealogical patterns, which are associated to the effects of genetic introgression in a putative hybrid zone between two major lineages in the complex. Altogether, these differences are attributed to distinct population and/or evolutionary dynamics of the markers, and offer a more accurate phylogeographic description for the *T. goettingensis* complex.

Introduction

Phylogeographic analyses to date have mostly been based on genealogies of mitochondrial markers (Avise, 2000). However, phylogeographic inferences based on single or genetically linked markers are hampered by processes of stochastic lineage sorting of haplotypes, hybridization, introgression or horizontal transfer, potentially leading to differences in the history of each marker (Nei, 1987; Brower *et al.*, 1996; Doyle, 1997; Maddison, 1997). Therefore it is desirable to use the evidence from several independent markers for a detailed phylogeographic perspective of a group of organisms. For instance, the information provided by markers from different genomic compartments, e.g., nuclear and cytoplasmic, showing discordant phylogeographic patterns could prove differences in historical dispersal among sexes. Such examples of incongruent phylogeographic patterns have been demonstrated, among others, for humans (Hammer *et al.*, 1998; Seielstad *et al.*, 1998; Karafet *et al.*, 1999; Pérez-Lezaun *et al.*, 1999) and *Heliconius* butterflies (Davies & Bermingham, 2002).

In the study of phylogeography, nested geographical clade analysis (NGCA) has proved to be a powerful analytical tool (Templeton *et al.*, 1995; list of references using this methodology for phylogeographic purposes in Gómez-Zurita, 2002). This methodology combines genealogical and geographical data for inferences of historical range distribution of haplotypes and range movements. However, inferences resulting from the application of NGCA to a set of sequences rely on a prior estimation of the genealogy of haplotypes. If these inferences are incongruent because of problems with

Correspondence: Jesús Gómez-Zurita, Molecular Systematics Laboratory, Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.
Tel.: +44 (0)207 942 5609; fax: +44 (0)207 942 5229;
e-mail: j.gomez-zurita@nhm.ac.uk

phylogenetic reconstruction or true differences in the phylogenetic history of various genes, the choice of markers will strongly affect the conclusion from this analysis in inferring demographical and phylogeographic population histories. Congruence and incongruence in multiple markers can be exploited to improve conclusions about historical geographical processes, in particular where markers with different modes of inheritance are used.

Here we are using leaf beetles in the genus *Timarcha* for such study of phylogeography from multiple markers. The *T. goettingensis* species-complex consists of an undetermined number of taxa with allopatric or parapatric populations mainly distributed in the Iberian Peninsula, and extending into Central Europe and the British Isles (Petitpierre, 1973). Gómez-Zurita *et al.* (2000a) obtained a comprehensive population sample of this species from the Iberian Peninsula based on mitochondrial cytochrome oxidase II (COII) sequences which were subjected to NGCA. This analysis revealed the existence of several lineages between which gene flow was restricted or absent, with short-range expansions as the most likely explanation for the observed geographical association of haplotypes. The two most divergent lineages in the complex, defining the western (central/north Iberian Peninsula) and eastern (north-eastern Iberian Peninsula to central Europe) clades were separated by an ancient fragmentation event according to this analysis (Gómez-Zurita *et al.*, 2000a,b).

The goal of the current study is to test the performance of NGCA in the *T. goettingensis*-complex employing a nuclear marker, the internal transcribed spacer region 2 (ITS2) of the rRNA cluster. The comparison with the existing COII data will improve our picture of the historical processes explaining the current distribution of genetic variation in the Iberian populations. In particular, a nuclear marker will clarify questions about hybridization of divergent lineages and potential cases of mitochondrial introgression, which have been hypothesised (Gómez-Zurita *et al.*, 2000a,b).

Material and methods

Samples

Geographical locations of all studied populations of the *T. goettingensis* complex are shown in Fig. 1. Table 1 includes the locality names and geographical co-ordinates for each sampled population along with their taxonomic assignment. Samples include 133 individuals and 31 localities studied in Gómez-Zurita *et al.* (2000a), plus six additional localities and 34 additional specimens.

DNA isolation, ITS2-PCR and sequencing

DNA was isolated using a standard phenol : chloroform extraction protocol. One microlitre of the DNA stock solution was used for polymerase chain reaction (PCR) amplification of ITS2 using the universal primers ITS3 (5′-GCATCGATGAAGAACGCAGC-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) (White *et al.*, 1990). PCR conditions were 4 min at 95 °C followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 50 °C for 30 s, and extension at 72 °C for 1 min, with a final extension step at 72 °C for 10 min. PCR products were checked in a 1.5% agarose gel and the products of the expected length were purified using the GeneClean II Kit (Bioline, La Joya, CA, USA). ITS2 PCR products were sequenced in both directions with the same primers used for amplification using the BigDye™ Terminator Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). Sequencing reaction products were subsequently ethanol precipitated and run on an ABI PRISM™ 3700 DNA Analyzer (Applied Biosystems). The COII data set represents 131 sequences of a fragment of 354 bp from Gómez-Zurita *et al.* (2000a) and was completed by 34 additional sequences.

Preparation of sequences for analysis

ABI sequence chromatograms were edited and complementary strands for each ITS2 PCR product were

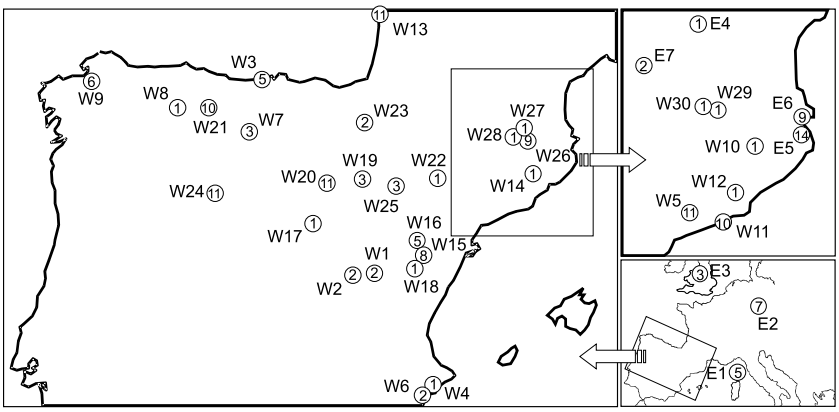

Fig. 1 Geographical location and number of specimens of the samples included in this study. Each locality is identified with a 'W' or a 'E' to identify the western and eastern clades, respectively.

