

Research article

Open Access

## Evidence for survival of Pleistocene climatic changes in Northern refugia by the land snail *Trochoidea geyeri* (Soós 1926) (Helicellinae, Stylommatophora)

Markus Pfenninger\*<sup>1,3</sup>, David Posada<sup>2</sup> and Frédéric Magnin<sup>1</sup>

Address: <sup>1</sup>IMEP URA-CNRS 1152 Case 451, Faculté des Sciences et Techniques de St. Jerome, F-13997 Marseille Cedex 20, France, <sup>2</sup>Department of Zoology, Brigham Young University, 574 WIDB, Provo, UT 84602-5255, USA and <sup>3</sup>Abteilung Ökologie und Evolution, Zoologisches Institut der J.W.Goethe-Universität, Bio-Campus Siesmayerstraße, D-60054 Frankfurt/Main, Germany

Email: Markus Pfenninger\* - Pfenninger@zoology.uni-frankfurt.de; David Posada - dposada@uvigo.es; Frédéric Magnin - frederic.magnin@univ.u-3mrs.fr

\* Corresponding author

Published: 29 April 2003

Received: 10 March 2003

BMC Evolutionary Biology 2003, 3:8

Accepted: 29 April 2003

This article is available from: <http://www.biomedcentral.com/1471-2148/3/8>

© 2003 Pfenninger et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** The study of organisms with restricted dispersal abilities and presence in the fossil record is particularly adequate to understand the impact of climate changes on the distribution and genetic structure of species. *Trochoidea geyeri* (Soós 1926) is a land snail restricted to a patchy, insular distribution in Germany and France. Fossil evidence suggests that current populations of *T. geyeri* are relicts of a much more widespread distribution during more favourable climatic periods in the Pleistocene.

**Results:** Phylogeographic analysis of the mitochondrial 16S rDNA and nuclear ITS-1 sequence variation was used to infer the history of the remnant populations of *T. geyeri*. Nested clade analysis for both loci suggested that the origin of the species is in the Provence from where it expanded its range first to Southwest France and subsequently from there to Germany. Estimated divergence times predating the last glacial maximum between 25–17 ka implied that the colonization of the northern part of the current species range occurred during the Pleistocene.

**Conclusion:** We conclude that *T. geyeri* could quite successfully persist in cryptic refugia during major climatic changes in the past, despite of a restricted capacity of individuals to actively avoid unfavourable conditions.

### Background

The predicted global climate change will undoubtedly have a major impact on the distribution ranges and survival of many animal and plant species [1]. Taxa with poor active dispersal abilities and fragmented habitats are especially likely to be affected from a shifting climate. The study of the reactions of such species on historical climate changes might help to understand the impact of future climatic changes on species distribution and biodiversity.

Land snails are suitable organisms to address this issue. While some snails have achieved a cosmopolitan distribution via passive anthropogenic dispersal (reviewed in [2]), most species show limited distribution ranges. Moreover, many species have particular habitat requirements, which result in a patchy, insular habitat distribution [3]. In general, active dispersal is quite restricted in snails (*e.g.*, [4–6]), preventing them to escape changing ecological conditions. So it is commonplace to assume that only few, if

any, land snail species were able to survive the pleniglacial phases of the Pleistocene in northern parts of Europe [7]. However, due to the good preservation of gastropod shells in loess deposits, snail fossils are relatively abundant [8]. Low dispersal capacities and presence in the fossil record are features that render an organism particularly amenable to phylogeographic study [9]. While low vagility preserves patterns of genetic variation arisen in the past, fossils allow for the integration of knowledge about past distributions into the formulation of phylogeographic hypotheses.

Despite their suitability for phylogeographic study, only a few such studies have been carried out in land snails. With one exception [10], these studies relied on large Helicoid snails of the genera *Cepaea* and *Helix* [11–13], which are especially prone to displacement by human activities. Unfortunately, anthropogenic dispersal may easily result in phylogeographic patterns that do not reflect the impact of historical climate changes. In order to understand the consequences of climate changes on the distribution and genetic structure of snail populations, the study of species with rare human dispersal should be preferred.

*Trochoidea geyeri* (Soós 1926) is a small land snail of the Helicellinae subfamily within the Helicidae. Its active dispersal capacity is about 3 m during its one-year lifetime [6]. The mating system of the hermaphroditic species is obligately outcrossing. Today, the species range comprises parts of Germany and the south of France, showing a discontinuous, patchy distribution. *T. geyeri* fossils are relatively abundant. In loess deposits, the presence of *T. geyeri* shells has been reported since the early Pleistocene [14]. The subfossilised shell deposits in southern England and large parts of France are correlated with the widespread occurrence of rather arid cold steppe vegetation formations [15]. These formations are associated with transitional phases of Pleistocene climate cycles, covering parts of Europe even during maximal glacial expansion [16] thus providing the potential for local refugia. Both Pleistocene interstadial and pleniglacial periods resulted in altitudinal and latitudinal shiftings of these formations, as well as in reductions in their extent. *T. geyeri* is found today in open calcareous or loessic grass and scrublands with a sparse vegetation cover on mountaintops, carstic highland plateaus and disturbed pastures, which are thought to constitute ecological refuges [17]. The fossil record suggests that the population history of *T. geyeri* is linked to palaeoclimate changes [15]. The latitudinal shifts of suitable habitat during Pleistocene across Europe, driven by climate change, were anticipated by *T. geyeri* in the fossil record with remarkably short time lags. In other words, the species can be detected in the fossil record very soon after the onset of a suitable climate phase [15]. This raises the question whether these recurrent range expansions originated

from one or few major refuges or from cryptic refuges strewn all over the species range. Such northern refugia were recently identified for several plants, fishes and small mammals [18,19].

These entire characteristics make *T. geyeri* an ideal organism for the study of the role of historical climate changes on species distribution and population structure. Here we explicitly addressed the question of whether *T. geyeri*, an organism with restricted dispersal abilities, survived the Pleistocene glacial periods *in situ* in local northern refugia, or instead went extinct to re-colonise in recent times the northern range of its present day distribution from the south.

To answer this question, we analysed mitochondrial 16S rDNA and nuclear ITS-1 variation within a statistical phylogeographic framework using coalescent and nested clade analyses to test hypotheses on population history.

## Results

### **Mitochondrial and nuclear haplotype diversity**

The amplified 16S rDNA sequences were on average 356 base pairs long. Thirty-one haplotypes were identified, with 39 variable sites (Table 2). Three indels were observed, all in loop coding regions of the 16S rDNA according to the molluscan consensus secondary structure model of Lydeard *et al.* [20]. The number of substitutions between haplotypes ranged from 1 to 23 (0.27–6.41%). The transition/transversion ratio and gamma shape parameter ( $\alpha$ ) estimates were 2.2597 and 0.0079. Such a small value of  $\alpha$  indicates that there is a strong rate variation among sites: a few sites are changing very fast, while the rest do not change, or change very slowly. With the exception of haplotype 20, which was found in two neighbouring sampling localities on the Causse de Larzac, all haplotypes were restricted to a single population (Table 2). The number of haplotypes per sampling site ranged from one to five, and decreased from south to north. The mean sequence divergence among populations was  $2.88\% \pm 1.52$  (s.d.), while the mean within population diversity was  $0.28\% \pm 0.31$  (s.d.).

