

PHYLOGEOGRAPHIC HISTORY OF THE LAND SNAIL *CANDIDULA UNIFASCIATA* (HELICELLINAE, STYLOMMATOPHORA): FRAGMENTATION, CORRIDOR MIGRATION, AND SECONDARY CONTACT

MARKUS PFENNINGER¹ AND DAVID POSADA^{2,3}

¹Abteilung Ökologie und Evolution, Zoologisches Institut der J.W.Goethe-Universität, Bio-Campus Siesmayerstraße, D-60054 Frankfurt/Main, Germany

E-mail: Pfenninger@zoology.uni-frankfurt.de

²Department of Zoology, Brigham Young University, Provo, Utah 84602

Abstract.—We studied sequence variation in 16S rDNA in 204 individuals from 37 populations of the land snail *Candidula unifasciata* (Poiret 1801) across the core species range in France, Switzerland, and Germany. Phylogeographic, nested clade, and coalescence analyses were used to elucidate the species evolutionary history. The study revealed the presence of two major evolutionary lineages that evolved in separate refuges in southeast France as result of previous fragmentation during the Pleistocene. Applying a recent extension of the nested clade analysis (Templeton 2001), we inferred that range expansions along river valleys in independent corridors to the north led eventually to a secondary contact zone of the major clades around the Geneva Basin. There is evidence supporting the idea that the formation of the secondary contact zone and the colonization of Germany might be postglacial events. The phylogeographic history inferred for *C. unifasciata* differs from general biogeographic patterns of postglacial colonization previously identified for other taxa, and it might represent a common model for species with restricted dispersal.

Key words.—Helicellinae, nested clade analysis, phylogeography, secondary contact.

Received July 2, 2001. Accepted June 11, 2002.

The formation of species or major phylogenetic lineages within species in Europe has been deeply influenced by a history of repeated glaciations during the Pleistocene period. Phylogeographic studies on organisms with efficient dispersal capacities have led to the recognition of a common set of processes resulting from the fragmentation into glacial refuges, range expansions via postglacial colonization routes, and suture or secondary contact zones among formerly disjointed lineages (Taberlet et al. 1998; Hewitt 2000).

All the same, similar studies on animal species with restricted vagility are still rare. As a matter of fact, the study of species with limited dispersal may reveal different historical patterns of refuges and migration routes than those of more mobile species, and therefore, it may lead to a more complete understanding of the relationship between evolutionary history and contemporary distribution of genetic variation (Cruzan and Templeton 2000). Indeed, the identification of cryptic northern refugia in plants and mammals has recently drawn additional attention to this issue (Stewart and Lister 2001).

Land snails are ideal organisms to study phylogeographic patterns and evolutionary processes in species with limited dispersal. Although a few land snails have attained a worldwide distribution via anthropogenic displacement (Godan 1979), the range of most land snails is usually quite restricted due to their poor dispersal ability and particular habitat requirements (Cowie 1984; Baur 1993; Pfenninger et al. 1996). These characteristics prevent land snails from actively escaping changing ecological conditions, and it is commonplace to assume that Pleistocene glaciations led to the extinction of most gastropod species and were followed by postglacial col-

onization from suitable refuges (Ant 1966). Indeed, this low dispersal capacity tends to preserve patterns of genetic variation that arose in the past, as these patterns are not blurred by contemporary gene flow, and it renders a species particularly amenable to phylogeographic study (Cruzan and Templeton 2000). In addition, the shells of land snails are well preserved in a subfossil state in aeolian sediments like loess (Goodfriend 1992), and this abundant fossil record provides useful knowledge about past geographical distributions. It is surprising that, despite the suitability of land snails for phylogeographic inference, only a few studies of this kind have been carried out (Ross 1999; Guiller et al. 2001).

Here we tried to decipher the phylogeographic history of the land snail *Candidula unifasciata* (Poiret 1801) (subfamily Helicellinae, family Helicidae). These are small snails (3.8 ± 0.6 mm height, 6.3 ± 0.8 mm breadth) that display substantial variation in shell morphology (Gittenberger 1993; Pfenninger and Magnin 2001), and with a life span that usually does not exceed one year. The average active dispersal during this period is estimated from mark-recapture experiments to be roughly 5 m (Bahl et al. 1996). The core distribution of this species comprises southeast France and Germany, with scattered populations reported from north and west France, Hungary, Poland, and Scandinavia (Kerney et al. 1983; M.Pfenninger, pers. obs.). Populations of this species are relatively abundant in the southern part of the species range, whereas density decreases towards the north. This pattern is explained by habitat preference, as populations of *C. unifasciata* often occur in dry grass- and scrublands with sparse vegetation cover, preferentially on calcareous soils or rocks (Magnin 1993; Bahl et al. 1996), which is less frequent in higher latitudes. Regarding the fossil record, the widespread occurrence of *C. unifasciata* in Pleistocenic loess deposits in southeastern France is associated with interstadial climate stages, where periods of warming climates allowed a rapid

³ Present address: Variagenics, Inc., Cambridge, Massachusetts 02139 and Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

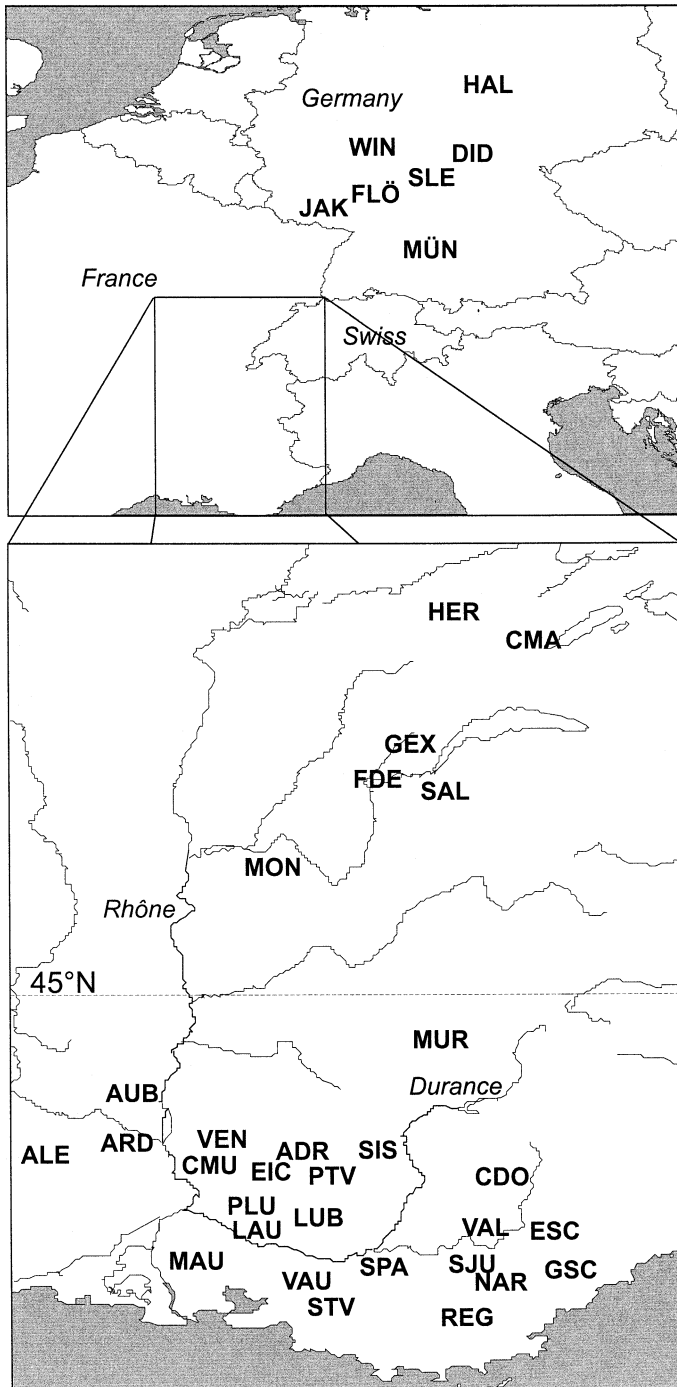


FIG. 1. Geographical location of *Candidula unifasciata* sampling sites. The lower map is a magnification of the framed part in the upper map; abbreviations correspond to those given in Table 1.

colonization of preboreal open lowland habitats (Magnin 1989, 1993). In the northern part of the present day distribution, *C. unifasciata* has only been reported from fossil assemblages attributed to Holocene deposits (Puissegur 1978).

