Delimitation of species is often complicated by discordance of morphological and genetic data. This may be caused by the existence of cryptic or polymorphic species. The latter case is particularly true for certain snail species showing an exceptionally high intraspecific genetic diversity. The present investigation deals with the *Trochulus hispidus* complex, which has a complicated taxonomy. Our analyses of the COI sequence revealed that individuals showing a *T. hispidus* phenotype are distributed in nine highly differentiated mitochondrial clades (showing p-distances up to 19%). The results of a parallel morphometric investigation did not reveal any differentiation between these clades, although the overall variability is quite high. The phylogenetic analyses based on 12S, 16S and COI sequences show that the *T. hispidus* complex is paraphyletic with respect to several other morphologically well-defined *Trochulus* species (*T. clandestinus*, *T. villosus*, *T. villosulus* and *T. striolatus*) which form well-supported monophyletic groups. The nc marker sequence (5.8S–ITS2–28S) shows only a clear separation of *T. o. oreinos* and *T. o. scheerpeltzi*, and a weakly supported separation of *T. dandestinus*, whereas all other species and the clades of the *T. hispidus* complex appear within one homogeneous group. The paraphyly of the *T. hispidus* complex reflects its complicated history, which was probably driven by geographic isolation in different glacial refugia and budding speciation. At our present state of knowledge, it cannot be excluded that several cryptic species are embedded within the *T. hispidus* complex. However, the lack of morphological differentiation of the *T. hispidus* mitochondrial clades does not provide any hints in this direction. Thus, we currently do not recommend any taxonomic changes. The results of the current investigation exemplify the limitations of barcoding attempts in highly diverse species such as *T. hispidus*.

Corresponding author: Luise Kruckenhauser, Central Research Laboratories, Natural History Museum Vienna, Burggring 7, 1010 Vienna, Austria. E-mail: Luise.Kruckenhauser@nhm-wien.ac.at

Luise Kruckenhauser, Natural History Museum Vienna, Central Research Laboratories, Burggring 7, Vienna, 1010, Austria. E-mail: Luise.Kruckenhauser@nhm-wien.ac.at

Michael Duda, 3rd Zoological Department, Natural History Museum Vienna, Burggring 7, 1010, Vienna, Austria. E-mail: Michael.Duda@nhm-wien.ac.at

Daniela Bartel, University of Vienna, Department of Integrative Zoology, Althanstraße 14, 1090, Vienna, Austria. E-mail: Daniela.Bartel@univie.ac.at

Helmut Sattmann, 3rd Zoological Department, Natural History Museum Vienna, Burggring 7, 1010, Vienna, Austria. E-mail: Helmut.Sattmann@nhm-wien.ac.at

Josef Harl, Sandra Kirchner, and Elisabeth Haring, Natural History Museum Vienna, Central Research Laboratories, Burggring 7, Vienna, 1010, Austria and University of Vienna, Department of Integrative Zoology, Althanstraße 14, 1090, Vienna, Austria. E-mails: Josef.Harl@nhm-wien.ac.at, Sandra.Kirchner@nhm-wien.ac.at, Elisabeth.Haring@nhm-wien.ac.at

© 2014 The Authors Zoologica Scripta published by John Wiley & Sons Ltd on behalf of Royal Swedish Academy of Sciences

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
Introduction

Recognising species as biological entities is, besides the human interest in describing and categorising nature, the basis for many biological investigations. Yet for many species groups, this task is not at all trivial. Cryptic as well as highly polymorphic species may hamper unambiguous species assignment, leaving biologists unsatisfied. This is especially valid for poorly studied species. Molecular genetic approaches brought invaluable progress in understanding the history of populations and hence of speciation. Sometimes, however, the results of molecular genetic analyses reveal inconsistencies between morphological differentiation and phylogenetic relationships based on DNA (mostly mitochondrial) sequences. A case in point is certain snail species exhibiting an exceptionally high intraspecific genetic diversity in mitochondrial (mt) sequences. (Hayashi & Chiba 2000; Haase et al. 2003; Van Riel et al. 2005; Dillon & Robinson 2008; Davison et al. 2009; Scheel & Hausdorf 2012).