The ITS-1 fragment was between 525 and 527 bp long. There were no signs of heterozygosity in the data; the sequences could therefore be treated as haplotypes. Six haplotypes, defined by six variable sites, were found. Within all populations only a single haplotype could be detected. There were no signs of recombination

### **Phylogenetic relationships of haplotypes**

In the statistical parsimony (SP) analysis of 16S variation, haplotypes separated by up to eight mutational steps had a greater than 95% probability of being connected in a parsimonious fashion (i.e., no over imposed mutations),

**Table 1: Sampled populations, abbreviations used, geographical location and number of sampled individuals.**

| Population              | Abbreviation | Location        | Sample size |
|-------------------------|--------------|-----------------|-------------|
| Schlüchtern/Hohenzell   | SHZ          | 50°21'N 09°33'E | 7           |
| Eisleben/Süßer See      | ESS          | 51°35'N 11°40'E | 5           |
| Münsingen               | MUN          | 48°25'N 09°30'E | 8           |
| Mont Ventoux/Sommet     | MVS          | 44°11'N 05°16'E | 8           |
| Mont Ventoux/Les Rabets | MVR          | 44°08'N 05°10'E | 5           |
| Montagne de Lure        | ML           | 44°07'N 05°28'E | 7           |
| Plateau de Vaucluse     | PTV          | 43°58'N 05°28'E | 7           |
| Grand Luberon           | LUB          | 43°49'N 05°28'E | 6           |
| Sainte Victoire         | STV          | 43°32'N 05°43'E | 5           |
| Chaine des Etoiles      | CDE          | 43°24'N 05°28'E | 9           |
| Sainte Baume            | STB          | 43°19'N 05°47'E | 5           |
| Cause de Larzac 1       | LAR1         | 44°00'N 03°20'E | 3           |
| Causse de Larzac 2      | LAR2         | 43°59'N 03°10'E | 5           |
| Causse de Larzac 3      | LAR3         | 44°01'N 03°15'E | 6           |

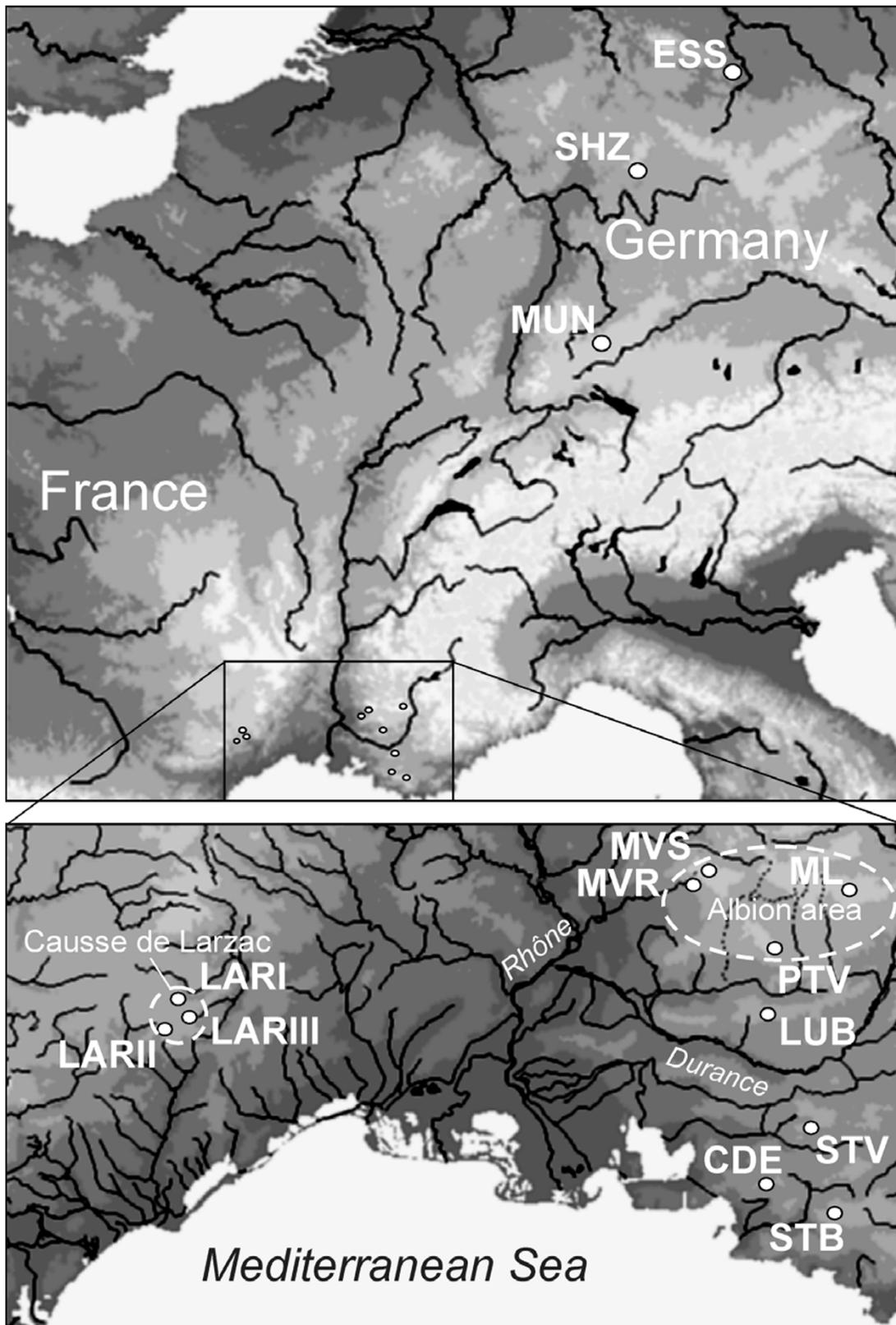
**Table 2: a) Distribution of *T. geyeri* 16S rDNA and b) ITS-I haplotypes (columns) at each sampled location (rows).**

a)

|      | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| LUB  | 2 | 1 | 1 | 2 |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| MVS  |   |   |   |   | 3 | 4 | 1 |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| MVR  |   |   |   |   |   |   |   | 5 |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| PTV  |   |   |   |   |   |   |   |   |   |    | 2  | 3  | 2  | 1  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| ML   |   |   |   |   |   |   |   |   | 6 | 1  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| STV  |   |   |   |   |   |   |   |   |   |    |    |    |    |    | 2  | 3  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| CDE  |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    | 9  |    |    |    |    |    |    |    |    |    |    |    |    |    |
| STB  |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    | 3  | 2  |    |    |    |    |    |    |    |    |    |    |    |
| LAR1 |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    | 3  |    |    |    |    |    |    |    |    |    |    |
| LAR2 |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    | 3  | 1  | 1  |    |    |    |    |    |    |    |    |
| LAR3 |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    | 2  | 1  | 1  | 1  | 1  |    |    |    |    |
| SHZ  |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 6  | 1  |    |    |
| MUN  |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 8  |    |
| ESS  |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 5  |