Phylogeography is the study of the history of gene flow in space and time (Avice 2000). Indeed, knowledge of gene-flow pattern is essential to understand the evolutionary processes acting when lineages are divided in space and time

(Cruzan and Templeton 2000) or when formerly divided lineages meet again (Hewitt 1996). Compelling advances in coalescent theory over the last twenty years have provided a sound statistical framework to test complex hypotheses about historical and present gene flow (Donnelly and Tavaré 1995; Fu and Li 1999; Templeton 1998; Edwards and Beerli 2000). Here we conducted phylogenetic, coalescent, and nested clade analyses upon 16S sequence variation in *C. unifasciata* in order to study the impact of Pleistocene glaciations on the temporal dynamics of species distribution and population structure of a poor disperser. We focus specifically on the identification of potential Pleistocene refuges, migration routes, and potential secondary contact zones.

MATERIALS AND METHODS

Sampling Collection and DNA Extraction

Thirty-seven sites, comprising the core species range of *C. unifasciata* in Germany, Switzerland, and France, were sampled (Fig. 1 and Table 1). Sampling effort represents roughly population density, as recommended by Cruzan and Templeton (2000). The taxonomic status of sampled individuals, especially their distinctness from the congeneric species *C. rugosiuscula* (Michaud 1831) and *C. gigaxii* (L. Pfeiffer 1850), was assessed through the application of a morphometric and genetic marker-based recognition approach (Pfenninger and Magnin 2001). A total of 204 individuals were crushed with their shells in 10% w/v laundry detergent solution for storage at room temperature and tissue digestion, following the protocol of Bahl and Pfenninger (1996). All samples were shaken for 24 h at 37°C in the laboratory prior to phenol/chloroform extraction of total DNA following a standard protocol (Sambrook et al. 1989).

Amplification of 16S rDNA, Sequencing and Alignment

The 16S target-DNA was amplified by PCR with standard universal primers for the sequence 16S1 (5' > CGC AGT ACT CTG ACT GTG C < 3' and 16S2 5' > GTC CGG TTT GAA CTC AGA TC < 3'). Amplification was performed with Boehringer Taq-polymerase in 12.5 ml total reaction volume with standard reaction conditions. Samples were amplified for 10 cycles (92°C for 50sec, 44°C for 50sec, and 72°C for 40sec) and 36 cycles (92°C for 30sec, 48°C for 40sec and 72°C for 40sec) after initial incubation of 90°C for 2 min, 30sec. Both strands of the purified amplification products were cycle-sequenced with the Perkin Elmer Taq DyeDeoxy terminator cycle sequencing kit after the protocol of the supplier and read automatically on the ABI Prism 377 sequencing device of the same manufacturer. Sequences (GenBank accession nos. AF407841–AF408058) were aligned with the Clustal option (Thompson et al. 1994) in the program Sequence Navigator (Perkin Elmer, Applied Biosystems Foster City, CA), and adjusted by eye. We attempted to sequence other genes as well, but available universal primer for fast evolving protein genes as *COI* or *Cytb* produced unspecific results.

Phylogenetic Estimation

The phylogeny of the 16S rDNA haplotypes was inferred using maximum-likelihood (ML), maximum-parsimony

TABLE 1. Sampling sites from south to north, region (Département, Kanton, Bundesland, respectively), abbreviations used, abbreviation of pooled population for coalescent analyses, geographical position and number of sampled individuals (n). The abbreviations for the pooled populations are: SD, South of river Durance; RR, Right of river Rhône; LR, Left of Rhône; B, Burgundy, and D, Germany.

Population	Region	Abbreviation	Pooled population	Position	n
Sainte Victoire	Bouches du Rhône	STV	SD	43°32'18"N, 05°34'66"E	3
Vauvenargue	Bouches du Rhône	VAU	SD	43°33'14"N, 05°35'06"E	5
Regasse	Var	REG	—	43°39'30"N, 06°08'00"E	2
Saint Paul	Var	SPA	SD	43°40'30"N, 05°41'15"E	5
Grotte Saint Cezaire	Alpes Maritimes	GSC	—	43°40'82"N, 06°48'57"E	9
Maussane	Bouches du Rhône	MAU	SD	43°41'71"N, 04°50'55"E	4
Escragnolles	Alpes Maritimes	ESC	—	43°43'78"N, 06°47'58"E	11
Lauris	Vaucluse	LAU	RL	43°45'35"N, 05°15'31"E	5
Grand Luberon	Vaucluse	LUB	RL	43°47'55"N, 05°26'10"E	6
Petit Luberon	Vaucluse	PLU	RL	43°48'18"N, 05°17'02"E	5
Valensole	Alpes de H ^{te} Provence	VAL	—	43°49'40"N, 05°58'50"E	4
Naverre	Alpes de H ^{te} Provence	NAV	—	43°52'55"N, 06°13'50"E	5
Saint Jurs	Alpes de H ^{te} Provence	SJU	—	43°53'55"N, 06°12'40"E	8
Col d'Orme	Alpes de H ^{te} Provence	CDO	—	43°54'40"N, 06°12'36"E	2
Col de Murs	Vaucluse	CMU	RL	43°58'50"N, 05°13'45"E	5
Plateau de Vaucluse	Vaucluse	PTV	RL	44°00'10"N, 05°32'05"E	5
Les Eicharettes	Vaucluse	EIC	RL	44°04'50"N, 05°26'45"E	6
Les Adriens	Vaucluse	ADR	RL	44°06'26"N, 05°30'43"E	5
Le Ventouret	Vaucluse	VEN	RL	44°07'45"N, 05°22'20"E	5
Alès	Gard	ALE	RR	44°07'45"N, 04°12'40"E	4
Sisteron	Drôme	SIS	RL	44°12'15"N, 05°54'39"E	4
Ardèche	Ardèche	ARD	RR	44°18'30"N, 04°33'22"E	5
Aubenas	Ardèche	AUB	RR	44°35'30"N, 04°25'10"E	1
La Mure	Isère	MUR	—	44°52'55"N, 05°50'62"E	15
Montalieu	Isère	MON	B	45°08'30"N, 05°23'25"E	5
Mons Salève	Haute Savoie	SAL	B	46°04'40"N, 06°07'30"E	6
Fort d'Ecluse	Ain	FDE	B	46°07'25"N, 05°53'15"E	5
Gex	Ain	GEX	B	46°21'05"N, 06°03'00"E	5
Col de Marchaiz	Fribourg	CMA	B	46°32'62"N, 06°15'15"E	6
Herticourt	Doubs	HER	B	47°15'30"N, 06°45'10"E	5
Münsingen	Baden-Württemberg	MÜN	D	48°25'00"N, 09°30'00"E	10
Jakobsberg	Rheinland-Pfalz	JAK	D	49°75'71"N, 07°59'08"E	3
Flörsheim	Hessen	FLÖ	D	50°00'01"N, 08°23'10"E	7
Schlüchtern/Elm	Hessen	SLE	D	50°21'40"N, 09°33'20"E	8
Winterscheid	Hessen	WIN	D	50°55'85"N, 09°01'91"E	4
Dierdorf	Thüringen	DID	D	51°10'66"N, 10°17'34"E	5
Halle	Sachsen-Anhalt	HAL	D	51°42'35"N, 11°53'15"E	5

(MP), and statistical-parsimony (SP) analyses (Templeton et al. 1992). For the MP and SP analyses, insertions and deletions were treated as a fifth state. The best-fit model of nucleotide substitution for the ML analyses was selected using the hierarchy of likelihood-ratio tests implemented in Modeltest 3.0 (Posada and Crandall 1998). Maximum-parsimony and ML heuristic searches were conducted with 1000 random sequence addition replicates. Nodal support was estimated using the bootstrap approach (Felsenstein 1985) with 1000 replicates. The MP and ML analyses were performed with PAUP 4.03b (Swofford 1998). The SP cladogram was constructed using TCS vers. 1.06 (Clement et al. 2000).