The genus Trochulus Chemnitz, 1786 represents an example of complicated taxonomy, species differentiation and delimitation. The most common species within this genus, *T. bispidus* (Linnaeus, 1758), is widely distributed in Europe. It occurs over a broad range of altitudes, up to 2300 m above sea level (asl) and habitats. The range covers large parts of Europe from Ireland and France in the west to western Russia (southern Urals) in the east (Loźek 1956; Kerney et al. 1983; Sysoev & Schileyko 2009). In the north, it reaches the Arctic Circle, in the south, the northern part of the Iberian Peninsula, the Mediterranean and the Balkan Peninsula. *T. bispidus* is a morphologically polymorphic species and, as a consequence, its systematics has long been a matter of controversy (Forcart 1965; Procków 2009; Welter-Schultes 2012; Procków et al. 2013), for example, concerning morphologically similar species, as *Trochulus plebeius* (Draparnaud, 1805), *Trochulus sericus* (Draparnaud, 1801) and *Trochulus coelambola* (Loccard, 1888) (e.g. Falkner 1982; Procków 2009).

One in the meanwhile clarified example, *T. oreinos* (Wagner 1915) was originally regarded as a regional subspecies of *T. bispidus*, but was later considered as a separate species (Falkner 1982). In our previous molecular study (Duda et al. 2011) on *T. oreinos*, we were able to confirm the latter assumption: The two taxa are morphologically and genetically clearly separated, and the high sequence divergence indicates that they split a long time ago. The genetic investigation, which comprised *T. bispidus* specimens collected in the areas surrounding the distribution range of *T. oreinos*, revealed monophyly for both *T. bispidus* and *T. oreinos* (Duda et al. 2011). However, an earlier analysis of *T. bispidus* using samples from Germany, France and Switzerland indicated high genetic variability with various, distinct mitochondrial lineages (Pfenninger et al. 2005). This raised the question whether there are additional *T. bispidus* lineages in the Eastern Alps. The taxonomy of *T. bispidus* is further complicated by *Trochulus sericus*: this species is characterised by a more globular shell and a narrow umbilicus, but turned out to be not clearly differentiated from *T. bispidus* (Pfenninger et al. 2005) on the basis of mtDNA sequence data. Therefore, the various lineages and morphotypes have been summarised under the designation *Trochulus bispidus/sericus* complex (Pfenninger et al. 2005). Further investigations of differentiated mtDNA lineages, nuclear (nc) markers and morphology indicated the presence of cryptic species within this complex (Pfenninger et al. 2005; Depraz et al. 2009). The results of Pfenninger et al. (2005) indicate that detailed geographic sampling is crucial for a meaningful interpretation of the phylogeography of this complex. In this context, note that the Austrian *T. bispidus* clade detected in our previous study (Duda et al. 2011) was highly differentiated from those described by Pfenninger et al. (2005). This finding suggests that the Alpine region might have played an important role for the diversification of the lineages of this complex. This calls for gathering more data from the Eastern Alps and surrounding lowland areas in eastern Austria. These regions are known to have served as glacial refugia for several invertebrates and vascular plants and as a consequence still display high levels of endemism (Trîbsch & Schönswetter 2003; Schönswetter et al. 2005; Rabitsch et al. 2009). In the current study, we performed for the first time an exhaustive analysis of the variability of *T. bispidus* and *T. sericus* in Austria using mitochondrial and nuclear DNA marker sequences. In addition, we analysed several samples from surrounding countries (Hungary, Italy, Slovenia, Switzerland and southern Germany) as well as from Sweden the type locality of *T. bispidus*, and the Netherlands. We also included several other Central European species of the genus *Trochulus* into the analysis to clarify their phylogenetic relationships and to compare the amount of intraspecific variation. The main objective of the study was to gain more insights into the complicated evolutionary history of *T. bispidus*. By assessing mtDNA variation, we wanted to obtain a clearer picture about the geographic distribution of haplotypes. This should, together with conchological and anatomical analyses, provide a basis to attempt a delineation of *T. bispidus*. Specifically, we addressed the following questions: (i) Are there additional clades with *T. bispidus* or *T. sericus* phenotype besides those reported in earlier studies (Pfenninger et al. 2005; Depraz et al. 2009) and are (some of) the latter distributed also in the Alpine region? (ii) What is the
distribution of clades? (iii) Are there regions where clades co-occur? (iv) Is there a pattern in the ncDNA analysis that reflects the results of the mtDNA? and (v) Do the data suggest the presence of cryptic species?

