Evidence for survival of Pleistocene climatic changes in Northern refugia by the land snail Trochoidea geyeri

David Posada
Markus Pfenninger
Frederic Magnin

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Biology Commons

Original Publication Citation
BMC Evolutionary Biology, Vol. 3, No. 1. (23)

BYU ScholarsArchive Citation
Posada, David; Pfenninger, Markus; and Magnin, Frederic, "Evidence for survival of Pleistocene climatic changes in Northern refugia by the land snail Trochoidea geyeri" (2003). Faculty Publications. 501. https://scholarsarchive.byu.edu/facpub/501

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
Evidence for survival of Pleistocene climatic changes in Northern refugia by the land snail Trochoidea geyeri (Soós 1926) (Helicellinae, Stylommatophora)
Markus Pfenninger*1,3, David Posada2 and Frédéric Magnin1

Address: 1IMEP URA-CNRS 1152 Case 451, Faculté des Sciences et Techniques de St. Jerome, F-13997 Marseille Cedex 20, France, 2Department of Zoology, Brigham Young University, 574 WIDB, Provo, UT 84602-5255, USA and 3Abteilung 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

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 shifts 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 mountain tops, arctic 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 palaeoclimatic 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 clad 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 (α) estimates were 2.2597 and 0.0079. Such a small value of α 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% ± 1.52 (s.d.), while the mean within population diversity was 0.28% ± 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.

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

Table 2: a) Distribution of *T. geyeri* 16S rDNA and b) ITS-1 haplotypes (columns) at each sampled location (rows).

### a)

| Location | 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 | 2 | 1 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| ML       |   |   |   |   |   | 6 | 1 |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| STV      |   |   |   |   |   |   | 2 | 3 | 2 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| CDE      |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    | 9 |    |    |    |    |    |    |    |    |    |    |    |    |    |
| STB      |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    | 3 | 2 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| LAR1     |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 3 | 3 | 1 | 1 | 1 | 1 |
| LAR2     |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| LAR3     |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| SHZ      |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 2 | 1 |
| MUN      |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 6 | 1 | 8 |
| ESS      |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    | 5 |

### b)

<table>
<thead>
<tr>
<th>Location</th>
<th>n1</th>
<th>n2</th>
<th>n3</th>
<th>n4</th>
<th>n5</th>
<th>n6</th>
</tr>
</thead>
<tbody>
<tr>
<td>LUB</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MVS</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MVR</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTV</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ML</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STV</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDE</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STB</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAR1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAR2</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAR3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SHZ</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUN</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESS</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1
Geographical location of *Trochoidea geyeri* sampling sites. For abbreviations see Table 1.
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.
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...
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 MUN. 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).

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

**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 karstic 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 monophony 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

---

**Page 9 of 13**

(page number not for citation purposes)
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 periglacial 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 CTG C < 3’ and 16S2 5’ > GTC CGG TTI 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
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 (± 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 (± 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 \( t = (d_{xy} - 0.5 * (d_x + d_y)) \) 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...

<table>
<thead>
<tr>
<th></th>
<th>3-1</th>
<th>3-2</th>
<th>3-3</th>
<th>3-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>0.004 ± (0.004)</td>
<td>0.475 ± (0.119)</td>
<td>0.823 ± (0.084)</td>
<td>0.990 ± (0.090)</td>
</tr>
<tr>
<td>3-2</td>
<td>0.034 ± (0.007)</td>
<td>0.011 ± (0.007)</td>
<td>0.383 ± (0.121)</td>
<td>0.587 ± (0.123)</td>
</tr>
<tr>
<td>3-3</td>
<td>0.051 ± (0.005)</td>
<td>0.030 ± (0.007)</td>
<td>0.006 ± (0.005)</td>
<td>0.180 ± (0.108)</td>
</tr>
<tr>
<td>3-4</td>
<td>0.061 ± (0.005)</td>
<td>0.042 ± (0.007)</td>
<td>0.016 ± (0.006)</td>
<td>0.007 ± (0.006)</td>
</tr>
</tbody>
</table>

Table 4: Nucleotide divergence and clade divergence times estimates. Lower half: nucleotide divergence (mean ± s.d) among 3-step clades, calculated as changes per site, diagonal: nucleotide diversity (mean ± s.d.) within 3-step clades and upper half: estimated divergence time in million years before present between clades (mean ± s.d.).
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 fluctuating population size model using the program Fluctuate [52]. We have then performed a maximum likelihood maximisation of this value in Genetree, giving 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