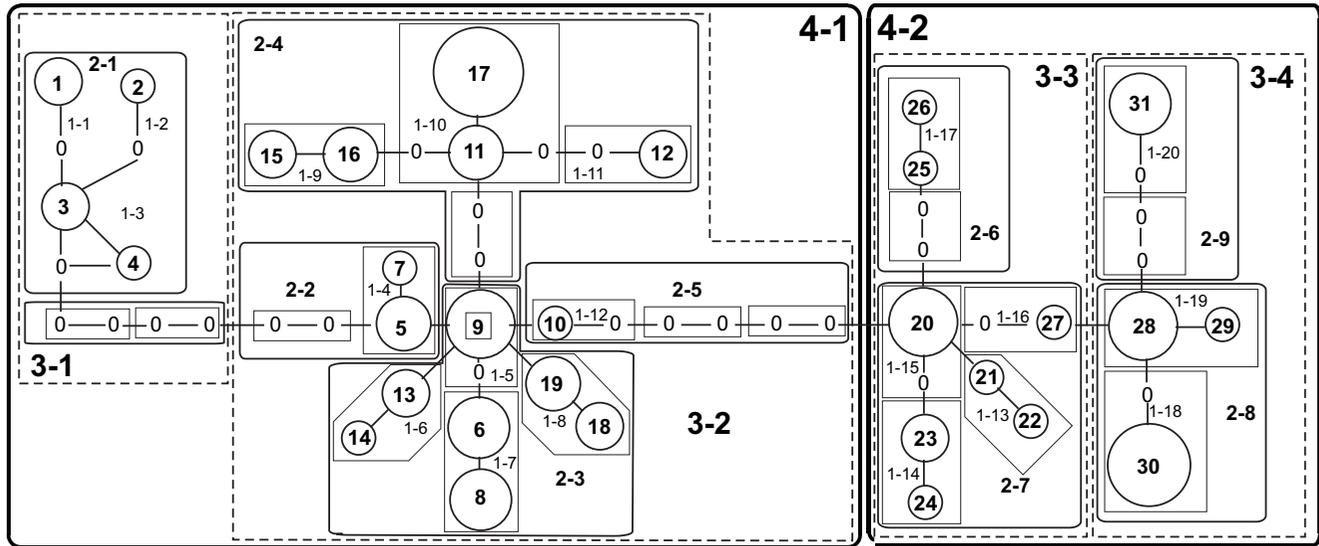
b)

|      | n1 | n2 | n3 | n4 | n5 | n6 |
|------|----|----|----|----|----|----|
| LUB  | 6  |    |    |    |    |    |
| MVS  |    | 8  |    |    |    |    |
| MVR  |    | 5  |    |    |    |    |
| PTV  |    |    | 7  |    |    |    |
| ML   |    | 7  |    |    |    |    |
| STV  |    |    | 5  |    |    |    |
| CDE  |    |    | 9  |    |    |    |
| STB  |    | 5  |    |    |    |    |
| LAR1 |    |    |    | 3  |    |    |
| LAR2 |    |    |    | 5  |    |    |
| LAR3 |    |    |    | 6  |    |    |
| SHZ  |    |    |    |    | 7  |    |
| MUN  |    |    |    |    | 5  |    |
| ESS  |    |    |    |    |    | 8  |