In a parallel approach, the same individuals as in this study were investigated morphologically to determine whether the differentiation found in the mtDNA sequences is accompanied by shell morphological diversification. These data are presented elsewhere (Duda et al. revised), but will be discussed in the context of the genetic results.

Material and methods

Study area and sampling

We collected living specimens of the genus Trochulus from 126 sampling sites (Fig. 1). Exact positions and elevations of collection sites were determined using GPS. Additional samples from Switzerland, Germany and Sweden were kindly provided by colleagues (Ulrich Schneppat, Ira Richling, Ted von Proschwitz and Zoltan Feher). Snails with a T. hispidus and T. sericeus phenotype, Trochulus bionicus (Eder, 1917), Trochulus villosus (Draparnaud, 1805), Trochulus coelomphala, Trochulus clandestinus (Hartmann, 1821),

Fig. 1 Sampling sites and distribution of clades. Circles indicate regions in which the sampling sites were too dense to be depicted with their actual distances and therefore had been manually decompressed. Triangles indicate sampling sites at which several clades co-occur. The dark blue lines indicate the maximum extent of glaciers (35–19 ka ago) during the Würm ice age. Abbreviations: Tr_bic: Trochulus bionicus; Tr_cla: Trochulus clandestinus; Tr_coe: Trochulus coelomphala; Tr_ore: T. o. oreinos; Tr_sch: T. o. scheerpeltzi; Tr_str: Trochulus striolatus; Tr_vil: Trochulus villosus; Tr_vus: Trochulus villosus.
\textit{Trochulus villosus} (Rossmässler 1838) and \textit{Trochulus striolatus} (Pfeiffer, 1828) were determined using morphological traits described in the literature (e.g. Ložek 1956; Kerney et al. 1983; Procków 2009). \textit{Trochulus oreinos oreinos} and \textit{Trochulus oreinos scheerpeletzi} were assigned using the original description (Wagner 1915; Mikula 1957) as well as by comparisons with reference specimens (paratypes) from the collections of the Natural History Museum Vienna (NHMW). Information on individuals and localities is compiled in Table S1.

Living specimens were drowned in the laboratory, and DNA was extracted according to the protocol of Krucken- hauer et al. (2011). If available, DNA was extracted from three adult individuals of each locality. In total, 389 individuals were used for this analysis, among them three \(T. \textit{bicornicus}\), eight \(T. \textit{villosus}\), four \(T. \textit{coelomphala}\), three \(T. \textit{clandestinus}\), six \(T. \textit{villosus}\), 39 \(T. \textit{striolatus}\), 32 \(T. \textit{o. oreinos}\) and 28 \(T. \textit{o. scheerpeletzi}\). Twenty-six individuals of \(T. \textit{o. oreinos}\) and 27 of \(T. \textit{o. scheerpeletzi}\) have also been used in earlier studies (Duda et al. 2010, 2011). The remaining 261 individuals were morphologically assigned to \textit{Trochulus bispinus} or tentatively to \textit{Trochulus sericus}; 79 of them were already included in earlier studies of Duda et al. (2010, 2011). Six individuals representing out-group taxa were used: three individuals of \textit{Iognomostoma iognomostomos} (Schröter, 1784), one of \textit{Monacha cantiana} (Montagu, 1803) and two of \textit{Plicuteria labomirski} (Ślòsarski, 1881).

\section*{Genetic analysis}

From all individuals, a partial region of the mt \textit{cytochrome c oxidase subunit I} (\textit{COI}) gene was analysed. In addition, from representatives of each clade, partial regions of the mt 16S rRNA (16S) and the 12S rRNA (12S) genes were also sequenced (98 individuals). As out-group taxa, \textit{Monacha cantiana} (one specimen) and \textit{Plicuteria labomirski} (two specimens) were analysed with all three markers. Primer binding sites correspond to those used by Gittenberger et al. (2004) for \textit{COI} and by Pfenninger et al. (2003) for 16S. Primers were optimised based on the alignments of several snail species and published in Duda et al. (2011).