Nested Clade Analyses

We carried out a nested clade analysis (NCA; Templeton et al. 1995; Templeton 1998) to infer the population history of *C. unifasciata*. The NCA is an objective statistical approach that first tries to reject the null hypothesis of no association between haplotype variation and geography, and then interprets the significant patterns using explicit criteria that include an assessment of sampling adequacy. This analysis uses the temporal information contained in the haplotype

tree to partition historical (fragmentation, range expansion) from current (gene flow, drift, system of mating) processes responsible for the observed pattern of genetic variation. In addition, the NCA allows for the inference of separate evolutionary events in space and time. The NCA nesting design was constructed by hand on the SP cladogram following the rules given in Templeton (1998) and Crandall (1996). The program GeoDis 2.0 (Posada et al. 2000) was used to calculate the various NCA distance measures and their statistical significance:

(1) Average clade distance (D_C) measures the average distance of all clade members from the geographical center of the clade distribution relative to other clades within the same nesting clade; (2) Nested clade distance (D_N) measures how widespread a particular clade is relative to the distribution of its nesting clade; (3) Interior-tip distances ($I-T_C$ and $I-T_N$) indicate how widespread younger clades (tips) are compared to their ancestors (interiors), relative to other clades within the same nesting clade; and (4) Average population clade distances (APCD) measures average clade distance from the geographical center for the involved haplotypes or clades found in a particular population.

The statistical significance of the distance measures was calculated by comparison with a null distribution derived from 10,000 random permutations of clades against sampling locality. Biological inferences for each clade with significant geographical association were drawn from the patterns of significant distance measures using the inference key given in Templeton (1998) and available at http://zoology.byu.edu/crandall_lab/geodis.htm.

To detect secondary contact between lineages for which the previous inference of fragmentation is a prerequisite, we used an extension of the NCA recently proposed by Templeton (2001). This strategy consists of calculating the average clade distance from the geographical center for the involved haplotypes or clades found in each population and each nesting level (APCD). In panmictic populations, all haplotypes and clades are expected to have the same geographical center, so that the average population clade distance should be the same for all populations. Under isolation by distance, the lower clade levels are expected to have small positive average population clade distances that approach zero with rising clade level. If haplotypes from previously fragmented clades are united in a single population, the average population clade distance is expected to remain high or even rise with clade level, until a maximum should be reached at the clade level where the fragmentation was inferred (Templeton 2001). The statistical significance of this distance measure was evaluated through 10,000 random permutations of clades against sampling locality.

Solving Cladogram Ambiguities

In some cases, and to construct the nesting, ambiguities in the cladogram (represented by loops) have to be solved (although see Templeton and Sing 1993). For that purpose, and at least under neutrality, we can apply several empirical predictions derived from coalescent theory (Crandall and Templeton 1993; Crandall et al. 1994; Posada and Crandall 2001). First, we expect that more frequent haplotypes are older. The probability that an allele that occurs n_i times in a sample of size n is the oldest is n_i/n (Watterson 1976; Kelly 1977; Watterson and Guess 1977; Donnelly and Tavaré 1996), and the expected rank of the alleles by age is equal to the rank of alleles by frequency (Donnelly and Tavaré 1996). If high frequency haplotypes have been present in the population for a long time, they had more chances of originating new haplotypes than did younger haplotypes. This implies not only that frequent haplotypes occupy often interior nodes in the cladogram whereas rare haplotypes are expected to occur at the tips of the cladogram, but also that frequent haplotypes have a greater number of mutational connections (Excoffier and Langaney 1989; Golding 1987). Second, we expect that newly arisen haplotypes are more likely to remain in the original population than to move to a distant population, unless high levels of gene flow occur (Watterson 1985; Takahata 1988). Therefore, singletons (haplotypes with a frequency of one) are more likely to be connected to haplotypes from the same population than to haplotypes from different populations.

These predictions from coalescent theory, validated in empirical data sets by Crandall and Templeton (1993), were summarized in three criteria that were used to decide among

alternative solutions of the loops: (1) Frequency criterion: haplotypes are more likely to be connected to haplotypes with higher frequency than to singletons; (2) Topological criterion: haplotypes are more likely to be connected to interior haplotypes than to tip haplotypes; and (3) Geographical criterion: haplotypes are more likely to be connected to haplotypes from the same population or region than to haplotypes occurring in distant populations.

Coalescence Analyses

Although the NCA may be very useful to infer different historical process, this approach does not allow for the estimation of population genetic parameters of interest like rates of gene flow. The coalescence approach (Kingman 1982) can be used to obtain maximum-likelihood estimates of gene flow among populations taking into account at the same time population structure and demographic processes. This class of estimators overcomes the limitations imposed by more conventional population genetic models, which usually rely on the biologically unrealistic assumptions of equal population sizes of subpopulations or symmetric gene flow (Bossart and Prowell 1998; Beerli and Felsenstein 2001). It is thus likely that these coalescent-based approaches of gene flow estimates deliver better estimates than F_{ST} based methods (Beerli and Felsenstein 1999).

Here we are interested in the relative likelihoods of different biologically plausible migration models within clade 4-2. We focus on this clade because it was the clade that expanded its range to the north. We defined a priori two biologically plausible gene-flow models, and compared these two with a third ad hoc model derived from the results of the NCA analysis:

(1) NCA Model, with unconstrained gene flow among SD, RL, RR, B, and expansion towards D. This posthoc model reflects the inferences obtained by NCA; (2) Island-like model, where only neighboring populations exchange migrants; and (3) Northern expansion model, from southern France to Germany.

Ideally one would simultaneously estimate all the relevant parameters for these migration models along with their likelihoods, but this is not computationally feasible. First, it does not seem possible to estimate these parameters for the 37 populations, as this implies many migration parameters to be estimated from low sample sizes. In such cases, data pooling can be a reasonable solution, so we pooled the 37 populations in five groups representing distinct geographical regions (Table 1). Second, even for five populations, it still does not seem possible to jointly estimate all the relevant migration parameters, thus we decided to split the estimation problem in several steps. First we estimated the migration matrices corresponding to the three models evaluated using Migrate 0.9.7 (Beerli and Felsenstein 1999). Second, we estimated the population mutation parameter and growth rate using Genetree 8.3 (Griffiths and Bahlo 1998), given the migration matrix estimated in the previous step. Finally, we estimated the likelihood of the competing models using Genetree 8.3 given the previously estimated SP cladogram, migration matrices, population mutation parameter, and growth rate.

For the Migrate analysis, the starting values of population

TABLE 2. Distribution of *C. unifasciata* haplotypes (columns) at each sampled location (rows). Abbreviations as in Table 1 and Figure 1.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
STV																				
VAU		2																		
MAU		3																		
SPA																				
REG																	1	1		
VAL																				
SJU															3					
NAV															4	1				
CDO																				
GSC								8												
ESC									1											
SIS	2	2																		
VEN		4		1																
EIC		3		3																
PTV		2		3																
ADR				3							1	1								
LUB		2		1																
CMU				1	1	1	1	1												
PLU				1																
LAU																				
ALE		4																		
ARD		5																		
AUB		1																		
MUR										8										3
MON		3												2						
GEX				5																
FDE		4							1											
SAL		2		2															1	1
HER		5																		
CMA		6																		
MÜN		8												2						
FLÖ		7																		
SLE		4																		
WIN		1	3																	
HAL		1	4																	
JAK		3																		
DID		1	4																	
Σ	2	73	11	20	1	1	1	1	9	9	1	1	2	2	7	1	1	1	4	1

mutation parameter θ and M , the ratio between immigration rate to the respective subpopulation and the mutation rate per generation, were estimated from F_{ST} values (Maynard Smith 1970, Nei and Feldman 1972, Beerli and Felsenstein 1999). We ran 10 short chains, each with a total of 50,000 genealogies and with a sampling increment of 20 genealogies, and five long chains, each with a total of 100,000 genealogies and a sampling increment of 20 genealogies. The first 10,000 genealogies in each chain were discarded. For the other settings, we used the default values. For the Genetree analysis first we jointly estimated the population mutation parameter θ and the growth rate β under a subdivided population model with migration and given the SP cladogram. To obtain these parameter estimates we iteratively completed 15 runs with 1,000,000 replications each. Given the θ and β estimates and the SP cladogram, the likelihoods of the three migration matrices estimated in Migrate were then calculated in a single run with 10,000,000 replications.