Table 1 Sampling localities, geographical decimal co-ordinates, taxonomic assignment of samples, observed internal transcribed spacer region 2 (ITS2) haplotypes and their frequencies (in parenthesis).

Taxon	Sample site*	Co-ordinates	ITS2 haplotype
Western clade			
Timarcha aurichalcea	Guadalaviar (W1)	40.40, −1.72	1 (1), 2 (1)
	Tragacete (W2)	40.35, −1.85	3 (2)
T. cyanescens	Cueto (W3)	43.48, −3.80	4 (1), 5 (2), 6 (2)
T. fallax	Serra d'Aitana (W4)	38.65, −0.32	7 (1)
	Alió (W5)	41.28, 1.30	8 (5), 9 (3), 10 (1), 11 (1), 12 (1)
	Puerto de la Carrasqueta (W6)	38.61, −0.48	13 (1), 14 (1)
T. geniculata	Peña Amaya (W7)	42.67, −4.17	15 (3)
	Puerto de Ventana (W8)	43.05, −6.00	16 (1)
T. gougeleti	Fiobre-Bergondo (W9)	43.33, −8.23	17 (6)
T. interstitialis	L'Esquirol (W10)	42.03, 2.37	18 (1)
	Garraf (W11)	41.27, 1.97	8 (8), 19 (1), 20 (1)
	Sant Just Desvern (W12)	41.37, 2.08	21 (1)
T. maritima	Le Canon (W13)*	44.68, −1.23	22 (11)
T. monserratensis	Collformic (W14)	41.80, 2.33	23 (1)
T. perezi	Puerto de Cuarto Pelado (W15)	40.55, −0.55	24 (2), 25 (2), 26 (1), 27 (1), 28 (1), 29 (1)
	Ejulve (W16)	40.78, −0.53	28 (1), 30 (1), 31 (1), 32 (1), 33 (1)
	Layna (W17)	41.10, −2.30	34 (1)
	Puerto de Linares (W18)	40.33, −0.53	24 (1)
	El Moncayo (W19)	41.80, −1.85	35 (2), 36 (1)
	Puerto de Oncala (W20)	41.95, −2.32	37 (9), 38 (1), 39 (1)
	Puerto de San Isidro (W21)*	43.07, −5.38	40 (9), 41 (1)
	Sariñena (W22)*	41.78, −0.15	42 (1)
	Tafalla (W23)*	42.52, −1.67	43 (1), 44 (1)
	Villanubla (W24)	41.72, −4.83	45 (2), 46 (2), 47 (1), 48 (4), 49 (1)
	Zaragoza (W25)	41.62, −0.92	50 (2), 51 (1)
T. sinuatocollis	Coll de la Creueta (W26)	42.32, 1.98	52 (9)
	Vall d'Eina (W27)*	42.42, 2.13	53 (1)
	Vall de Núria (W28)	42.40, 2.13	54 (1)
	Queixans (W29)	42.40, 2.07	23 (1)
	Viliella (W30)	42.42, 1.72	52 (1)
Eastern clade			
T. cornuta	Zicavo (E1)*	41.90, 9.12	55 (3), 56 (2)
T. goettingensis	Jena (E2)	50.85, 11.52	57 (5), 58 (1), 59 (1)
	Stinchcombe (E3)	51.68, 2.38	60 (3)
T. interstitialis	Caramany (E4)	43.53, 1.75	61 (1)
	L'Escala (E5)	42.13, 3.13	62 (1), 63 (3), 64 (1), 65 (5), 66 (4)
	Sant Pere de Roda (E6)	43.32, 3.17	61 (2), 62 (4), 67 (1), 68 (1), 69 (1)
T. recticollis	Pla de l'Artiga (E7)	42.75, 0.75	70 (1), 71 (1)

*Samples identified with an asterisc were not included in the mitochondrial phylogeography of the complex by Gómez-Zurita *et al.* (2000a).

assembled using the Sequencher 4.1 software (Gene Codes Corp, Ann Harbor, MI, USA). Sequences were preliminarily aligned using the Clustal W algorithm (Thompson *et al.*, 1994) and a manual refinement was necessary to account for several variable microsatellite-like repeated motifs among the sequences (see Gómez-Zurita *et al.*, 2000c for a detailed description). From the refined alignment, all microsatellites were recoded so that each di- or trinucleotide repeat unit would be treated as a single mutational event; other indels in the alignment were also recoded as single mutational events. A recoding strategy is necessary in this case to avoid branch lengthening in the networks due to considering each position of the microsatellites as independent mutations.

Each distinct ITS2 sequence obtained has been deposited in the EMBL database under accession numbers AJ512523–AJ512608. Additional COII sequences newly generated for this study have been assigned the accession numbers AJ512512–AJ512522.

Measures of genetic variation and of population structure

Sequence diversity indices, including number of segregating sites, number of haplotypes, haplotype diversity, nucleotide diversity, and gene flow (F_{ST}, Nm) were estimated using DnaSP version 3.50 (Rozas & Rozas, 1999).