Primer sequences for the 12S fragment were designed by Cadávia et al. (2013): 12SGast_fwd2 5'-AGTGAAGGGGC GATTGGTGT-3', 12SGast_rev3 5'-TAAGCCTGTGGGTC ATATAAC-3'. Resulting fragment sizes (including primers) were 705 bp (COI), 391–399 bp (16S) and 689–703 bp (12S).

As a nuclear (nc) marker, we analysed a region including partial sequences of two rRNA genes and the spacer region in between (5.8S–ITS2–28S) using two overlapping fragments generated with the two primer combinations: 5.8S_LSU-1fw 5'-CTAGCCTGGCAGAATTAATGTGAG-3', 28S_LSU-2fw 5'-GGGT TTGTGGGGAATGCACG-3' and 28S_LSU-3rv 5'-ACTTTCCCTCAGGATCTTG-3', 28S_LSU-4rv 5'-GTTAGACTCCTTGGTTCCGTG-3' (all from Wade & Mordan 2000). The combined alignments resulted in a sequence of ~1360 bp. PCR was performed on a Master Gradient thermocycler (Eppendorf) in 50 µl with 1 unit Taq DNA polymerase (Roche), 1 µM of each primer and 0.2 µM of each dNTP (Boehringer Mannheim). Each PCR comprised 35 reaction cycles with the following annealing temperatures: 50°C (COI, both nuclear fragments) and 55°C (16S and 12S). Control reactions for both DNA extractions and PCR amplifications were carried out. PCR products were purified using the QIAquick PCR Puri-fication kit (Qiagen) and analysed by direct sequencing (both directions). Within the PCR fragment 5.8S_LSU-1fw/28S_LSU-3rv, six heterozygous individuals with length polymorphisms were detected, and hence, their sequence could not be determined by direct sequencing. The corresponding PCR fragments were cloned: PCR products were extracted from agarose gels using the QIAquick Gel Extraction Kit (Qiagen, Düsseldorf, Germany) and cloned (TOPO TA Cloning Kit, Invitrogen, Carlsbad, CA, USA); four clones per individual were sequenced. Sequencing of both strands was performed by LGC Genomics (Berlin, Germany) using the original PCR primers or (for cloned PCR products) Mi13 universal primers.

\section*{Data analysis}

Sequences were edited in BioEdit version 5.0.9 (Hall 1999). This software was also used to translate the DNA sequences of the \textit{COI} into amino acid sequences using the invertebrate mitochondrial code. For the \textit{COI} sequences, the alignment was straightforward because there were no insertions or deletions. Alignments of the mt sequences 16S and 12S as well as for the nc sequences (5.8S–ITS2–28S) were performed with ClustalX 2.0.12 (Larkin et al. 2007) using the default parameters. Lengths of the alignments were 377 bp (16S), 699 bp (12S) and 1364 bp (5.8S–ITS2–28S). The 16S and the 12S alignments were trimmed using the automated option in trimAl v.3.1 (Capella-Gutierrez et al. 2009) at the Phytemon 2 server (Sánchez et al. 2011), resulting in a 351-bp alignment for the 16S and a 654-bp alignment for the 12S sequences, and these trimmed alignments were used for all further analysis.

A test for substitution saturation (Xia et al. 2003) was performed with DAMBE 5.2.68 (Xia & Xie 2001) for the complete \textit{COI} alignment, as well as for all single codon positions separately. Additionally, using the DAMBE graphic tool, transitions were plotted against transversions to obtain a graphic representation of the saturation. The results of the saturation test suggested using all characters, although the third codon position showed a moderate degree of saturation in the plots.
Average p-distances (pairwise exclusion of gaps) were calculated using MEGA version 4 (Tamura et al. 2007), which was also used to calculate neighbour-joining (NJ) trees (Saitou & Nei 1987). For the NJ tree of the nc sequences, we used mid-point rooting. Nodal support was evaluated with non-parametric bootstrapping based on 1000 replicates. A search for the best-fitting substitution model was performed using the Akaike information criterion corrected for small sample size (AICc) as implemented in the jModeltest 0.1.1. (Posada 2008).