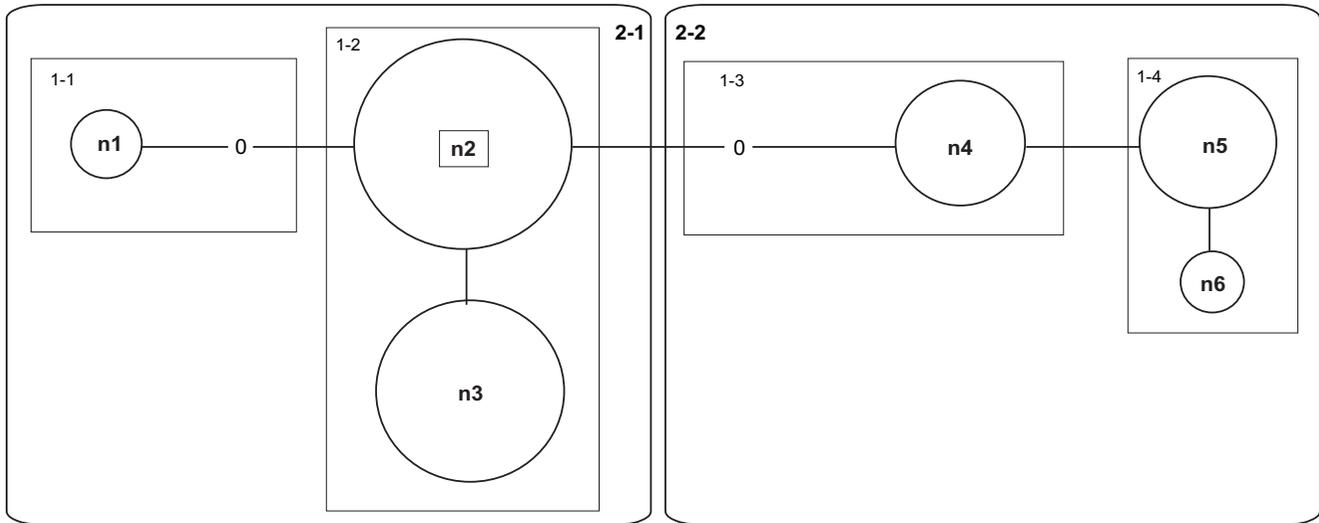


**Figure 1**  
Geographical location of *Trochoidea geyeri* sampling sites. For abbreviations see Table 1.

# mt 16S rDNA



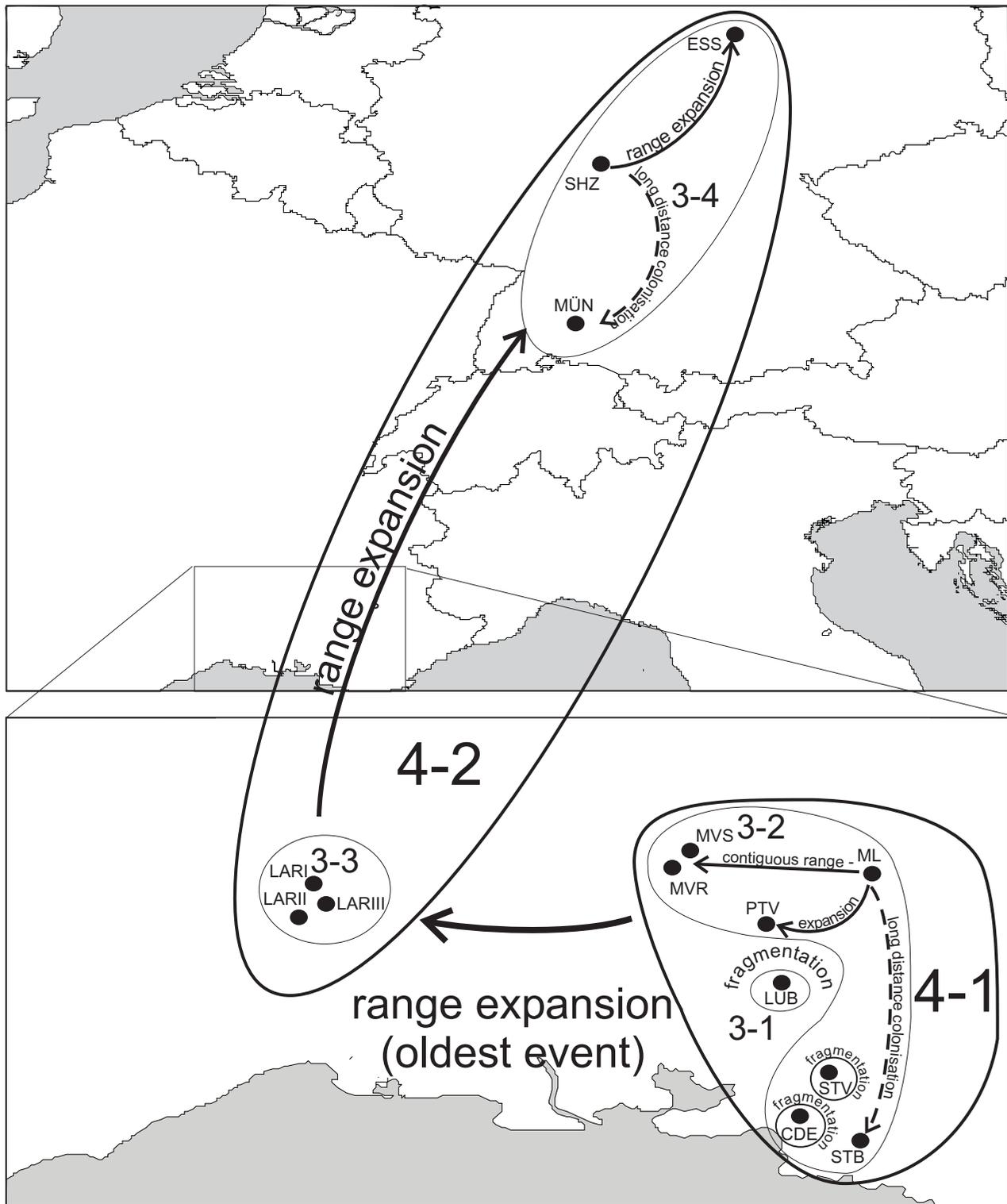
# nuc ITS-1



**Figure 2**  
**Statistical parsimony network and associated nested design.** Statistical parsimony network and associated nested design. Haplotypes are designated by names as 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. 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. a) Nested cladogram for 16S rDNA and b) for ITS-1 variation.

so that the criterion of parsimony could be assumed for all mutational connections in the network. Haplotype 9 had the highest relative root probability ( $p = 0.12$ ) [21], which suggests that this haplotype could be the oldest among those sampled. The SP network was almost fully resolved (Fig. 3), except for the position of haplotype 4.

The ITS-1 haplotype network was completely resolved. The longest connections were two mutational steps, which indicates that all connections were parsimonious, because in this marker connections longer than 9 steps have a probability of less than 95% of being



**Figure 3**  
**Geographical distribution of 3 and 4-step clades of 16S rDNA variation and inferred events in *Trochoidea geyeri*.** The font size of the inferences decreases with nesting level.

**Table 3:  $\chi^2$ -test of geographical association of clades and inferences of biological causes for association for the different clades for a) 16S rDNA and b) ITS-1. The inferences were obtained following the newest version of the inference key given in Templeton (1998).**

| a)                         |          |                 |  |
|----------------------------|----------|-----------------|--|
| Clade                      | $\chi^2$ | <i>p</i> -value | Inference  |
| Haplotypes nested in 1–8   | 9.000    | 0.008           | Long distance colonisation of MVR from MVS   |
| Haplotypes nested in 1–10  | 12.000   | 0.002           | Past fragmentation between PTV and CDE   |
| Clades nested in 2–3       | 68.000   | <0.001          | Long distance colonization of STB from Albion area   |
| Clades nested in 2–4       | 9.000    | 0.008           | Past fragmentation between PTV and STV   |
| Clades nested in 2–7       | 11.700   | 0.043           | Restricted gene-flow with isolation by distance between populations on the Causse de Larzac  |
| Clades nested in 2–8       | 15.000   | <0.001          | Long distance colonisation of MUN from SHZ   |
| Clades nested in 3–2       | 46.384   | <0.001          | Continuous range expansion with subsequent fragmentation in the Albion area  |
| Clades nested in 3–3       | 3.111    | 0.308           | No geographical association of clades  |
| Clades nested in 3–4       | 20.000   | <0.001          | Range expansion of southern German populations to ESS, but impossible to discriminate between contiguous range expansion and long distance dispersal |
| Clades nested in 4–1       | 42.000   | <0.001          | Past fragmentation between LUB and all other Provence populations, confirmed by longer than average mutational connection                            |
| Clades nested in 4–2       | 34.000   | <0.001          | Range expansion, but impossible to discriminate between contiguous range expansion and long distance dispersal from LAR to northern areas            |
| Clades in entire cladogram | 77.000   | <0.001          | Range expansion from Provence to Causse de Larzac  |
| b)                         |          |                 |  |
| Clade                      | $\chi^2$ | <i>p</i> -value | Inference  |
| Haplotypes nested in 1–2   | 46.000   | <0.001          | Isolation by distance among Provence populations except LUB  |
| Haplotypes nested in 1–4   | 20.000   | <0.001          | Inadequate sampling design to discriminate between fragmentation, range expansion or isolation by distance among German populations                  |
| Clades nested in 2–1       | 52.000   | <0.001          | Past fragmentation of LUB from the rest of the Provence populations  |
| Clades nested in 2–2       | 34.000   | <0.000          | Range expansion from the Causse de Larzac to Germany   |
| Total cladogram            | 86.000   | <0.001          | Contiguous range expansion from Provence to the Causse de Larzac   |

parsimoniously connected. Haplotype n2 had the highest root probability ( $p = 0.27$ ).

Two major clades were identified in both phylogenetic analyses. The first major clade comprised all haplotypes from populations in the Provence, east of the river Rhône (clade 4-1 in Fig. 2a and clade 2-1 in Fig. 2b). The second major clade included all haplotypes from the Causse de Larzac and the populations from Germany (clade 4-2 and clade 2-2 in Fig. 2). Within clade 4-1, there were two clearly separated groups: the LUB population (clade 3-1 in Fig. 3), and the rest of populations (clade 3-2). This relation is also reflected in the ITS-1 network (clades 1-1 and 1-2). Phylogenetic relationships in clade 4-2 suggest that haplotypes found in the LAR populations from the Causse de Larzac (clade 3-3 and 1-3, respectively) are ancestral to the German haplotypes (clade 3-4 and 1-4, respectively).