Divergence Times

The molecular clock hypothesis was tested with a likelihood ratio test (LRT) upon the estimated tree and best-fit model (Felsenstein 1981). To calibrate the molecular clock

we used an estimate of 0.056 changes per site and million years (Pfenninger and Magnin 2001). This relatively high molecular clock rate for 16S was estimated from the presumed age of the split of *C. unifasciata* and its sister species, *C. rugusiosculca*. Such rates might be typical for terrestrial snails (Thomaz et al. 1996). Divergence times were estimated according to Nei (1987). Patristic pairwise distances among haplotypes were inferred from the statistical parsimony network. The divergence time t between clades was computed as $t = (d_{xy} - 0.5 * (d_x + d_y)) \times$ substitution rate), where d_x and d_y denote the average sequence diversity within and d_{xy} the sequence divergence among clades, respectively (Nei 1987). An advantageous feature of this approach is that it accounts for divergence in the ancestral population and does not require reciprocal monophyly (Edwards and Beerli 2000).

RESULTS

16S Haplotype Diversity

The amplified mitochondrial 16S rDNA sequences from 204 individuals of *C. unifasciata* were between 321 and 326 base-pairs long. We observed two indels, both in loop coding regions of the 16S rDNA, according to a molluscan consensus

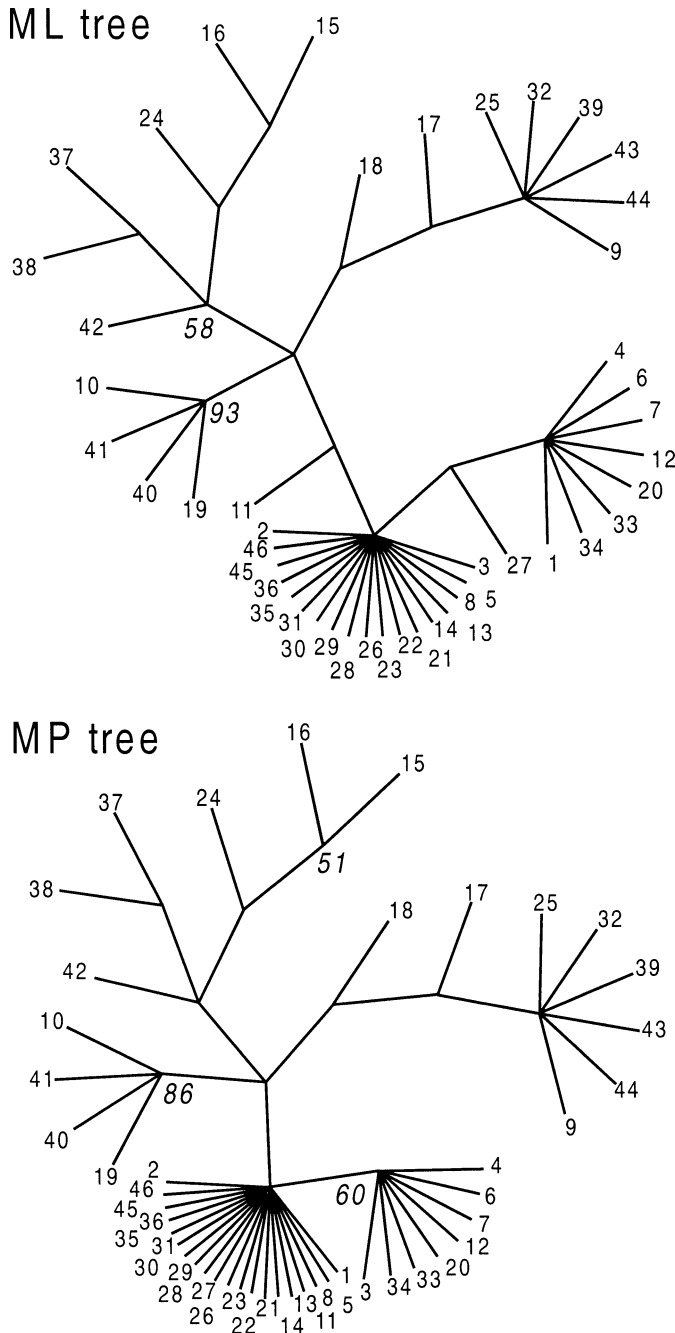


FIG. 2. Maximum-likelihood and maximum-parsimony estimates of haplotype phylogeny. Above: strict consensus of three maximum-likelihood trees under the HYK-model of sequence evolution. Below: strict consensus of 1062 equally parsimonious trees. Haplotypes refer to those listed in Table 2. Italicized numbers represent the percentage bootstrap values from 500 replications. Bootstrap values lower than 50% are not shown.

STV, than to haplotype 13, less frequent and found far away. Loop 2 can be solved in a very similar fashion to loop 1, whereas loop 3 is easily solved by appealing to the frequency and geographical criteria. However, the remaining loop could be broken at four different places (4A–4D, see Fig. 3) with similar probabilities according to the different criteria. The four resulting cladograms led to different nesting designs and,

thus, potentially different inferences about population history. We explored the inferences drawn from all possible solutions of loop 4, and despite the differences in the resulting nesting designs, the interpretation of the population history was essentially identical. We present here the nesting solution corresponding to the removal of connection 4A (see Fig. 4). This nesting solution was chosen because it removes two uncertainties at a time and seems more plausible.

Inferred Population History and Coalescence Analysis

There were significant associations between haplotype clades and geographic distribution on all clade levels (Table 3). The oldest inferred event was a past fragmentation between clade 4-1 (populations in the foothills of the French Sea-Alps) and clade 4-2 (all other populations) (Fig. 5). Within clade 4-1, a continuous range expansion (CRE) to the north was inferred. Within clade 4-2, there was indication for isolation by distance (IbD). The clades nested in 3-4 and 3-5 were also characterised by recurrent gene flow restricted by distance. In clade 3-1, no distinction between isolation by distance and past fragmentation was possible. Due to the lack of significant clade distances, no inference was possible for clades 3-2 and 3-3. The most interesting inference on the lower clade levels was the isolation by distance with some long distance colonizations of clade 1-17 into Germany. Geographical distributions of major clades and inferences of population history are plotted in Figure 5.

There were no significant APCD values at the haplotype level. At the one-step level, several values were significant; however, since no fragmentation was inferred for this level, no secondary contact was identified. On the two-step level, past fragmentation of clades within clade 3-4 was inferred. The significantly high APCD for population LAU is therefore indicative of a secondary contact zone. With increasing clade level, APCDs for four populations remained significant up to the three-step level. However, because no fragmentation event was inferred on this level for the populations MUR and LAU, these significant values did not signify a secondary contact zone. On the four-step level, only in populations VAL, SAL, and FDE did the APCDs remain significantly different from zero. Because past fragmentation events were inferred for populations at the four-step level, the significant APCDs indicate two secondary contact zones at the highest clade level. These are situated around the Geneva Basin and in the Sea Alps.

The island model of gene flow had a significantly higher likelihood than the NCA or the northern expansion model (LRTs, $P < 0.05$) (Fig. 7). A common feature of the NCA and Island-like models was a net gene flow from the populations left of the Rhône (RL) to all other neighboring regions. Thus, both coalescence analysis and NCA suggested a source-sink model of gene flow with the region RL as source.

DISCUSSION

Inferring Intraspecific Sequence Evolution

The ML and MP trees and the SP cladogram exhibited similar topologies (Figs. 2A, B, and 4). However, both ML and MP trees failed to resolve most of the relationships