Bayesian analyses (BI) were performed using MrBayes 3.1.2 (Huelsenbeck & Ronquist 2001), applying the models of sequence evolution for nucleotide sequences as suggested from jModeltest (HKY + I + G for the three partitions: COI, 16S and 12S). Runs were started with random trees and performed for 3 million generations each with four Markov chains and a sampling frequency of every 100th generation. Those trees generated prior to the stationarity were discarded as burn-in and were not included in the calculation of the consensus trees.

A full median-joining (MJ) network (Bandelt et al. 1999) was constructed with Network 4.6.0.0 (available at www. fluxus-engineering.com), putting equal weight on each site and using the postprocessing option ‘mp calculation’ for the following data sets: COI of clade 2A, COI of T. striolatus as well as for the nuclear sequences. The number of haplotypes was determined with ARLEQUIN 3.11 (Excoffier et al. 2005).

The sequences determined in this study are deposited at GenBank under the accession numbers COI: KJ151294 - KJ151548, 16S: KJ151549 - KJ151617, 12S: KJ151618 - KJ151718, 5.8S-ITS2-28S: KJ151719 - KJ151767. The material is deposited in the mollusc collection of the Natural History Museum Vienna (Mollusca NHMW 109000/AL, individual Ids see Table S1).

Results

Variation in COI sequences

The analysis of the partial mt COI gene (660 bp) amplified from 389 individuals revealed 203 different haplotypes. In the deduced amino acid sequence, most (274 of 383) individuals are identical, with a maximum number of five replacements recorded between two individuals. The most common amino acid sequence occurs in specimens from six different clades and in three species, but there is variation within clades and species (data not shown).

The sampling area (altogether 126 localities) is displayed in Fig. 1. Fig. S1 shows the NJ tree of the COI DNA sequences with several quite distinct and highly supported clades. Two clades (4 and 7) are represented by only one individual each; the others constitute groups of up to 158 (clade 2A) individuals. The clades obtained maximum support, with the exception of clade 3A. The relationships between the clades, however, are not well resolved. Specifically, the more basal nodes are poorly supported.

Seven clades represent taxa that are clearly classified according to morphological features: the species T. villosus, T. clandestinus, T. villosulus, T. striolatus, T. biconicus, as well as the two subspecies of T. oreinos (T. o. oreinos and T. o. scheerpeltzii). Another clade was tentatively assigned to the species T. coelomphala, although this assignment is ambiguous because one of the four individuals in this clade is morphologically very similar to T. bipidius. Besides these taxa, the nine remaining mtDNA clades (1–9) comprise individuals that displayed a high variation in shell size and morphology and were mainly assigned to T. bipidius, while some displayed a more T. sericeus-like phenotype. Morphological variation in each clade was high, and none of the clades was composed of T. sericeus-like phenotypes exclusively. In the following, we refer to the nine clades as ‘T. bipidius complex’ (for details of the morphometric study and the taxon T. sericeus, see Duda et al. revised). Four of these clades are further subdivided into subclades (Fig. S1). The definition of clades and subclades was based on the criteria: high branch support (above 95%), limiting the maximal p-distance within a clade (7%). Interspecific distances and distances within clades are presented in Table S2.

Partial trees showing all individuals of each clade or subclade are presented in figures (Figs S2–S4) and described in detail below.

Phylogenetic relationships between mtDNA clades

Two additional mtDNA marker sequences (12S and 16S), which in general contain conserved parts with lower substitution rates, were sequenced from representatives of each clade (101 individuals). In the BI tree calculated from the concatenated mtDNA sequences (12S, 16S and COI; altogether 1660 bp; Fig. 2), most nodes are well supported, and only two have posterior probabilities (pp) of less than 95%. Hence, concerning the phylogenetic relationships between the clades, we will refer only to the concatenated tree.

In this tree, the basal node separates T. oreinos with its two monophyletic subspecies, and the next node separates T. biconicus from a paraphyletic group comprising the T. bipidius complex and morphologically well-separated species. This paraphyly is supported with high confidence. This paraphyletic assemblage contains two main groups: one is formed by T. villosus, T. clandestinus and clade 8 of the T. bipidius complex, the latter being the sister group of T. villosus and the second main group contains (i) clade 6, (ii) the highly supported sister clades 1 and 9 and (iii) a group with the two sister species T. striolatus and T. villousulus as well as the remaining clades of the T. bipidius complex (2, 3, 4, 5, 7) together with T. coelomphala.
**Variation and geographic distribution of the mtDNA clades of the *T. hispidus* complex**