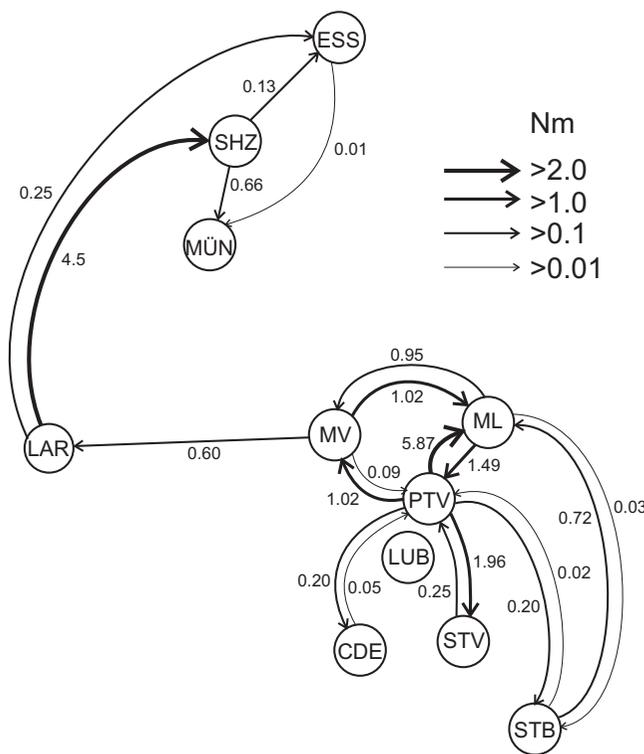
#### **Inferred population history from NCA**

Nested clade analysis showed for both loci a highly significant association between genetic and geographic distribution whenever there was geographic and genetic variation. The inferred population history is essentially

the same for both loci, except for a lower resolution of the ITS-1 locus. For both loci, the oldest inferred event (corresponding to the 4-step level in 16S and the 2-step level in ITS-1) was a range expansion from populations east of the Rhône to the Causse de Larzac. After this range expansion, a fragmentation was inferred between the population on the Luberon mountain chain (clade 3-1, 1-1, respectively) and all other populations east of the Rhône (clade 3-2, 1-2). Within the western clade (4-2, 2-2) a range expansion event to the north (clade 3-4, 1-4) was suggested. On lower (i.e., younger) clade levels, several episodes of range expansions and fragmentation events were inferred with the 16S rDNA marker, which exhibited due to its larger variability the better temporal resolution. The results are summarised in Table 3. Geographical distributions of major clades and inferences of population history derived from the 16S rDNA locus are represented in Figure 3.

#### **Past gene-flow estimates**

Even though every possible pairwise gene-flow rate was estimated, most were far below 0.001 4 Nm, indicating a high degree of population subdivision. The unilateral gene-flow from MV to LAR supports the NCA inference of



**Figure 4**  
**Past gene-flow estimates among populations inferred from 16S rDNA and ITS-1 in *Trochoidea geyeri*.** Estimates of 4 Nm  $\geq$  0.001 are indicated by arrows.

the populations in Provence being ancestral to the LAR population (Fig. 4). The ML and the PTV populations appear to be the source for most other populations in Provence, except LUB that is fragmented from other populations. In the West, the LAR region was the source to the SHZ and ESS population, which was in turn the origin of MÜN. The major events on which the argumentation is based like the northern range expansion from the LAR to SHZ, the fragmentation of the LUB population and other events inferred by the NCA are thus congruent with the gene-flow pattern suggested by *Migrate* (Fig. 4).

**Divergence times among 3-step clades**

Divergence times estimates for the 16S rDNA locus following Nei [22] for 3-step clades ranged from 180,000  $\pm$  108,000 years (mean  $\pm$  s.d.) between clades 3-3 and 3-4, to 990,000  $\pm$  90,000 years between clades 3-1 and 3-4. A population mutation parameter ( $\theta$ ) of 11 was estimated for a model with population subdivision and fluctuating population size in *Genetree*. Applying our previously estimated substitution rate of 0.056 changes per site and million years, an effective population size of 270,000 was

obtained. Using these estimates, coalescence analysis suggested an age of 232,000  $\pm$  36,600 (mean  $\pm$  95% confidence limit) for the mutation defining the split between the lineages in clade 3-3 (Causse de Larzac) and 3-4 (Germany). The average clade sequence divergence on the 3-step level of 0.039 is 5.4 fold higher than the within clade diversity of 0.007. According to the 'three times rule' [23], we can thus expect the majority of the nuclear loci having reached mutual monophyly, which is for ITS-1 indeed the case among these clades.

**Discussion**

**Combined use of mitochondrial and nuclear markers for phylogeographic inference**

A more complete understanding of processes shaping the pattern of genetic variation can be gained from the simultaneous use of independent loci. This is due to the fact that potential selection will affect only certain loci, whereas population demography and -history will leave a common imprint across all neutral loci [24]. The concordance among 16S and ITS-1 phylogenies is thus strong evidence that they reflect indeed the species history.

ITS-1 showed a lower apparent substitution rate than did 16S. This is at least partially due to the four times larger effective population size in the nuclear genome compared to mitochondrial markers. Application of the 'three-times-rule' [23] to the sequence divergence among mitochondrial 3-step clades showed, however, that a reciprocal monophyly among these entities could be expected in the nuclear marker (Table 3).

**Population history**

Nested clade analysis indicated that there was significant population structure at all clade levels (Table 3). The NCA inferences emphasise the role of historical events in shaping the distribution of haplotypes that we see today. This is evidenced by the relatively ancient fragmentation of the Luberon (LUB) population on the 3-step level from all other populations east of the Rhône. Because the Luberon lies between the Albion area and Sainte Victoire (STV), it seems likely that the haplotypes of the Luberon had their origin elsewhere and colonised the Luberon after the fragmentation of the former populations (Fig. 4).

Populations in the Albion area are found today at high altitude (above 1000 m) on Mont Ventoux (MVR, MVS) and Plateau de Vaucluse (PTV). NCA suggests that these populations expanded from the Montagne de Lure and were isolated from each other since then. This is congruent with the idea previously suggested by [15] from fossil evidence that the warming climate after one of the last glacial phases has prevented these populations to pertain at lower altitudes and therefore to exchange migrants. The mountain range Sainte Baume seems to have been colo-

nised in a distinct event from the Albion area, despite being closer to both Sainte Victoire and Luberon populations (Fig. 4).

There is little indication for current gene flow between populations of *T. geyeri*. The only exceptions from this general picture are the LAR populations of clade 3-3 on the Causse de Larzac. This area is a large carstic highland plateau with an unfragmented habitat for *T. geyeri*, where large herds of sheep may have served as vectors for passive transportation [25]. It is therefore plausible that some (restricted) gene flow occurs among these populations. The German populations seem to have originated from a northern range expansion from the populations on the Causse de Larzac. (Fig. 3, 4). All haplotypes currently found in the northern species range form a monophyletic clade, suggesting that German populations originated from a single expansion event. It is possible that this range expansion went over an intermediate, now extinct, population in the north of France. The current distribution of the German populations (clade 3-4) seem to have originated from SHZ in the centre of their range, following a northern expansion founding the ESS population and a younger long distance colonization to the south that originated the MUN population.

#### **Northern cryptic refugia?**

To speculate about the potential causes for the inferred population events, we need at least a rough idea about when they occurred. The reciprocal monophyly of 4- and 3-step clades (Figs. 2 & 3), and the appreciable sequence divergence (3.9%) between the 3-step clades, suggests that the events causing the geographical fragmentations of these clades are relatively old. The time estimates obtained suggest that the splits between 3-step clades were most likely driven by climatic changes associated with several different glacial cycles of the Pleistocene, ranging back almost one million years.