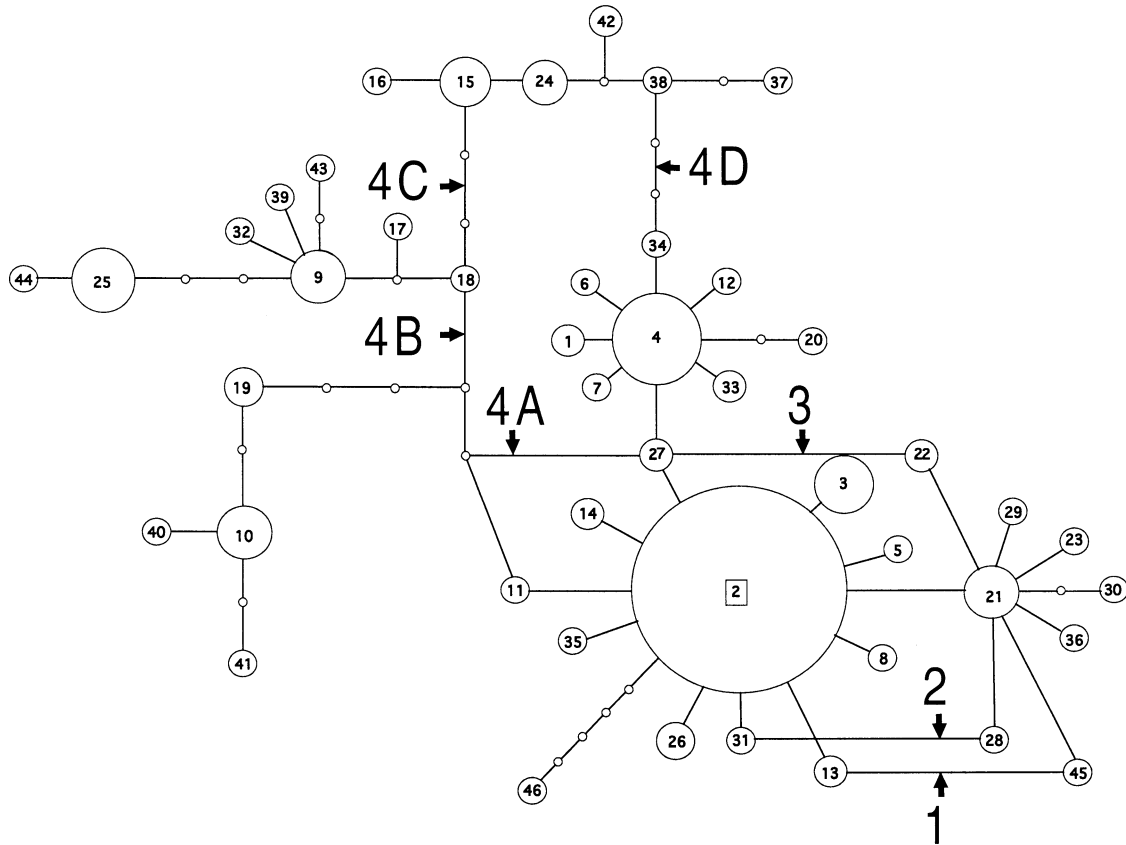


FIG. 3. Statistical-parsimony cladogram and loop solutions. Potential connections to resolve loops 1–4 are indicated by arrows. Haplotypes are designated by names defined in Table 2. Zeros indicate haplotype states that are necessary intermediates but were not present in the sample. Each line represents a single mutational step connecting two haplotypes.

among haplotypes as well as among higher-level clades. The polytomy in the center of the MP and ML trees prevents a meaningful interpretation, whereas the SP cladogram is more resolved, although it includes four loops. This result highlights how network approaches as the SP method may portray more effectively intraspecific evolution. In the ML or MP representations, all taxa are located at the tip of the branches. This appears to be an inappropriate depiction of evolutionary relationships within species, where many haplotypes can be derived from the same ancestral haplotypes, and ancestral haplotypes are commonly extant (Posada and Crandall 2001).

Population History

The population history of *C. unifasciata* as inferred by NCA appeared to be complex, involving historic events as well as recurrent gene flow on virtually all clade levels. We will focus the discussion on the inferences with implications on a broader geographic scale.

The oldest event that can be traced is a fragmentation, separating populations in the foothills of the French Sea-Alps from populations closer to the river Rhône. The fact that this fragmentation is not marked by several mutation steps in the cladogram might be due to the retention of ancient polymorphisms in each of the major clades, which appears to be typical for the population structure of terrestrial gastropods (Thomaz et al. 1996). Given that the molecular clock was

not rejected (LRT P-value > 0.05), we could estimate the time of the population fragmentation to have occurred $280,000 \pm 20,000$ years ago. This estimate suggests that the fragmentation might have been caused by a Pleistocene climate event, perhaps the Riss glaciation, isolating different *C. unifasciata* populations in separate refuges.

A continuous range expansion from the French Sea-Alps to the North seems to have taken place after the fragmentation. This expansion most likely occurred along the valleys of Durance, Isère, and upper Rhône Rivers, following the north-south running mountain ranges of the French pre-Alps to the Geneva Basin (Fig. 5). Around the Geneva Basin, haplotypes belonging to clade 4-1 and 4-2 were found in sympatry in two different sampling sites. Significantly large average population clade distances at the four-step clade level of fragmentation for populations FDE and SAL suggest a secondary contact zone of the formerly fragmented four-step clades in this area (Figs. 5, 6). Inferences for clade 4-2 and the clades nested within suggest an overall pattern of isolation by distance, except for the range expansion into Germany.

The fact that the island model provided a better fit than the model suggested by the NCA does not contradict these inferences, as indeed an island model of gene flow is equally restricted by distance. In fact, the migration matrix associated with the island model suggested gene flow towards the German populations, which is in agreement with the northern

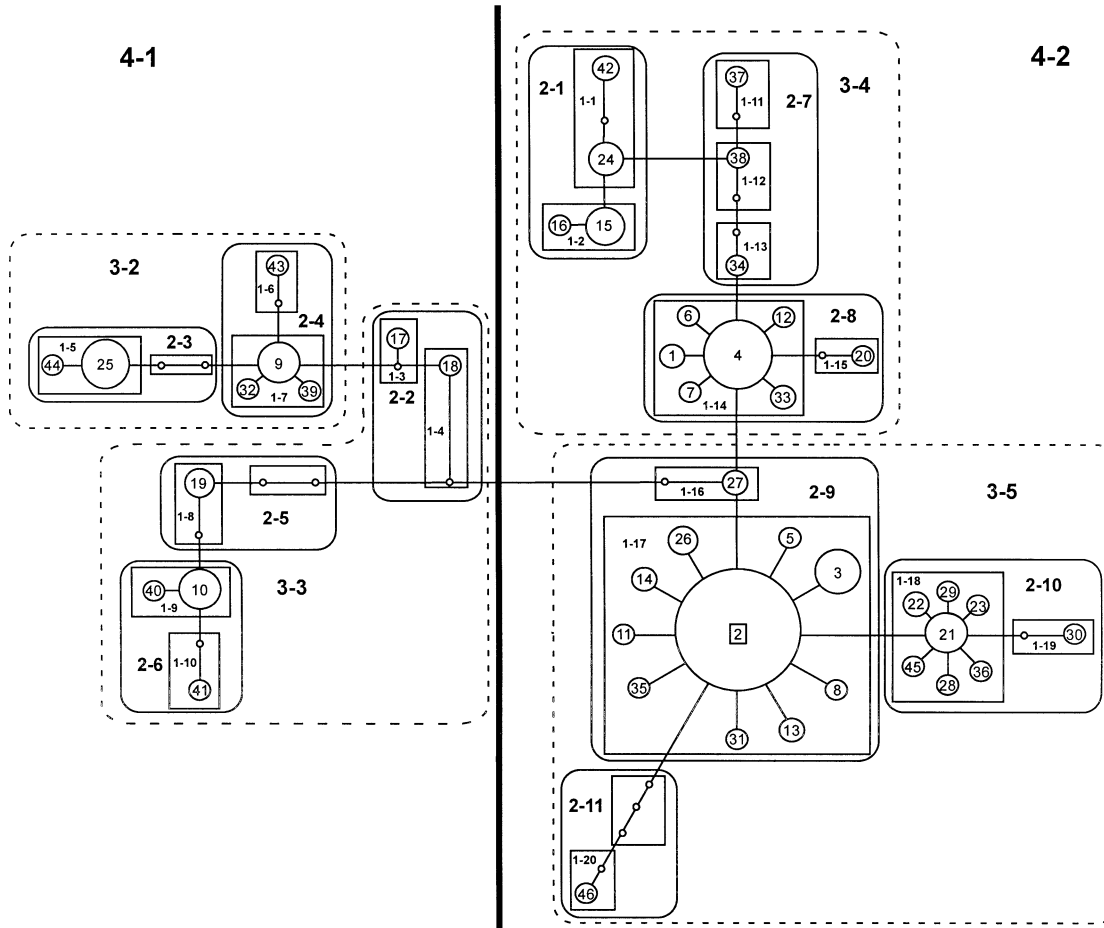


FIG. 4. Statistical-parsimony cladogram with completed nesting design. Haplotypes belonging to the same clade level are boxed up to clade level 4-x. Clade level designations are given within each box that contains observed haplotypes.

expansion suggested by NCA (Fig. 7B). In addition, the coalescent analysis suggested that the region west of the Rhône and north of the Durance (RL) might have been acting as the source for all other regions within clade 4-2 (Fig. 7). Indeed, haplotype 2, which is the most probable root for the entire cladogram, is well represented in the RL region. This sug-

gests that RL is probably the area where the most recent common ancestor of the sample lived, and that the presence of *C. unifasciata* in the Burgundy area might have originated from a northward expansion along the Rhône Valley. A post-glacial expansion scenario seems plausible, because the cover of the Geneva Basin by the Rhône glacier during the Würm

TABLE 3. Chi-squared test of geographical association of clades and biological inference from the NCA analysis. Probability *P* is the probability of obtaining a χ^2 -statistic larger than or equal to the observed statistic by randomly permuting the original contingency table 9999 times. Inferences were obtained following the key in Templeton (1998). Abbreviations for the inferences are: CRE, continuous range expansion; IbD, isolation by distance; and LDC, long distance colonization.