Nested geographical clade analysis

The recoded ITS2 data matrix was used to calculate the 'limits of parsimony' (*sensu* Templeton *et al.*, 1992) and to obtain the marker genealogy applying statistical parsimony (Templeton *et al.*, 1992) as implemented in the software TCS version 1.3 (Clement *et al.*, 2000), coding gaps a fifth character state. The hierarchical structure of the tests for geographical association of haplotypes was defined on the ITS2 genealogy using the nesting rules described in Templeton *et al.* (1987) and additional rules to deal with ambiguity (Templeton & Sing, 1993). Permutational contingency tests and geographical distance analyses based on the nests design to test for alternative hypotheses of nonrandom spatial distribution of haplotypes were performed using GeoDis version 2.0 (Posada *et al.*, 2000). The interpretation of the significance patterns obtained for the statistics of interest, i.e. clade distances (Dc), nested clade distances (Dn), and the difference of both measures for interior and tip clades in the network ([I-T] Dc and [I-T] Dn, respectively), was performed using a revised version of the inference key proposed in Templeton *et al.* (1995) (http://bioag.byu.edu/zoology/crandall_lab/dposada/documents/). This inference key allows discriminating among several hypotheses potentially explaining an observed geographical association of haplotypes, including recurrent (restricted gene flow) and historical (population range expansions, past fragmentation of populations) processes, and also identifies weaknesses in sampling design to be addressed in future genetic analysis.

Results

ITS2 variability and statistical parsimony analysis

The entire ITS2 region flanked by small portions of the 5.8S and 28S rRNA genes was selected for analysis. This fragment ranged in size from 633 to 662 bp, and after sequence alignment and recoding of several variable di- and tri-nucleotide repeats and longer insertion/deletion features as single mutational events, the resulting matrix contained 639 characters (619 excluding gaps) distinguishing 71 ITS2 haplotypes from a sample of 167 individuals. The estimation of standard parameters of genetic variation required the exclusion of gaps, and hence only 65 haplotypes remained defined by 86 polymorphic sites, producing a value for haplotype diversity = 0.971 and nucleotide diversity = 0.015.

Statistical parsimony analysis of the 71 ITS2 haplotypes within the limits of parsimony (*sensu* Templeton *et al.*, 1992), i.e. the number of steps by which two haplotypes have a 95% statistical probability of being linked without homoplasy, was calculated to be 10 steps or less. This analysis resulted in four independent networks within which the connections between each of the haplotypes fulfilled this criterion (Fig. 2). Network I includes most of the haplotypes obtained from Iberian individuals, here referred to as the western clade. Network II links haplotypes from three Pyrenean localities and the haplotypes obtained from nonIberian populations (eastern clade). Two divergent haplotypes, III and IV, correspond to the two western-most localities W8 and W9, respectively.

Nested geographical clade analysis of ITS2 data and estimation of gene flow

The topology of the networks served as basis for establishing the hierarchical nesting design for subsequent statistical tests of geographical association of haplotypes. Figure 2 shows the three lowest levels of nesting, zero- (haplotypes), one- and two-step groups of haplotypes with both geographical and genetic variation, for which testing for the association of both variables is meaningful. Figure 3 shows a simplification of the same haplotype networks with the highest hierarchical levels of nesting (up to six-step groups). Figure 3 also shows how the four networks are related to each other through connections exceeding 10 mutational steps, i.e., with a probability lower than 95% for the connections being affected by homoplasy.

The results of permutational contingency tests implementing geographical distances between sample localities estimated from their geographical co-ordinates are provided as supplementary web material (Tables S1 and S2). Statistics of NGCA Dc, Dn, [I-T] Dc and/or [I-T] Dn, showed significant deviations from a situation of random geographical distribution of haplotypes when considering the entire network as shown in Fig. 3: Network I ($Dc = 213.92^{S}$; $Dn = 232.18^{S}$), Network II ($Dc = 489.28^{L}$; $Dn = 548.69^{L}$), Network III (haplotype 16: $Dc = 0$; $Dn = 502.35$) and Network IV (haplotype 17: $Dc = 0^{S}$; $Dn = 685.32^{L}$). The test for differences between interior and tip haplotypes for the total network, assigning the root on Network I, is highly significant as well ([I-T] $Dc = -214.70^{S}$; [I-T] $Dn = -331.68^{S}$). Historical or population structure inferences made based on these significant results using the modified inference key in Templeton *et al.* (1995) suggest 'restricted gene flow with isolation by distance' as the predominant process to explain patterns of haplotype distribution, plus a few cases of 'contiguous range expansion' (Table 2).

A high proportion of inferences made from the spatial structure of ITS2 haplotypes and their genealogical relationships is consistent with a population genetic model of isolation by distance, where gene flow is restricted but not prevented (Table 2). In those cases where the pattern is interpreted to indicate restricted gene flow and sample sizes are large enough to estimate population parameters with confidence, accepting that genetic exchange is recurrently occurring and that no reproductive incompatibilities exist among populations within the complex, we have calculated the level of