The age of clade 3-4 was of special interest, because it could indicate whether *T. geyeri* survived pleniglacial phases in periglacial refuges in the north, as suggested by the fossil record [15]. We applied two different approaches to estimate divergence times for the German populations. The method of Nei [22] takes in account ancient polymorphisms in order to distinguish between the divergence of genes and the divergence of populations, whereas the coalescent approach uses the temporal information contained in the gene tree [26]. Results from both methods remarkably agreed; the point estimate for gene divergence from coalescence analysis (232,000 years BP) preceded the point estimate for the population divergence by approximately 50,000 years, as we expect if we consider that gene divergence happened before the population split [26]. The age of the German

split suggests that snails arrived to the north during the Riss-Wuermian interglacial, a view supported by the fossil record from Burgundy [14], and that they survived *in situ* posterior glaciations in local periglacial refugia with suitable microclimate. We based this interpretation upon three arguments:

1) Divergence time estimates: Indeed, the accuracy of the divergence times estimates depends on the reliability of the substitution rate assumed. The estimated rate of 5.6% sequence divergence per million years is fast compared to the widely used of 2% for invertebrate mitochondrial sequences (e.g. [27,28]). If we were estimating the substitution rate in this gene, this only means that the inferred gene and population divergence events date even further back in time. On the other hand, only a substitution rate of about 80% or an extreme rate heterogeneity among sites would give a population divergence estimate that would fall, with 95% confidence, into the Holocene. The accuracy of the presented molecular clock estimate is thus not critical to arrive to the conclusion of a pre-Holocene split between clade 3-3 and 3-4.

2) Extinct source populations: On the other hand, an alternative interpretation of the data would be that the population split may have occurred somewhere else and that the colonisation of the German area was accomplished at the beginning of the Holocene by ancient haplotypes from a now extinct population. Potential candidate areas, with the necessary calcareous underground, could be the Northeast of France and the Rhône valley. The former area had a pleniglacial vegetation comparable to the potential refugial areas in Germany [16]. Our conclusions about the existence of glacial refuges in the North would thus not change, but shift to the West (from Germany to Northeast France or Central Belgium). The lower Rhône valley was isolated from Germany by the Rhône glacier at the beginning of the Holocene [29], thus blocking a recolonisation from the South perhaps until the reforestation of the landscape extinguished the habitat of *T. geyeri*. The inferred gene flow pattern supports this view. Gene flow can be traced only from the SHZ population to the MUN population, but not *vice versa*, as it should be expected if a recolonisation occurred from the South (Fig. 5).

3) Distribution of northern genetic variation: In addition, if the northern colonisation originated from the south in France in recent times, we would expect to find shared haplotypes among the northern populations, as it is the case for another Helicellinae species, *Candidula unifasciata* [10] and for the marine gastropod *Acanthinucella spirata* [30]. The fact that the German populations are monophyletic with respect to both nuclear and mitochondrial loci, suggests that German populations have completed

the process of lineage sorting, which in turn suggests that their origin is quite old. Especially the nuclear monophyly provides strong evidence for an ancient split. It seems unlikely that the repeated Pleistocene climatic fluctuations would have allowed the accumulation of the observed divergence in a scenario of colonisations from the same source population.

The microspatial structuring within snail populations as *i.e.* shown for the SHZ population [6], is likely to maintain genetic variation longer than expected from population size alone, as pointed out by Thomaz *et al.* [12] and Ross [31]. The reciprocal monophyly of populations within clade 3-3 is therefore indicative of a relatively ancient isolation, most likely in the region where they are found today. The alternative explanation, postglacial leptokurtic dispersal from now extinct southern refugia, appears therefore as an unnecessary *ad hoc* explanation.

As far as it can be traced back into the Pleistocene, the population history of *T. geyeri*, can be described as a repeated suite of range expansions out of local refuges and subsequent fragmentations. It seems that *T. geyeri* was capable of using the spread of favourable habitat during the transitional climate phases for recurrent range expansions. The range expansions, even on lower clade levels, involved large distances compared to the dispersal capacity of the snail, so that it is unlikely that the colonisations were achieved by active dispersal. We can only speculate about the vectors for a passive transportation, but the rich Pleistocene fauna of large mammals in this region [32] is a likely candidate for such a task [25], even though other animals like birds or even wind have been implicated in dispersal of snails [33,34].

## Conclusions

*T. geyeri* seems to have survived in local refugia the reduction of the favourable steppe-like habitat due to climatic extremes during the pleniglacial and interstadial periods, as it is the case today. Pfenninger & Bahl [35] suggested that snail species with restricted dispersal might survive in habitats of a size in the magnitude of few square meters. There is increasing evidence that such small spots with a favourable microclimate existed in the periglacial area of central Europe [19] and were presumed to have provided refuges for comparatively cold resistant snail species [7]. The present study fits thus well into the increasing evidence that the well-studied southern and eastern European refugia were supplemented by cryptic sanctuaries in northern Europe during the late Pleistocene in shaping present day species composition.

## Methods

### Sampling collection and DNA extraction

Fourteen sites were sampled in Germany and France, comprising all known extant populations of *T. geyeri* (Fig. 1, Table 1). Eighty-eight 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 & Pfenninger [36]. 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 [37].

### Amplification of 16S rDNA, sequencing and alignment

The 16S target-DNA from 79 individuals was amplified by PCR with standard universal primers of 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 50 s, 44°C for 50 s and 72°C for 40 s) and 36 cycles (92°C for 30 s, 48°C for 40 s and 72°C for 40 s.) after initial incubation of 90°C for 2 min 30 s. The nuclear ITS1 locus was amplified with primers obtained from Armbruster *et al.* [38]. Amplification was performed with 40 cycles (92°C for 1 min, 48°C for 1 min, and 72°C for 1 min 30 s). 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 were deposited in GenBank (Accession nos. XXXX-XXXX). Sequences were aligned with the Clustal option [39] in the computer program SequenceNavigator (Perkin Elmer, Applied Biosystems) and manually adjusted.

### Nested clade analyses

The phylogeny of the mitochondrial and nuclear haplotypes was inferred using statistical parsimony (SP; [40]). Indels were regarded as a fifth state. The SP networks were constructed using TCS v1.06 [41].

We used nested clade analysis (NCA; [42]) to infer the population history of *T. geyeri*. The NCA is an objective statistical analysis that first tries to reject the null hypothesis of no association between haplotype variation and geography, and then interprets the significant association 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, genetic drift, system of mating) processes responsible for the observed pattern of genetic variation. In addition, the NCA allows for the inference of separate

**Table 4: Nucleotide divergence and clade divergence times estimates Lower half: nucleotide divergence (mean and s.d) among 3-step clades, calculated as changes per site, diagonal: nucleotide diversity (mean  $\pm$  s.d.) within 3-step clades and upper half: estimated divergence time in million years before present between clades (mean  $\pm$  s.d.).**

|     | 3-1                 | 3-2                 | 3-3                 | 3-4                 |
|-----|---------------------|---------------------|---------------------|---------------------|
| 3-1 | 0.004 $\pm$ (0.004) | 0.475 $\pm$ (0.119) | 0.823 $\pm$ (0.084) | 0.990 $\pm$ (0.090) |
| 3-2 | 0.034 $\pm$ (0.007) | 0.011 $\pm$ (0.007) | 0.383 $\pm$ (0.121) | 0.587 $\pm$ (0.123) |
| 3-3 | 0.051 $\pm$ (0.005) | 0.030 $\pm$ (0.007) | 0.006 $\pm$ (0.005) | 0.180 $\pm$ (0.108) |
| 3-4 | 0.061 $\pm$ (0.005) | 0.042 $\pm$ (0.007) | 0.016 $\pm$ (0.006) | 0.007 $\pm$ (0.006) |

evolutionary events in space and time. The NCA has been successfully applied to diverse sets of data (e.g. [10,43,44]).