Clades nested with	Permutational χ^2 statistic	<i>P</i>	Chain of inference	Inference
Clade 1-1	7.00	0.046	2-11-12-13-14-NO	CRE or LDC
Clade 1-14	66.93	0.044	2-3-4-9-10-NO	IbD or past fragmentation
Clade 1-17	459.69	0.000	2-3-5-6-7-YES	IbD with some LDC
Clade 2-1	7.47	0.024	2-only interior clades	Inconclusive outcome
Clade 2-9	64.65	0.020	No significant clade distances	—
Clade 3-1	17.00	0.016	2-3-4-9-10-NO	IbD or past fragmentation
Clade 3-2	21.43	0.000	No significant clade distances	—
Clade 3-4	22.42	0.022	2-3-4-9-NO	Past fragmentation
Clade 3-5	124.90	0.044	2-3-4-NO	IbD
Clade 4-1	97.99	0.000	2-11-12-NO	CRE
Clade 4-2	83.70	0.000	2-3-4-NO	IbD
Entire cladogram	187.65	0.000	2-3-4-9-NO	Past fragmentation

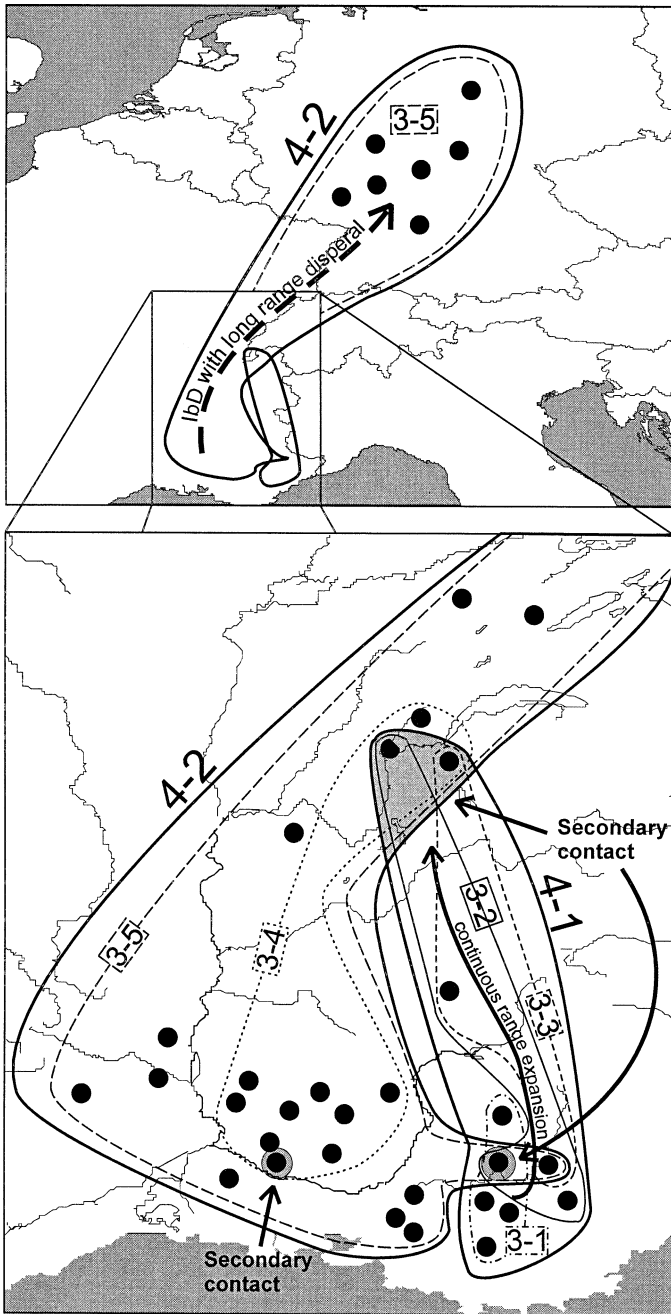


FIG. 5. Geographical distribution of three- and four-step clades and inferred events. Area of inferred secondary contact between four-step clades is shaded.

glaciation (Frenzel 1981) likely prevented the presence of *C. unifasciata*. Further evidence for a corridor migration are the clade distributions along the migration axis (Fig. 5) and their bondage to the corridor by geographic barriers (Cruzan and Templeton 2000). These conditions were met, because the distribution of clade 4-2 is limited to the Rhône Valley in the area in question and it is restricted to it by unsuitable habitat in the west and high mountains in the east, where no populations could be found.

The presence of haplotypes from both four-step clades in

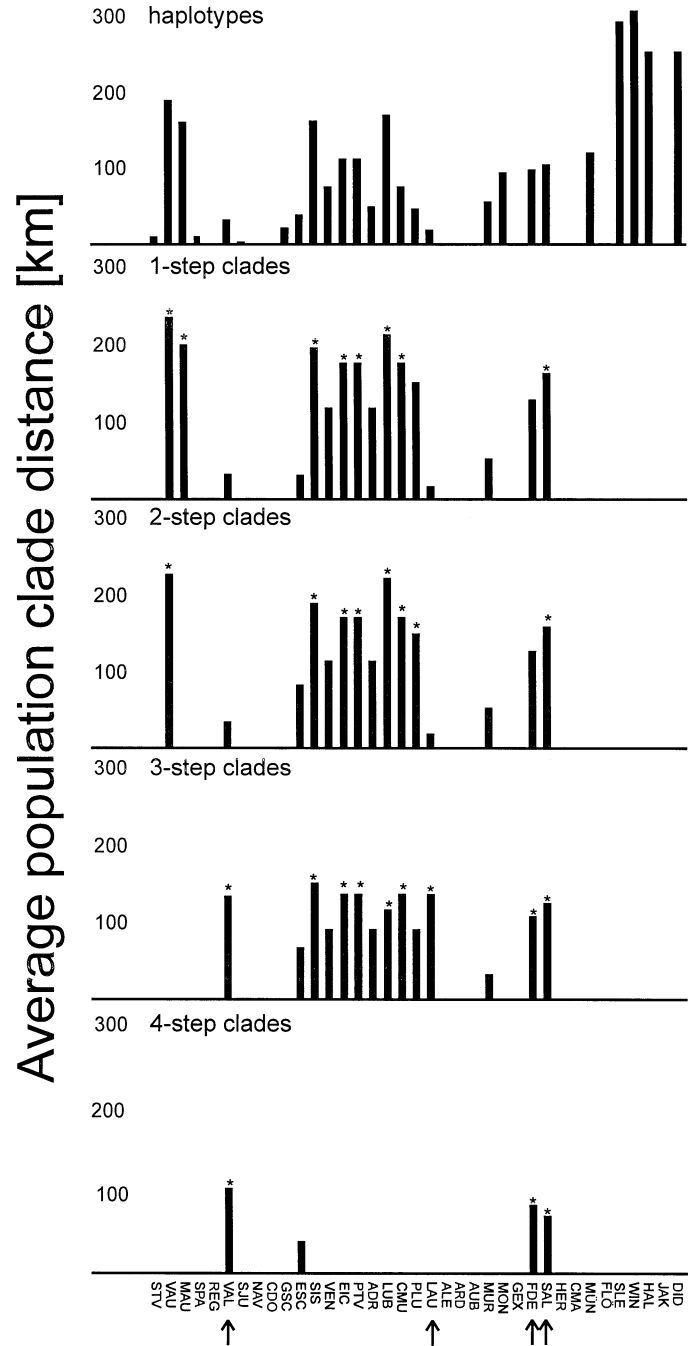


FIG. 6. Average population clade distance for all clade levels. Significantly large clade distances at the 1% level ($P < 0.01$) are marked by an asterisk. Populations, where secondary contact was inferred, are marked by an arrow.