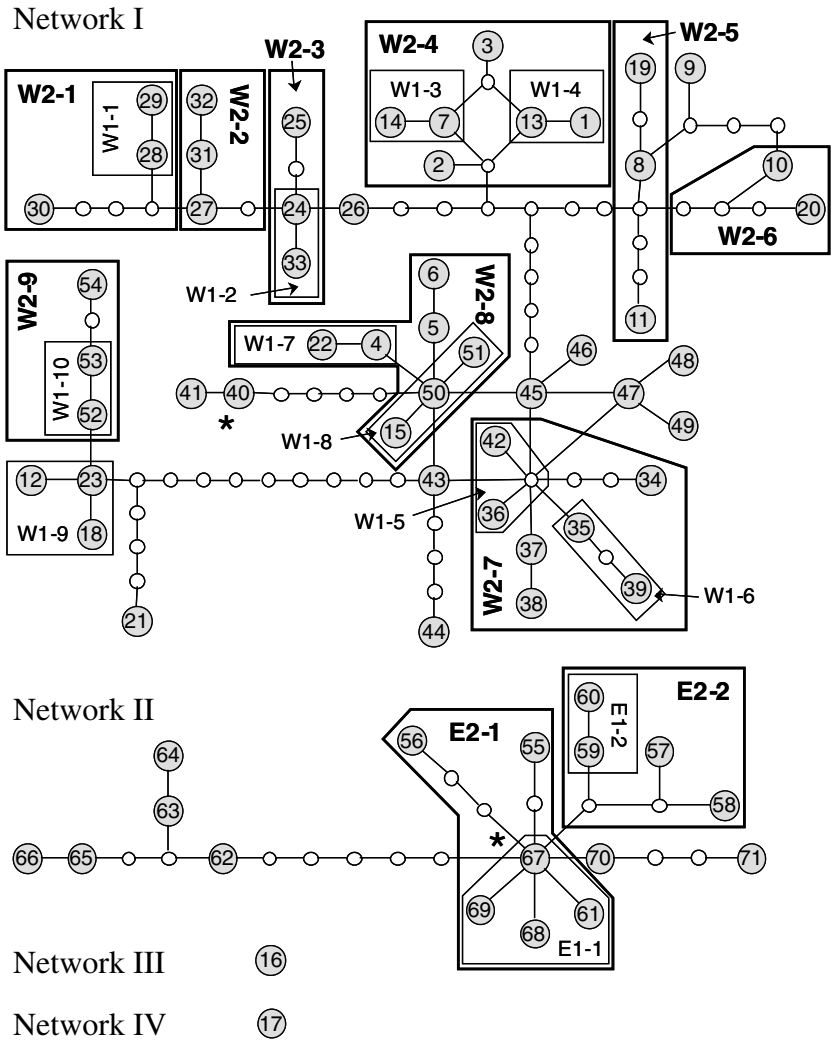

Fig. 2 The *Timarcha goettingensis* species complex internal transcribed spacer region 2 ribotype statistical parsimony networks and two first levels of the nesting design, one- and two-step groups, including both genetic and geographical variation ('W' and 'E' design nesting categories in the western and eastern lineages, respectively). Haplotype numbers are the same used in Table 1. An asterisk identifies the 'rooting' haplotype in each case, deduced from the results of the phylogenetic analysis of the genus (Gómez-Zurita *et al.*, 2000b,c).

population structuring and the associated migration rates both for ITS2 and COII, defining the populations as described by the nesting design on the ITS2 genealogy (Table 3). The estimated ITS2 F_{ST} values range between 0.36 and 0.84, and the range for COII is 0.14–0.62.

NGCA of COII data

Statistical parsimony network reconstruction, nesting design and interpretation of patterns based on NGCA were also conducted for COII using the expanded data matrix (not shown). The mitochondrial populations of the *T. goettingensis* complex show a highly significant geographical structure at all hierarchical levels of the haplotype network, and in most cases (11 of 17 informative inferences) the inferences of spatial structuring are 'contiguous range expansions'. One instance of 'restricted gene flow with isolation by distance' is found for the mitochondrial data set, for a group of Iberian populations in the eastern clade. All other inferences from the COII analysis indicate that the current spatial distribution of genetic variation is because of a series of historical events of population expansions.

Mitochondrial data separate two major allopatric clades with a distribution similar to that found for ITS2, i.e. a western group of haplotypes in the Iberian Peninsula and an eastern group distributed in the Northeast of the Peninsula and the rest of Europe. According to NGCA, this spatial distribution of COII haplotypes affecting the separation of the western and eastern clades can be related to a process of contiguous range expansion from the western to the eastern areas. This particular inference does not fully agree with the published results using the smaller COII data set, as the NGCA inference produced a signal for a past fragmentation event not recovered in the new analysis (Gómez-Zurita *et al.*, 2000a). However, the new inference regarding the separation of the eastern and western clades is in agreement with those from the

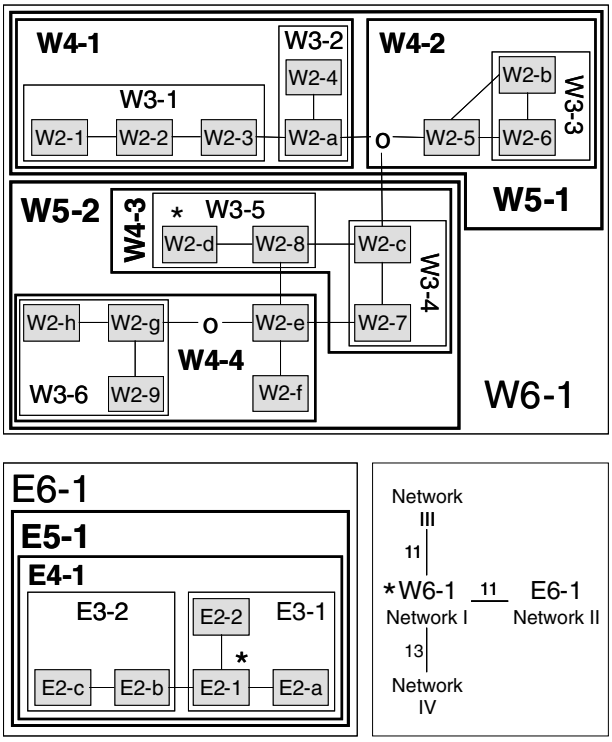

Fig. 3 Simplified statistical parsimony networks from Fig. 2 showing the four last nesting levels, three-, four-, five- and six-step groups, with both genetic and geographical variation ('W' and 'E' identify the western and eastern clades, mentioned on the text throughout). Bottom-right inset shows the relationships between the four independent subnetworks of Fig. 2 with the lengths of the branches connecting them. An asterisk identifies the 'rooting' nesting in each case.

analysis of the independent ITS2 nuclear marker, and as it is built on a more comprehensive sampling, we accept the new inference as a more robust explanation for the observed pattern.

Discussion

ITS2 phylogeography of the *Timarcha goettingensis* species complex

The use of nuclear ITS2 revealed that variation in populations of the *T. goettingensis* complex is highly structured. This confirmed the conclusions from NGCA of mitochondrial COII (Gómez-Zurita *et al.*, 2000a). The strong phylogeographic structure of these insect populations is consistent with their well-studied natural history, revealing very reduced dispersal capabilities (Petitpierre, 1970, 1973). *Timarcha* beetles are apterous and show habitat and host fidelity (González-Megías & Gómez, 2001), which limits dispersal. In addition, the orography of the Iberian Peninsula imposes geographical barriers, favouring speciation (Gómez-Zurita *et al.*, 2000a,b).

Despite the remarkable geographical structure, the NGCA of ITS2 sequences suggests that population separation is not complete, at the level of the geographical distance between sampling locations. However, gene flow estimates based on the two markers show F_{ST} values above 0.35. Applying an island model of population structure, these F_{ST} values are agreed to represent low genetic exchange between populations (Avise, 2000, p. 75).