The NCA nesting design was constructed by hand upon the SP network following the rules given in Templeton [44] and Crandall [45]. The program GeoDis 2.0 [46] was used to calculate the various NCA distance measures and their statistical significance. The statistics calculated for all clades were i) the nested clade distance ( $D_C$ ), that measures the average distance of all clade members from the geographical centre of distribution, ii) the nested clade distance ( $D_N$ ), that measures how widespread a particular clade is relative to the distribution of its sister clades in the same nesting group, and iii) the interior-tip distances ( $I-T_C$  and  $I-T_N$ ), that indicate how widespread evolutionary younger clades (tip clades) are relative to their ancestors clades (interior clades). The statistical significance of the different NCA distance measures was calculated by comparison with a null distribution (i.e., no association between genetic variation and geographical distribution) constructed from 10000 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 [44].

#### Migration rates

While the NCA may be very useful to infer different historical processes, it does not allow for the estimation of standard population genetic parameters. The coalescence approach [47] can be used to obtain maximum likelihood estimates of various population parameters. In particular, estimates of gene flow among populations can be calculated taking population structure and demography into account. We used both loci combined in a single coalescent analysis to estimate migration patterns with Migrate 1.2.4 [48]. Even though pooling of populations violates certain assumptions of Migrate, it can be a reasonable solution to keep computation feasible [10]. All LAR populations and the populations MVR and MVS were pooled because of their geographical proximity. The Migrate approach to estimate gene-flow rates has advantages over equilibrium approaches, because it takes history and

asymmetrical gene flow into account [49]. To obtain past gene-flow estimates for a full migration model among all populations, we used 30 short chains with 500 steps and 100,000 sampled genealogies, and 25 long chains with 5000 steps and 1,000,000 sampled genealogies.

#### Divergence times

The substitution rate for the 16S gene was estimated from the estimated ML tree for *T. geyeri*, and three other species from the same subfamily: *Cernuella cespitum*, *Candidula unifasciata* and *C. rugosiuscula* (Sequences from Steinke & Pfenninger, unpublished). Homogeneity of substitution rates among lineages was not rejected for this four species (LRT;  $P > 0.05$ ; [50]), and we calibrated the molecular clock using a divergence time of approximately 3.2 ( $\pm 0.5$ ) million years for the split between *Candidula unifasciata* and *Candidula rugosiuscula*, obtained from fossil evidence (unpublished). The estimated substitution rate was 0.056 ( $\pm 0.011$ ) changes per site per million years.

Average sequence divergence between inferred 3-step clades in the nested cladogram was used to estimate divergence times. Patristic pairwise distances among haplotypes were inferred from the statistical parsimony network. The divergence time  $t$  between clades was computed as  $t = (d_{xy} - 0.5 * (d_x + d_y)) * \text{substitution rate}$ , where  $d_x$  and  $d_y$  denote as the average sequence diversity within and  $d_{xy}$  as sequence divergence among clades, respectively [22].

We also explore a second approach for the estimation of divergence times. A coalescent-likelihood framework was used to obtain estimates of gene divergence times for the mutations defining the haplotypes of populations SHZ, ESS and MUN. The mean coalescent times and standard deviations of all mutation events defining a haplotype tree were estimated under different models of demographic history and population structure, using an advanced simulation approach [51] implemented in the computer program *Genetree*, version 8.3 (R. C. Griffiths). Analysis of mutational ages was restricted to the previously inferred clade 4-2, because mutations in this clade conform to the infinite sites model, which is a condition

required for the analysis, and to keep computing time reasonable. Moreover, all pairwise connections in the cladogram were parsimonious at the 95% confidence level, which assures lack of homoplasy in the data set. Mutational ages were estimated for a demographic model with population subdivision under fluctuating population size. The migration matrix for clade 4-2 obtained with *Migrate* was used to specify the migration rates for *Genetree*. Calculations were based on 10,000,000 simulations. To convert coalescence time (T) to real time (t), the relationship  $t = 2T N_e G_t$  was used, where  $G_t$  stands for the generation time in years and  $N_e$  for the effective population size [51]. A minimum generation time of one year was used [6]. The parameter  $N_e$  was estimated from the relationship  $N_e = \theta / 2 \mu$ . The population mutation parameter  $\theta$  was initially estimated under a fluctuating population size model using the program *Fluctuate* [52]. We have then performed a maximum likelihood maximisation of this value in *Genetree*, given the migration matrix estimated in *Migrate*.

### Authors' contributions

Markus Pfenninger and Frédéric Magnin designed the study and sampled the material together. M.P. carried out the molecular work and analysed the sequences together with David Posada. All three authors contributed to the preparation of the manuscript. They read and approved the final version.