populations VAL and ESC in southeastern France indicated a secondary contact (Fig. 5). Even though the APCDs in these populations are quite large at the four-step level, only the values for VAL were significantly high (Fig. 6). The presence of the haplotypes on the Plateau de Valensole and the Sea-Alps (VAL, ESC) found otherwise in populations further to the west (LUB, VAU, STV, SPA) may be explained by the annual transhumance of sheep from winter pastures on the Plain de la Crau close to the Rhône delta to summer pastures

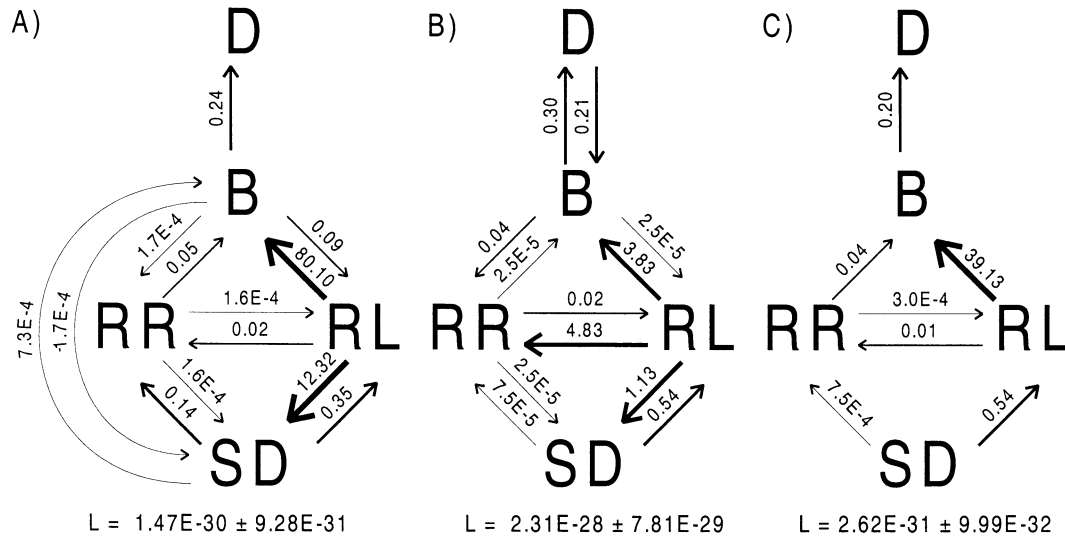


FIG. 7. Effect of differing migration matrices on estimated likelihoods in the coalescence analysis for pooled populations (Table 1) of clade 4-2. (A) NCA model: migration rate estimates based on a gene-flow model reflecting the inferences of NCA. (B) Island model: migration rate estimates based on a gene-flow model reflecting an island-migration pattern. (C) Expansion model: migration rate estimates based on a gene-flow model reflecting an expansion pattern. The estimated migration rates given as Nm (where N is the effective population size, and m is the migration rate per individual per generation) are represented by arrows whose line width is proportional to the gene-flow rate. Likelihoods for the migration patterns having created the data are given below each pattern as the mean, errors are 95% confidence limits of the mean. Model B has a significantly larger likelihood than the other two models ($P < 0.05$).

in the higher Sea-Alps during the last centuries. These trails connected suitable habitat for *C. unifasciata*, and passive transport by sheep has been suggested to be a major source of dispersal in land snails (Dörge et al. 1999; Pfenninger and Magnin, 2001).

All inferred range expansions in *C. unifasciata* involved large distances compared to the dispersal capacity of the individual snail (Bahl et al. 1996), so that we can hypothesize that these long distance colonizations were achieved by passive dispersal. We can only speculate about the vectors, but the most likely seem to be larger mammals, even though other agents like birds or even wind have been implicated in the dispersal of snails (Boag 1986; Kirchner et al. 1997).

Colonization of the Northern Species Range

Passive transport by migrating animals was probably also the mechanism for the expansion to the German populations from southern sources. Additional but indirect support for an inference of a colonization or a range expansion in general arose from the significant APCDs for German populations on the haplotype level (Fig. 6). A large APCD indicates that a population consists of at least some haplotypes that are far away from their clade center. A significantly large APCD is therefore indicative of migration towards this particular population, even if the overall evidence for the respective clade points to another inference (see inference for clade 1-17 in Table 3).

The inference at the one-step level suggests that the expansion of the range into Germany is a relatively recent event. This colonization occurred probably soon after the onset of the climate warming after the last Pleistocene glaciation. The habitat preference of sparsely covered, open grass- and scrubland suggests that the expansion happened mainly in the pre-

boreal phase. This is in concordance with a first occurrence of *C. unifasciata* shells in early Holocene sediments of the upper Rhine Valley (Puissegur 1978). The migration route went presumably through the Rhône Valley and then along the Jura Mountains into Germany. The findings for the southern species range stress the importance of river valleys as migration corridors for *C. unifasciata*. This and the location of the German populations in or close to river valleys suggest that river corridors had also relevance for the spread into Germany.

Conclusions

The phylogeographic analysis of *C. unifasciata* suggested the occurrence of a fragmentation of the species in different refuges in south eastern France 300,000 years ago, followed by and independent range expansion via corridor migration and a secondary contact zone in the Geneva Basin. The inferred long distance colonization of Germany is most likely of postglacial origin.

An unexpected result from this study is the location of refugial populations in the south of France. Many phylogeographic studies conclude that the three main glacial refuges for current central and northern European biota were located south of the Pyrenees, south of the Alps, and on the Balkans (Taberlet et al. 1998; Hewitt 2000). Although these are the traditionally recognized refuges for land molluscs as well (Ant 1966), *C. unifasciata* does not seem to have occurred so far south. Indeed the inferences from one species is not sufficient to draw wide-ranging conclusions, nevertheless these results may add new details to the general biogeographic patterns described by Taberlet (1998) and Hewitt (2000). Further studies with more species are needed to show whether the inferred biogeographic pattern of refuges in the south of

France represents a common model for land gastropods and other taxa with restricted vagility.

The increasing availability of powerful statistical techniques, well rooted in population genetics, allows today for a very welcomed "statistical phylogeography." Within this framework, the NCA is an approach based on expectations closely linked to the biological inference. The coalescent analyses estimate different parameters under explicit population genetic models, but the value of the statistics need to be linked afterwards to some biological process. Here we have utilized and combined these two statistical approaches in order to obtain detailed insights on the history of our species of interest. Indeed, the use of rigorous statistics on top of phylogeographic information will increase during the next years. We hope that this study encourages the use of such a powerful framework to better understand the natural history of populations.

ACKNOWLEDGMENTS

We thank K. Schwenk, J. W. Sites and K. A. Crandall for helpful discussions. The comments of two anonymous referees and K. Shaw improved the manuscript further.