With regard to NGCA inferences, lower level nestings, particularly 1-step clades, offer very little resolution, and the reliability of the inferences is compromised by the small sample sizes (two to four genotypes for two to 12 specimens). In the second and third levels of nesting, where sample sizes range from two to 23 individuals for two-step clades, and five to 33 individuals for three-step clades, the method still detects deficiencies in the sampling design, but a pattern of spatial structure related to restricted gene flow with isolation by distance starts to emerge.

A past fragmentation event is inferred for the group W3-2, detaching the southern-most samples (W1, W2, W4 and W6) from the populations situated at higher latitudes, and two contiguous range expansion events are inferred for the groups of haplotypes W3-5 and E2-1. The W3-5 case involves the colonization of a vast area in the north of the Iberian Peninsula and across the western Pyrenees from northwestern Iberian Peninsula, while the E2-1 case implies an expansion of populations from the Northeast of the Iberian Peninsula to Corsica. This latter inference should be considered only tentatively given the under-representation of other non-Iberian European samples in this study and it might change by including samples from other geographically close areas (e.g. Italy, Southern France). However, the markers divergences between the Corsican and its sister continental samples are compatible with a recent dispersal across water. The K2P average genetic distances are 0.007 ± 0.003 for ITS2 and 0.031 ± 0.003 for COII. Assuming a molecular clock and a vicariance caused by the split of the Sardo-Corsican plate initiated about 29 Mya (Oosterbroek & Arntzen, 1992) these divergences would relate to very slow evolutionary rates (0.00024 and 0.00106 changes/site/My for ITS2 and COII, respectively), at least an order of magnitude smaller than what would be compatible with the geological date of this separation.

The two major clades obtained in the analysis, the so-called western (W6-1) and eastern (E4-1) clades, and the major nestings within, reflect predominantly geographical structure because of restricted gene flow with isolation by distance (Table 2), and relatively low migration rates ($0.13 < Nm < 0.45$). Considering the entire network, with the four identified sub-networks and rooting the analysis in Network I, it is inferred a long distance colonisation event giving rise to population W9 well outside the area of distribution of the other

Table 2 Phylogeographic inferences of ITS2 data based on nested geographical clade analysis. Sample sizes (*n*) on which the inference is based are also indicated.

Clade	*n*	Inference chain	Inferred process
W1-2	4	1-2-11-17-NO	Inconclusive outcome
W1-3	2	1-2-11-17-NO	Inconclusive outcome
W1-4	2	1-2-11-17-4-NO	Restricted gene flow (RGF) with isolation by distance (IBD)
W1-6	3	1-2-11-17-NO	Inconclusive outcome
W1-7	12	1-2-11-17-4-9-10-NO	Geographical sampling scheme inadequate to discriminate between fragmentation and IBD
W1-8	7	1-2-3-4-9-NO	Past fragmentation
E1-1	6	1-2-11-12-13-14-NO	Sampling design inadequate to discriminate between contiguous range expansion (CRE) and long distance colonization (LDC)
W2-1	4	1-2-(no tip/interior)	Inconclusive outcome
W2-4	7	1-2-11-17-4-NO	RGF with IBD
W2-6	2	1-2-11-17-NO	Inconclusive outcome
W2-7	16	1-2-3-4-NO	RGF with IBD
W2-8	23	1-2-3-5-15-16-18-NO	Geographical sampling scheme inadequate to discriminate between fragmentation, range expansion, and IBD
E2-1	11	1-2-11-12-NO	CRE
E2-2	10	1-2-3-5-6-(clades ≤ 2)-7-8-NO	Insufficient genetic resolution to discriminate between range expansion/colonization and restricted dispersal/gene flow (impossibility to discriminate between short and long distance movements)
W3-2	8	1-2-11-17-4-9-NO	Past fragmentation
W3-3	5	1-2-(no tip/interior)	Inconclusive outcome
W3-4	26	1-2-(no tip/interior)	Inconclusive outcome
W3-5	33	1-2-11-12-NO	CRE
W3-6	17	1-2-3-4-9-10-NO	Geographical sampling scheme inadequate to discriminate between fragmentation and IBD
E3-1	23	1-2-11-12-13-14-NO	Sampling design inadequate to discriminate between CRE and LDC
E3-2	18	1-2-3-4-NO	RGF with IBD
W4-1	21	1-2-3-4-NO	RGF with IBD
W4-2	20	1-2-11-17-NO	Inconclusive outcome
W4-3	59	1-2-3-4-NO	RGF with IBD
W4-4	19	1-2-3-4-9-10-NO	Geographical sampling scheme inadequate to discriminate between fragmentation and IBD
E4-1	41	1-2-3-4-NO	RGF with IBD
W5-1	41	1-2-11-12-13-14-NO	Sampling design inadequate to discriminate between CRE and LDC
W5-2	78	1-2-3-4-NO	RGF with IBD
W6-1	119	1-2-3-4-NO	RGF with IBD
Total	167	1-2-11-12-13-YES	LDC (CRE, excluding Network IV)

haplotypes. This particular event, characterized by a significantly small Dc (restricted geographical distribution of haplotype 17) and a significantly large Dn (haplotype 17 geographically located at a great distance from the other haplotypes in the total network), might overpower other inferences from the remaining samples. This possibility was further explored by repeating the analysis excluding haplotype 17. In this case, the values obtained were: Network I ($Dc = 213.92^{S}$; $Dn = 232.79^{S}$), Network II ($Dc = 489.28^{L}$; $Dn = 536.07^{L}$) and Network III ($Dc = 0$; $Dn = 521.54$); [I-T] $Dc = -263.72^{S}$ and [I-T] $Dn = -302.93^{S}$. These results correspond to a situation of contiguous range expansion of populations, accounting for the structured division of groups W6-1 and E4-1, the western and eastern clades of this study. The direction of colonization would have occurred from central and north Iberia to the Northeast of the Peninsula, across the Pyrenees and into the rest of Europe. The described scenario is compatible with the population range expansion in Europe following ice retreat after the glacial periods (Hewitt, 1993, 1999; Taberlet *et al.*, 1998). However, as discussed above, several *T. goettingensis* samples from other European populations should be included to test this hypothesis convincingly. Limited sample density north of the Pyrenees could obscure an alternative pattern resulting from vicariance, more congruent perhaps with the low vagility of these insects. Figure 4 shows a summary of the inferred phylogeographic processes using NGCA and the ITS2 genealogy superimposing the topology of this network on the geography.