### References

- Gates DM: **Climate change and its biological consequences** *Sunderland: Sinauer Ass* 1993.
- Godan D: **Schadschnecken und ihre Bekämpfung** *Stuttgart: Ulmer* 1979.
- Kerney MP, Cameron RAD and Jungbluth JH: **Die Landschnecken Nord- und Mitteleuropas** *Hamburg: Berlin: Paul Parey* 1983.
- Cowie RH: **Density, Dispersal and Neighborhood Size In the Land Snail *Theba pisana*** *Heredity* 1984, **52**:391-401.
- Baur A and Baur B: **Daily movement patterns and dispersal in the land snail *Arianta arbustorum*** *Malacologia* 1993, **35**:89-98.
- Pfenninger M, Bahl A and Streit B: **Isolation by distance in a population of a small land snail *Trochoidea geyeri*: Evidence from direct and indirect methods** *Proceedings of the Royal Society of London Series B-Biological Sciences* 1996, **263**:1211-1217.
- Ant H: **Die Bedeutung der Eiszeiten für die rezente Verbreitung der europäischen Landgastropoden** *Malacologia* 1966, **5**:61-62.
- Goodfriend GA: **The Use Of Land Snail Shells In Paleoenvironmental Reconstruction** *Quaternary Science Reviews* 1992, **11**:665-685.
- Cruzan MB and Templeton AR: **Paleoecology and coalescence: phylogeographic analysis of hypotheses from the fossil record** *Trends in Ecology & Evolution* 2000, **15**:491-496.
- Pfenninger M and Posada D: **Phylogeographic history of the land snail *Candidula unifasciata* (Poirlet 1801) (Helicellinae, Stylommatophora): fragmentation, corridor migration and secondary contact** *Evolution* 2002, **56**:1776-1788.
- Guiller A, Madec L and Daguzan J: **Geographical Patterns of Genetic Differentiation in the Landsnail *Helix aspersa* Müller (Gastropoda, Pulmonata)** *Journal of Molluscan Studies* 1994, **60**:205-221.
- Thomaz D, Guiller A and Clarke B: **Extreme divergence of mitochondrial DNA within species of pulmonate land snails** *Proceedings of the Royal Society of London Series B-Biological Sciences* 1996, **263**:363-368.
- Guiller A, Coutellec-Vreto MA, Madec L and Deunff J: **Evolutionary history of the land snail *Helix aspersa* in the Western Mediterranean: preliminary results inferred from mitochondrial DNA sequences** *Molecular Ecology* 2001, **10**:81-87.
- Schlickum RK and Puissegur JJ: **Die Molluskenfauna der Schichten mit *Viviparus burgund* und *Pyrgula nodotiana* von Montagny-les-Beaumes (Dép. Côte-d'Or)** *Archiv für Molluskenkunde* 1978, **109**.
- Magnin F: **Les distributions pléistocène et actuelle de *Trochoidea (Xerochloa) geyeri* (Soós 1926) dans le Sud de la France: un exemple de disjonctions d'aire liée au réchauffement post-glacière** *Bulletin de la Société géologique de la France* 1989, **8**:779-786.
- Frenzel B: **The history of flora and vegetation during the Quaternary** *Fortschritte der Botanik* 1981, **43**:1-98.
- Magnin F: **Competition between 2 Land Gastropods Along Altitudinal Gradients in South-Eastern France – Neontological and Paleontological Evidence** *Journal of Molluscan Studies* 1993, **59**:445-454.
- Hanfling B, Hellemans B, Volckaert FAM and Carvalho GR: **Late glacial history of the cold-adapted freshwater fish *Cottus gobio* as revealed by microsatellites** *Molecular Ecology* 2002, **11**:1717-1729.
- Stewart JR and Lister AM: **Cryptic northern refugia and the origins of the modern biota** *Trends in Ecology and Evolution* 2001, **16**:608-613.
- Lydeard C, Holznagel WE, Schnare MN and Gutell RR: **Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures** *Molecular Phylogenetics and Evolution* 2000, **15**:83-102.
- Castelloe J and Templeton AR: **Root probabilities for intraspecific gene trees and neutral coalescent theory** *Molecular Phylogenetics and Evolution* 1994, **3**:102-113.
- Nei M: **Molecular Evolutionary Genetics** *New York: Columbia University Press* 1987.
- Palumbi SR, Cipriano F and Hare MP: **Predicting nuclear gene coalescence from mitochondrial data: The three-times rule** *Evolution* 2001, **55**:859-868.
- Hare MP: **Prospects for nuclear gene phylogeography** *Trend Ecol Evol* 2001, **16**:700-706.
- Dörge N, Walther C, Beinlich B and Plachter H: **The significance of passive transport for dispersal in terrestrial snails (Gastropoda, Pulmonata)** *Zeitschrift für Ökologie und Naturschutz* 1999, **8**:1-10.
- Edwards SV and Beerli P: **Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies** *Evolution* 2000, **54**:1839-1854.
- Gomez A, Carvalho GR and Lunt DH: **Phylogeography and regional endemism of a passively dispersing zooplankton: mitochondrial DNA variation in rotifer resting egg banks** *Proceedings of the Royal Society of London Series B-Biological Sciences* 2000, **267**:2189-2197.
- Masta SE: **Phylogeography of the jumping spider *Habronattus pugillis* (Araneae: Salticidae): Recent vicariance of sky island populations?** *Evolution* 2000, **54**:1699-1711.
- Wohlfarth B, Gaillard MJ, Haeberli W and Kelts K: **Environment and Climate in Southwestern Switzerland During the Last Termination, 15-10-Ka-Bp** *Quaternary Science Reviews* 1994, **13**:361-394.
- Hellberg ME, Balch DP and Roy K: **Climate-Driven Range Expansion and Morphologic Evolution in a Marine Gastropod** *Science* 2001, **292**:1707-1710.
- Ross TK: **Phylogeography and conservation genetics of the Iowa Pleistocene snail** *Molecular Ecology* 1999, **8**:1363-1373.
- Brugal JP and Crégut-Bonnouere E: **Le Paléolithique moyen en Vaucluse** *Avignon* 1994.
- Boag DA: **Dispersal in pond snails: potential role of waterfowl** *Canadian Journal of Zoology* 1986, **64**:904-909.
- Kirchner CH, Kratzner R and Welter-Schultes FW: **Flying snails – How far can *Truncatellina* (Pulmonata:Vertiginidae) be blown over the sea?** *Journal of Molluscan Studies* 1997, **63**:479-487.
- Pfenninger M and Bahl A: **Influence of habitat size on the viability of spatially structured populations of the land snail *Trochoidea geyeri*** *Verhandlungen der Gesellschaft für Ökologie* 1997, **27**:469-473.

36. Bahl A and Pfenninger M: **A rapid method of DNA isolation using laundry detergent** *Nucleic Acids Research* 1996, **24**:1587-1588.
37. Sambrook J, Fritsch EF and Maniatis T: **Molecular Cloning: A Laboratory Manual** Cold Spring Harbour: Cold Spring Harbour Laboratory 1989.
38. Armbruster GFJ, van Moorsel CHM and Gittenberger E: **Conserved sequence patterns in the non-coding ribosomal ITS-1 of distantly related snail taxa** *Journal of Molluscan Studies* 2000, **66**:570-573.
39. Thompson JD, Higgins DG and Gibson TJ: **Clustal-W – Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice** *Nucleic Acids Research* 1994, **22**:4673-4680.
40. Templeton AR, Crandall KA and Sing CF: **A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation** *Genetics* 1992, **132**:619-633.
41. Clement M, Posada D and Crandall KA: **TCS: a computer program to estimate gene genealogies** *Molecular Ecology* 2000, **9**:1657-1659.
42. Templeton AR, Routman E and Phillips CA: **Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander, *Ambystoma tigrinum*** *Genetics* 1995, **140**:767-782.
43. Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR and Zegura SL: **Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation** *Molecular Biology and Evolution* 1998, **15**:427-441.
44. Templeton AR: **Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history** *Molecular Ecology* 1998, **7**:381-397.
45. Crandall KA: **Multiple interspecies transmissions of human and simian T-cell leukemia/lymphoma virus type I sequences** *Molecular Biology and Evolution* 1996, **13**:115-131.
46. Posada D, Crandall KA and Templeton AR: **GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes** *Molecular Ecology* 2000, **9**:487-488.
47. Kingman JFC: **The coalescent** *Stochastic Proceedings and Applications* 1982, **13**:235-248.
48. Beerli P and Felsenstein J: **Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach** *Genetics* 1999, **152**:763-773.
49. Beerli P and Felsenstein J: **Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:4563-4568.
50. Felsenstein J: **Evolutionary Trees from DNA sequences: a maximum likelihood approach** *J Mol Evol* 1981, **17**:368-376.
51. Griffiths RC and Tavaré S: **Ancestral Inference in Population-Genetics** *Statistical Science* 1994, **9**:307-319.
52. Kuhner MK, Yamato J and Felsenstein J: **Maximum likelihood estimation of population growth rates based on the coalescent** *Genetics* 1998, **149**:429-434.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