Distances between clades exceed in several cases those between undisputed species; this is evident in Table S2, which shows the uncorrected mean p-distances of the COI sequences. Distances within clades and subclades ranged up to 6.7% (Table S2) and are illustrated in partial trees (Figs S2–S4). Concerning the distribution of the clades of the *T. hispidus* complex, some geographic patterns become apparent (Fig. 1). Clade 1 represents the western and northernmost localities with specimens from Sweden, which is the type locality of *T. hispidus*, and from the Netherlands. Clade 2 is further divided into two subclades: subclade 2A represents by far the largest sample. It is widely distributed in Austria from the eastern part of the Austrian Alps and adjacent hills and flatland areas up to Tyrol in the west. In the east, it was found in localities along the Danube River, and in the south, it was found close to the Italian and Slovenian borders. In Central Austria exclusively this subclade was found. Subclade 2B shows a disjunct distribution in the north and south of Austria, at the margins of the area covered by subclade 2A. Clade 3 is also divided into two subclades: 3A has a disjunct distribution and shows quite high distances within clades (up to 6.4%). In the west, it was found in mountainous regions of western Austria (Tyrol, Salzburg) and along the Danube in southern Germany (Bavaria); in the south-east, it was detected in the Mecsek Mountains in southern Hungary. Unlike in clade 2A, the individuals in clade 3A cluster according to geographic areas. Subclade 3B was found at only one locality in northern Austria, in the Waldviertel region of Lower Austria. Clade 4 is represented by a single individual, which was found close to the Danube River in Upper Austria, the same locality in which the two individuals forming Clade 5 were also collected. Clade 6 seems to be restricted to the Danube and some of its larger tributaries: subclade 6A, which shows quite low within-clade distances, is restricted to the Danube River floodplains (it was found in localities close to Vienna and in southern Hungary), while 6B occurs along the Inn River in western Austria and the Tauber River in Baden-Württemberg (Germany). Clade 7 consists of a single individual from Sweden, which was found at a locality close to that of clade 1. Subclade 8A was detected in south-western Switzerland.

**Fig. 2** Bayesian tree of the concatenated COI, 16S and 12S sequences. Posterior probabilities are given for all nodes. Black dots indicate maximum support. The scale bar indicates the expected number of substitutions per site according to the model of sequence evolution applied. The colour code is the same as in Fig. 1.
and at the Rhine River (Baden-Württemberg, Germany), and subclade 8B consists of two individuals from one locality in eastern Switzerland. Finally, Clade 9 was found at three localities in Tyrol (Austria) and shows quite low within-group distances.

In summary, nine of the 13 clades and subclades described in the present study occur also in Austria. Furthermore, several quite distinct clades co-occur at the same localities. Nonetheless, some of the clades apparently have disjunct distributions, and in some cases, very similar sequences were found in quite distant localities. Although only few individuals were analysed at most of the 126 sampling sites, at 12 sampling sites, individuals from different (up two-three) mtDNA clades or subclades were obtained (see triangles in Fig. 1). All but one of those 12 sampling sites are located at the margins of the Austrian sampling area. In the central region, there is a wide area in which only subclade 2A was recorded. The following clades occurred together: 2A + 2B, 2A + 9, 3A + 9, 3A + 6B + 9, 4 + 5, 2A + 2B + 3B, 2A + 3B, 2A + 6A, 2A + 3A + 6A.

To assess possible geographic structures in more detail, COI networks were calculated for the subclades 2A and 3A, which include sufficient numbers of individuals (158 and 33, respectively). Fig. 3 shows a median-joining network of subclade 2A, where the colours represent eight geographic regions. The 72 haplotypes form five haplogroups, but none of them has a central position. Group 1 consists almost exclusively of individuals from the north-eastern Alps. However, many individuals from that region occur also in three other clades (2, 3 and 4). Groups 2–5 cannot be assigned to a specific geographic region. Individuals from each of the other seven regions are found within two to four haplogroups. Altogether, no clear geographic clustering of haplotypes is evident within subclade 2A. The network of clade 3A (data not shown) consists of 20 haplotypes arranged in a geographic pattern. The three individuals from the westernmost locality, ‘Stubaier Alpen’ (spID 244), have different haplotypes, which are all in the centre of the network. They are separated by only few (up to six) substitutions from most of the other south-western individuals in this clade. In contrast, the individuals from the eastern and northern localities have quite distinct haplotypes (18–23 substitutions), which are well separated from each other. Only the individuals from the ‘Berchtesgadener Land’ (spID 407 and 412) are present in two different groups: one closer to the western localities (spID 407 with eight substitutions) and the other very distantly related (spID 412 with 18 substitutions).