Table 3 Estimates of F_{ST} and migration rates (Nm) for the internal transcribed spacer region 2 (ITS2) clades identified with nested geographical clade analysis suggesting the recurrent process of restricted gene flow with isolation by distance.

		ITS2		COII	
Clade	*N*	F_{ST}*	Nm*	F_{ST}*	Nm*
E2-2	10	0.84	0.05	0.37	0.85
E3-2	18	0.39	0.39	0.37	0.85
W4-1	21	0.36	0.45	0.62	0.30
W4-3	59	0.39	0.39	0.37	0.85
E4-1	41	0.46	0.29	0.44	0.63
W5-2	78	0.66	0.13	0.25	1.53
W6-1	119	0.47	0.28	0.14	3.00

*Gene flow estimators are calculated following Hudson *et al.* (1992). In each case, the sub-populations are defined as the lower level nesting clades within the clade of interest. Identical estimates based on the same specimens and sub-populations defined on the basis of the ITS2 genealogy using the cytochrome oxidase II (COII) marker variation are also shown for comparison.

ITS2 and COII evolutionary processes are discordant

Phylogenetic signals of the ITS2 and COII data sets show a highly significant level of incongruence in ILD tests (Farris *et al.*, 1994) ($P < 0.001$). But how is this incongruence reflected in the independent assessment of phylogeographic processes estimated from each marker? NGCA of the COII data revealed highly structured mitochondrial populations with episodic events of 'contiguous range expansions' whereas the ITS2 phylogeography suggested that spatial distribution is explained predominantly by 'restricted gene flow with isolation by distance'. The COII pattern is hence explained as the result of a series of historical events, i.e. unique, rare occurrences affecting populations of *T. goettingensis* in the past. In contrast, the ITS2 patterns appear to be determined by an on-going, recurrent process of gene flow. These differences in the inferences from two markers are intriguing, and could have several explanations. First, they could simply be an analytical artefact related to different degrees of resolution offered by each marker. The degree of divergence between genotypes in COII and ITS2 differs, and it is conceivable that the NGCA method results in different inferences given the amount of variation available.

Alternatively, the results could reflect true differences in the history of genetic markers within organisms. While difficult to reconcile within a single linkage group, it is conceivable that the great differences in the modes of inheritance in maternally inherited mtDNA, and biparentally inherited nuclear markers which undergo recombination and (in the case of tandemly repeated rDNA loci) various processes of concerted evolution leading to homogenization, could indeed lead to substantial differences in the phylogenetic history of loci. Also, since the mtDNA markers exclusively retain the history of the female genealogies, conflicting phylogeographic inferences, as those reported here, would be expected if males and females differed in their natural histories, e.g. phylopatric females and dispersing males (Davies & Bermingham, 2002).

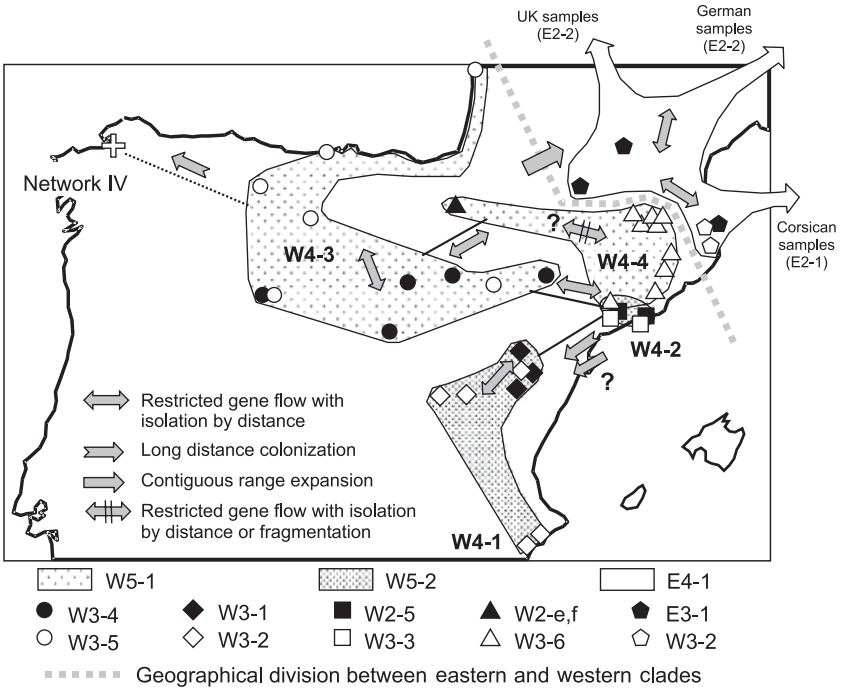

Fig. 4 Overlay on the geography of the topological structure of the internal transcribed spacer region 2 statistical parsimony networks of the *Timarcha goettingensis* species complex based on the nesting design. The names of the groups of samples defined by the nesting design are the same shown in Figs 2 and 3. Different types of arrows signify different types of inferences for each deduced case of geographically structured populations based on nested geographical clade analysis; a question mark indicates ambiguity in the inference. The genetic breakage between western and eastern clades in the complex is identified by a thick dotted line.