LITERATURE CITED

- Ant, H. 1966. Die Bedeutung der Eiszeiten für die rezente Verbreitung der europäischen Landgastropoden. *Malacologia* 5: 61–62.
- Avise, J. C. 2000. *Phylogeography: the history and formation of species*. Harvard Univ. Press, Cambridge, MA.
- Bahl, A., and M. Pfenninger. 1996. A rapid method for isolation of DNA using laundry detergent. *Nucleic Acids Res.* 24: 1587–1588.
- Bahl, A., M. Pfenninger, H. Bamberger, M. Frye, and B. Streit. 1996. Survival of snails in fragmented landscapes. Pp. 329–343 in J. Settele, C. R. Margules, P. Poschlod, and K. Henle, eds. *Species survival in fragmented landscapes*. Kluwer, Dordrecht, The Netherlands.
- Baur, B. 1993. Population structure, density, dispersal, and neighbourhood size in *Arianta arbustorum* (Linnaeus 1758) (Pulmonata: Helicidae). *Ann. Naturhist. Mus. Wien* 94:307–321.
- Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and population numbers of two populations using a coalescent approach. *Genetics* 152:763–773.
- . 2001. Maximum-likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA* 98: 4563–4568.
- Boag, D. A. 1986. Dispersal in pond snails: potential role of waterfowl. *Can. J. Zool.* 64:904–909.
- Bossart, J. L., and D. Pashley Prowell. 1998. Genetic estimates of population structure and gene flow: limitations, lessons, and new directions. *Trends Ecol. Evol.* 13:202–206.
- Castelloe, J., and A. R. Templeton. 1994. Root probabilities for intraspecific gene trees and neutral coalescent theory. *Mol. Phylog. Evol.* 3:102–113.
- Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. *Mol. Ecol.* 9:1657–1660.
- Cowie, R. H. 1984. Density, dispersal and neighbourhood size in the land snail *Theba pisana*. *Heredity* 52:391–401.
- Crandall, K. A. 1996. Multiple interspecies transmissions of human and simian T-cell leukaemia/lymphoma virus type I sequences. *Mol. Biol. Evol.* 13:115–131.
- Crandall, K. A., and A. R. Templeton. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134:959–969.
- Crandall, K. A., A. R. Templeton, and C. F. Sing. 1994. Intraspecific phylogenetics: problems and solutions. Pp. 273–297 in R. W. Scotland, D. J. Siebert, and D. M. Williams, eds. *Models in phylogeny reconstruction*. Clarendon Press, Oxford, U.K.
- Cruzan, M. B., and A. R. Templeton. 2000. Paleogeography and coalescence: phylogeographic analysis of hypothesis from the fossil record. *Trends Ecol. Evol.* 15:491–496.
- Donnelly, P., and S. Tavaré. 1996. The ages of alleles and a coalescent. *Adv. Appl. Probab.* 18:1–19.
- Donnelly, P., and S. Tavaré. 1995. Coalescents and genealogical structure under neutrality. *Ann. Rev. Genet.* 29:401–421.
- Dörge, N., C. Walther, B. Beinlich, and H. Plachter. 1999. The significance of passive transport for dispersal in terrestrial snails (Gastropoda, Pulmonata). *Z. Ökol. Natursch.* 8:1–10.
- Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.
- Excoffier, L., and A. Langaney. 1989. Origin and differentiation of human mitochondrial DNA. *Am. J. Hum. Genet.* 44:73–85.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum-likelihood approach. *J. Mol. Evol.* 17:368–376.
- . 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791.
- Frenzel, B. 1981. The history of flora and vegetation during the Quaternary. *Fortschr. Bot.* 43:1–98.
- Fu, Y. X., and W. H. Li. 1999. Coalescing into the 21st century: an overview and prospects of coalescent theory. *Theor. Popul. Biol.* 56:1–10.
- Gittenberger, E. 1993. On *Trochoidea geyeri* (Soós 1926) and some conchological similar taxa (Mollusca: Gastropoda Stylommatophora: Hygromiidae). *Zool. Med.* 67:303–320.
- Godan, D. 1979. *Schadschnecken und ihre Bekämpfung*. Ulmer, Stuttgart, Germany.
- Goodfriend, G. A. 1992. The use of land snail shells in paleoenvironmental reconstruction. *Quatern. Sci. Rev.* 11:665–685.
- Golding, G. B. 1987. The detection of deleterious selection using ancestors inferred from an evolutionary history. *Genet. Res.* 49: 71–82.
- Griffiths, R. C., and M. Bahlo. 1998. Genetree 8.3. available at: <http://www.maths.monash.edu.au/~mbahlo/mpg/gtree.html>.
- Guiller, A., M. A. Coutellec-Vreto, L. Madec, and J. Deunff. 2001. Evolutionary history of the land snail *Helix aspersa* in the western Mediterranean: preliminary results inferred from mitochondrial sequences. *Mol. Ecol.* 10:81–87.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* 22:160–174.
- Hewitt, G. 1996. Some genetic consequences of ice ages and their role in divergence and speciation. *Biol. J. Linn. Soc.* 58: 247–276.
- . 2000. The genetic legacy of the Quaternary ice ages. *Nature* 405:907–913.
- Kelly, F. P. 1977. Exact results from the Moran neutral allele model. *Adv. Appl. Prob.* 9:197–201.
- Kerney, M. P., R. A. D. Cameron, and J. H. Jungbluth. 1983. *Die Landschnecken Nord- und Mitteleuropas*. Parey, Hamburg-Berlin.
- Kingman, J. F. C. 1982. The coalescent. *Stoch. Proc. Appl.* 13: 235–248.
- Kirchner, C. H., R. Kratzner, and F. W. Welter-Schultes. 1997. Flying snails: How far can *Truncatellina* (Pulmonata: Vertiginidae) be blown over the sea? *J. Moll. Stud.* 63:479–487.
- Lydeard, C., W. E. Holzner, M. N. Schnare, and R. R. Gutell. 2000. Evolutionary patterns and phylogenetic content of the molluscan mitochondrial LSU rDNA sequences and rRNA secondary structure. *Mol. Phylog. Evol.* 15:83–102.
- Magnin, F. 1989. Les distributions pléistocène et actuelle de *Trochoidea (Xeroclausa) geyeri* (Soós 1926) (Gastropoda, Helicidae) dans le Sud-Est de la France: un exemple de disjonction d'aire liée au réchauffement post-glaciaire. *Bull. Soc. Géol. France* 4:779–786.
- . 1993. Competition between two land gastropods along altitudinal gradients in south-eastern France: neontological and palaeontological evidence. *J. Mol. Stud.* 58:445–454.

- Maynard Smith, J. 1970. Population size, polymorphism, and the rate of non-Darwinian evolution. *Am. Nat.* 104:231–237.
- Nei, M. 1987. *Molecular evolutionary genetics*. Columbia Univ. Press, New York.
- Nei, M., and M. W. Feldman. 1972. Identity of genes by descent within and between populations under mutation and migration pressures. *Theor. Popul. Biol.* 3:460–465.
- Pfenninger, M., A. Bahl, and B. Streit. 1996. Isolation by distance in a population of a small land snail *Trochoidea geyeri*: evidence from direct and indirect methods. *Proc. R. Soc. Lond. B* 263:1211–1217.
- Pfenninger, M., and F. Magnin. 2001. Phenotypic evolution and hidden speciation in *Candidula unifasciata ssp.* (Helicellinae, Gastropoda) inferred by 16S variation and quantitative shell traits. *Mol. Ecol.* 10:2541–2554.
- Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- . 2001. Intraspecific gene genealogies: trees grafting into cladograms. *Trends Ecol. Evol.* 16:37–45.
- Posada, D., K. A. Crandall, and A. R. Templeton. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Mol. Ecol.* 9:487–488.
- Puissegur, J. J. 1978. Les mollusques des séries loessiques à Achenheim. *Rech. Géogr. Strasbourg* 7:71–96.
- Ross, T. K. 1999. Phylogeography and conservation genetics of the Iowa Pleistocene snail. *Mol. Ecol.* 8:1363–1373.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Stewart, J. R., and A. M. Lister. 2001. Cryptic northern refugia and the origins of the modern biota. *Trends Ecol. Evol.* 16:608–613.
- Swofford, D. L. 1998. PAUP*. *Phylogenetic analysis using parsimony (*and Other Methods)*. Vers. 4. Sinauer, Sunderland, MA.
- Taberlet, P., L. Fumagalli, A. G. Wust-Saucy, and J. F. Cosson. 1998. Comparative phylogeography and postglacial colonisation routes in Europe. *Mol. Ecol.* 7:453–464.
- Takahata, N. 1988. The coalescent in two partially isolated diffusion populations. *Genet. Res.* 52:213–222.
- Templeton, A. R. 1998. Nested clade analyses of phylogeographic data: testing hypothesis about gene flow and population history. *Mol. Ecol.* 7:381–397.
- . 2001. Using phylogeographic analyses of gene trees to test species status and processes. *Mol. Ecol.* 10:779–791.
- Templeton, A. R., and C. F. Sing. 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.
- Templeton, A. R., K. A. Crandall, and C. F. Sing. 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., E. Routman, and C. A. Phillips. 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.
- Thomaz, D., A. Guiller, and B. Clarke. 1996. Extreme divergence of mitochondrial DNA within species of pulmonate land snails. *Proc. R. Soc. Lond. B* 263:363–368.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 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.
- Watterson, G. A. 1985a. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 27:256–276.
- Watterson, G. A. 1985. The genetic divergence of two populations. *Theor. Popul. Biol.* 27:298–317.
- . 1976. Reversibility and the age of an allele. I. Moran's infinitely many neutral alleles model. *Theor. Popul. Biol.* 10:239–253.
- Watterson, G. A., and H. A. Guess. 1977. Is the most frequent allele the oldest? *Theor. Popul. Biol.* 10:239–253.
- Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analysis. *Trends Ecol. Evol.* 11:367–372.

Corresponding Editor: K. Shaw