Genetic variation within T. striolatus and T. oreinos
Concerning the other species, which form well-supported monophyletic groups in the trees, the data allow meaning-

Nuclear marker sequence
The results of the mtDNA analysis raised the question whether the various clades of the T. bispidus complex belong to a single species or might in fact represent several cryptic species. To further approach this question, we analysed a nc marker sequence (5.8S–ITS2–28S) from a subset of individuals representing all clades (except clade 9) and all other species included in the concatenated tree (except the outgroup species).

From six individuals, more than one sequence was obtained by sequencing several clones. At most, three different copies differing by up to five substitutions were found within individuals. The overall mean p-distance for the ITS2 region was 1.7%, while for the 28S region, it was 0.8%. The NJ tree based on these sequences (Fig. 4) reveals a very low differentiation of most of the clades found in the mtDNA tree. However, there are some exceptions: T. oreinos is clearly separated, like in the mtDNA trees. Also, its two subspecies are well differentiated. Furthermore, T. danestinus forms a separate branch. The remaining sequences are present in one clade without any clear pattern. They represent individuals from all analysed clades of the T. bispidus complex as well as members of the species T. villosus, T. villonculus, T. striolatus and T. coelomphala.
Fig. 3 Median-joining network of the COI sequences from subclade 2A. The branches are not drawn to scale, but the number of substitutions is given in red numbers. The size of the circles corresponds to the number of individuals possessing the same haplotype. The geographic origin is reflected by the colours as shown in the map. The five main haplogroups are numbered.
Fig. 4 Neighbour-joining tree and median-joining network of the nuclear 5.8S–ITS2–28S sequences. In the median-joining network, the branches are not drawn to scale, but the number of substitutions is given in red numbers. The size of the circles corresponds to the number of sequences possessing the same haplotype. Abbreviations are the same as in Fig. 1.
Most sequences were identical, while 12 sequences differ from this most frequent haplotype by only one substitution. The remaining sequences differ at up to nine positions (substitutions and indels). The same data set was used to calculate a network (Fig. 4), which illustrates the lack of differentiation between most of the mtDNA clades.

**Discussion**

**Paraphyly of T. hispidus**

The most prominent outcome of this study is the existence of nine highly distinct mtDNA clades, which all are composed of snails with either a *T. hispidus* or a *T. sericeus* morphology. These clades form a paraphyletic assemblage as they are intermingled with clades representing morphologically clearly defined species (e.g. *T. villosus*, *T. villosulus* and *T. striolatus*), each of them forming a well-supported clade. At the current state of knowledge we have no reason to question their species status. Their clear morphological and anatomically differentiation can be interpreted as a reflection of the overall separation of the nc genomes, although this separation is not found in the ncDNA tree. There are several explanations for the incongruence between mtDNA and ncDNA data: incomplete lineage sorting, ongoing or several explanations for the incongruence between mtDNA and ncDNA data: incomplete lineage sorting, ongoing or

Species delimitation in the *T. hispidus* complex

In general, the variation in shell morphology and size is high in all clades of the *T. hispidus* complex, and most clades comprise individuals with ‘typical’ *T. hispidus* habitus as well as individuals displaying the *sericeus* type (narrow umbilicus, globular shell). This was confirmed in a parallel study of morphological and anatomical characters (Duda et al. revised), where the *T. sericeus* phenotype could not be exclusively assigned to a specific clade. However, the specimens analysed here might be specifically distinct from the true *T. sericeus*, and this species has not been covered by our sampling. Therefore, especially investigations of samples from France, which is the type locality for *T. sericeus*, would be very elusive.

There was no clade comprising all individuals with a *T. sericeus* phenotype, but also none of the nine clades representing the *T. hispidus* complex was differentiated in the shell morphological or anatomical analyses (Duda et al. revised). Moreover, the mtDNA clades were not recovered in the ncDNA tree.