Finally, the two sets of markers might expose the same underlying evolutionary history, but differences in evolutionary rates and coalescent dynamics leave a different genetic signature in either type of genetic marker. The substitution rates in ITS2 were lower than in COII at the level of the entire species complex (K2P corrected mean pairwise divergences – ITS2: 0.017 ± 0.007; COII: 0.042 ± 0.015), and in a separate comparison of the eastern and western clades ($ITS2_E$: 0.007 ± 0.003 and $COII_E$: 0.026 ± 0.015; $ITS2_W$: 0.015 ± 0.006 and $COII_W$: 0.033 ± 0.013). In addition, the effective allele population size is four times larger for a Mendelian nuclear marker compared with the haploid mitochondrial genome, and hence their expected coalescent times are greater (Kingman, 1982). Time to coalescence in ITS2 may be further prolonged because of the tandemly repeated, multi-copy arrangement of the rDNA locus and its complex homogenization dynamics (Ohta & Dover, 1984). The shorter times to fixation of the mtDNA variants therefore might lead to inferences of high spatial structuring with no evidence of gene flow, whereas ITS2 retain evidence for past recurrent gene flow between populations in the complex. This situation is more likely to emerge when migration rates are sufficiently small, as deduced for the *T. goettingensis* complex, where fast evolving markers will have the chance to coalesce before any migration event occurs (Hudson *et al.*, 1992).

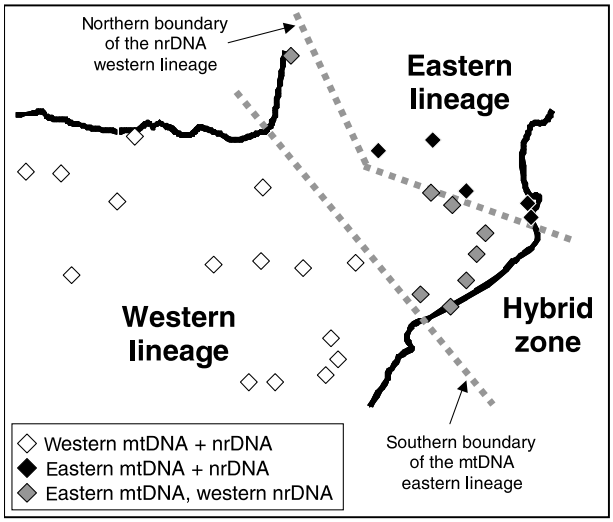

Fig. 5 Graphical representation of the inference of a hybrid zone in the area of parapatry of the western and eastern lineages of the *Timarcha goettingensis* species complex. Thick dotted lines indicate the inferred geographical boundaries between these lineages with nested geographical clade analysis of internal transcribed spacer region 2 and COII sequences. The region delimited by these shifting boundaries is tentatively regarded as a hybrid zone, containing individuals with mosaic genomes.

Geographical discordance of major ITS2 and COII genotypes

Whatever the precise inference about the processes leading to the spatial patterns, for both markers we observed that geographically close populations bear evolutionarily close haplotypes. However, this is not the case for populations in or near contact zones where eastern and western clades produce different genetic marker distributions, with the position of the boundary of the western clade in the ITS2 further north relative to the break in COII (Fig. 5). It is possible that this marker distribution is due to the secondary contact between formerly subdivided eastern and western lineages, as this region is a well established post-glacial hybrid zone for many organisms, including several invertebrate species (Hewitt, 1993). In *T. goettingensis* there is independent evidence for this scenario from a clinal pattern of morphological variation across this putative area of secondary contact of the western and eastern lineages (Petitpierre, 1970). The identification of this potential hybrid zone now can be further corroborated by identifying hybrids responsible for the increase of homoplasy in the comparison of data sets and a major source of incongruence. However, more specific phylogeographic tests are needed to evaluate and characterize hybridization events from sequences of multiple loci showing incongruence (J. Gómez-Zurita & A. P. Vogler, in preparation).

Conclusions

In this study we have shown that the NGCA methodology is quite sensitive to sampling design, as the previously published COII phylogeography of the *T. goettingensis* complex (Gómez-Zurita *et al.*, 2000a) and the reanalysis here show some disparities. More importantly, the analysis of an identical set of populations but using mitochondrial and nuclear markers with distinct evolutionary characteristics also has shown substantial discrepancies in the inferred processes of evolution, suggesting that individual loci provide only a partial picture of the phylogeography of the group. These results would generally caution against the inferences from phylogeographic analyses based on single genetic markers or groups of linked loci.

Alternatively, it could be argued that NGCA methodology *per se* is an unreliable procedure for inferring processes of phylogeographic differentiation. As has been claimed recently, this method might be inaccurate in the reconstruction of historical inferences because of lack of explicit reference to stochastically derived expectations and it does not provide confidence limits for the inferred phylogeographic processes (Knowles & Maddison, 2002). A more general criticism is that inferences of organismal history from patterns of relatedness and frequencies of genotypes within the species boundary are problematic (Brower *et al.*, 1996).

Nevertheless, NGCA is currently the only method for explicit analysis of phylogeographic patterns and as such has been very helpful in this study. Specifically, the NGCA procedure remains useful for direct comparisons between different gene markers obtained from the same set of populations. In this situation, it provides an explicit analytical procedure to reveal any differences in gene history and geographical distribution of genetic markers, and make them amenable to further interpretation. Here, this permitted the analysis of marker distributions with regard to differences in coalescent history (gene tree–species tree incongruences) and also revealed evidence for species hybridization. Hence the combination of several markers can result in a more complete and detailed picture of the evolutionary processes affecting a species. The conclusion is therefore straightforward: if our interests are the phylogeographies of organisms rather than genes, phylogeographic analyses should be based on diverse markers, otherwise the inferences may be specific to the history of a particular gene only.

Acknowledgments

The authors wish to thank Andrés Baselga, Mauro Daccordi, Koniev Desender, Javier de Diego, Frank Fritzlar, Fernando Murria, Eduard Petitpierre, Ross Piper, Joan Pons, Antoni Sacarés, Matthias Schöller, Eduard Vives and Joan Vives, for their help in collecting the samples. Eduard Petitpierre and Carlos Juan are acknowledged for their encouragement and ideas for this work. We are also indebted to Prof. Peter van Tienderen and an anonymous referee for their useful comments to improve this manuscript. This work has been supported by the European Community through Access to Research Infrastructures Programme (HPRI-CT-1999-00062) and a Marie Curie Fellowship (HPMF-CT-2000-00744) to JGZ.

Supplementary material

The following material is available from: http://www.blackwellpublishing.com/products/journals/suppmat/JEB/JEB599/JEB599sm.htm

Table S1 Results of nested geographical cladistic analysis for the 3 lower levels in the nesting series, including haplotypes, 1-step and 2-step groups of haplotypes. An asterisk identifies the ''interior'' haplotype(s) according to the network topology, and significance of Dc and Dn estimates is represented by a superscript ''S'' for values significantly smaller and superscript ''L'' for values significantly larger than expected under panmixia.

Table S2 Results of nested geographical cladistic analysis for the 4 higher levels in the nesting series, including 3-, 4-, 5- and 6-steps groups of haplotypes. An asterisk identifies the ''interior'' haplotype groups according to the network topology, and significance of Dc and Dn estimates is represented by a superscript ''S'' for values significantly smaller and superscript ''L'' for values significantly larger than expected under panmixia.

References

Avise, J.C. 2000. *Phylogeography. The History and Formation of Species*. Harvard University Press, Cambridge, MA, USA.

Brower, A.V.Z., DeSalle, R. & Vogler, A.P. 1996. Gene trees, species trees, and systematics: a cladistic perspective. *Ann. Rev. Ecol. Syst.* **27**: 423–450.

Clement, M., Posada, D. & Crandall, K.A. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* **9**: 1657–1660.

Davies, N. & Bermingham, E. 2002. The historical biogeography of two Caribbean butterflies (Lepidoptera: Heliconiidae) as inferred from genetic variation at multiple loci. *Evolution* **56**: 573–589.

Doyle, J.J. 1997. Trees within trees: genes and species, molecules and morphology. *Syst. Biol.* **46**: 537–553.

Farris, J.S., Källersjö, M., Kluge, A.G. & Bult, C. 1994. Testing significance of incongruence. *Cladistics* **10**: 315–319.

Gómez-Zurita, J. 2002. Quand la phylogéographie rencontre l'histoire des gènes: l'analyse cladistique emboîtée (nested clade analysis) et l'inférence de processus populationnels récurrents et épisodiques. *Biosystema* **20**: 95–114.

Gómez-Zurita, J., Petitpierre, E. & Juan, C. 2000a. Nested cladistic analysis, phylogeography and speciation in the *Timarcha goettingensis* complex (Coleoptera, Chrysomelidae). *Mol. Ecol.* **9**: 557–570.

Gómez-Zurita, J., Juan, C. & Petitpierre, E. 2000b. The evolutionary history of the genus *Timarcha* (Coleoptera, Chrysomelidae) inferred from mitochondrial COII gene and partial 16S rDNA sequences. *Mol. Phylogenet. Evol.* **14**: 304–317.

Gómez-Zurita, J., Juan, C. & Petitpierre, E. 2000c. Sequence, secondary structure and phylogenetic analyses of the ribosomal internal transcribed spacer 2 (ITS2) in the *Timarcha* leaf beetles (Coleoptera: Chrysomelidae). *Insect Mol. Biol.* **9**: 591–604.

González-Megías, A. & Gómez, J.M. 2001. Adult and larval plant range and preference in *Timarcha lugens* (Coleoptera: Chrysomelidae): strict monophagy on an atypical host. *Ann. Entomol. Soc. Am.* **94**: 110–115.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.C. & Zegura, S.L. 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**: 427–441.

Hewitt, G.M. 1993. Postglacial distribution and species substructure: lessons from pollen, insects and hybrid zones. In: *Evolutionary Patterns and Processes* (D.R. Lees & D. Edwards eds), pp. 97–123. Academic Press, London.

Hewitt, G.M. 1999. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* **68**: 87–112.

Hudson, R.R., Slatkin, M. & Maddison, W.P. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R.C., Templeton, A.R. & Hammer, M.F. 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**: 817–831.

Kingman, J.F.C. 1982. The coalescent. *Stoch. Proc. Appl.* **13**: 235–248.

Knowles, L.L. & Maddison, W.P. 2002. Statistical phylogeography. *Mol. Ecol.* **11**: 2623–2635.

Maddison, W.P. 1997. Gene trees in species trees. *Syst. Biol.* **46**: 523–536.

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

Ohta, T. & Dover, G.A. 1984. The cohesive population genetics of molecular drive. *Genetics* **108**: 501–521.

Oosterbroek, P. & Arntzen, J.W. 1992. Area-cladograms of circum-mediterranean taxa in relation to Mediterranean paleogeography. *J. Biogeogr.* **19**: 3–20.

Pérez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martínez-Arias, R., Clarimón, J., Fiori, G., Luiselli, D., Facchini, F., Pettener, D. & Bertranpetit, J. 1999. Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* **65**: 208–219.

Petitpierre, E. 1970. Variaciones morfológicas y de la genitalia en las *Timarcha* Lat. (Col. Chrysomelidae). *P. Inst. Biol. Apl.* **48**: 5–16.

Petitpierre, E. 1973. *Estudios sistemáticos, citogenéticos y evolutivos sobre el género Timarcha (Coleoptera, Chrysomelidae)*, PhD Thesis. Publ. Univ. Barcelona.

Posada, D., Crandall, K.A. & Templeton, A.R. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* **9**: 487–488.

Rozas, J. & Rozas, R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.

Seielstad, M., Minch, E. & Cavalli-Sforza, L.L. 1998. Genetic evidence for higher female migration rate in humans. *Nat. Genet.* **20**: 278–280.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G. & Cosson, J.F. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* **7**: 453–464.

Templeton, A.R. & Sing, C.F. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* **134**: 659–669.

Templeton, A.R., Boerwinkle, E. & Sing, C.F. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics* **117**: 343–351.

Templeton, A.R., Crandall, K.A. & Sing, C.F. 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, A.R., Routman, E. & Phillips, C.A. 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.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.

White, T.J., Bruns, T., Lee, S. & Taylor, J.W. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: a Guide to Methods and Applications* (M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White eds), pp. 315–322. Academic Press, New York.

Received 15 March 2003; accepted 16 June 2003