The high evolutionary divergence in the mt markers could be accompanied by genomic incompatibility preventing gene flow between co-occurring clades. This was described for the marine copepod *Tigriopus californicus* (Rawson & Burton 2002; Willett 2006). Yet, the number of amino acid replacements in the *COI* sequence within all *Trochulus* clades is rather limited, and they do not correspond to the mt clades. Hence, genomic incompatibility due to the high mitochondrial divergence probably had no major impact in *Trochulus*.
clades. Such analyses were conducted by Depraz (2009) for the Swiss endemic species approach following the biological species concept (Mayr consequences are drawn.

Although this would be in accordance with the phylogenetic species concept (e.g. Cracraft 1983; Nixon & Wheeler 1992), the fact that T. hispidus appears paraplyetic in our mtDNA tree is in our opinion not a sufficient argument for splitting it into different species. There are already several names available for shell variants in the T. hispidus complex, which are currently considered as synonyms; therefore, the identity of already existing names has to be clarified before any other nomenclatural consequences are drawn.

Concerning species delimitation, a straightforward approach following the biological species concept (Mayr 1942, 1970) is to test gene flow between co-occurring clades. Such analyses were conducted by Depraz et al. (2009) for the Swiss endemic species T. picardi and the partially co-occurring T. hispidus/serviceus lineage F. According to that study, restricted gene flow (assessed by microsatellite analysis and mt sequences) indicated that, despite rather small mtDNA divergence, the two lineages are actually distinct species. This finding corroborates earlier morphological analyses by Pfenninger & Pfenninger (2005). In a tree combining the COI data of the present investigation with those of Depraz et al. (2009), T. picardi is nested within our clade 8A, where it is most closely related to four Swiss individuals of our study (data not shown). This might be interpreted as a hint that more cryptic species might be hidden among the clades of the T. hispidus complex.

In our sample, there are several localities where representatives of two or even three clades occur together (black-lined triangles in Fig. 1), and it is likely that there are many more regions where the distribution of clades overlaps. During our sampling, we did not observe that specimens from different clades, which actually occur syntopically, have any recognisable difference in their local environment from which they were taken (e.g. plants). However, we are aware that representatives of these clades might be differentiated in biological parameters not investigated so far. Most of the localities harbouring different clades are situated around the distribution range of subclade 2A. Such localities are ideal settings to test whether or to which extent there is gene flow between individuals belonging to different mtDNA clades. Sympathy of distinct clades without gene flow would be evidence for the presence of distinct species, which might be differentiated by biochemical or life history traits. Currently, we are planning further investigations of the T. hispidus complex with an explicit focus on species delimitation which will be a prerequisite for taxonomic long-term decisions. Concerning the conservation issue, we have to mention that several clades within the T. hispidus complex and local populations of T. striolatus are under pressure, and this should be taken into consideration (for details see Duda et al. revised).

Age of the clades

As mentioned above, the existence of highly divergent mtDNA clades suggests that the radiation of the T. hispidus complex started long before the Pleistocene. The earliest fossil record in Central Europe is from the early Pleistocene (according to Frank 2006; Lozek 1964). Still, the fossil record for the Pliocene and the early Pleistocene is in general scarce, and the assignment of fossils is very problematic due to small conchological differences. Therefore, a reliable calibration to date the splits in our tree is hardly possible, and we refrained from performing a molecular clock analysis. The high mtDNA distances between clades could also be due to an accelerated mutation rate as it was suggested for other snails (e.g. Thomaz et al. 1996). In our nDNA data set, only the two subspecies of T. oreinos form clear clades in both the mt and n trees. They are separated by 0.9% mean p-distance in the nDNA and by a mean distance of 13.7% (15 times higher) in mtDNA. This is in the same order of magnitude as the relation between mt and n rates reported for other organism groups (e.g. 10 times for mammals; Li & Graur 1991). Thus, as our data do not allow to propose an accelerated mt substitution rate, one might ask for explanations for the persistence of these highly divergent clades over a presumably very long time. One possibility is that they (or some of them) might represent cryptic species. This possibility will be tested in a forthcoming study. Alternatively, what we observe as the T. hispidus complex might be indeed a paraplyetic species that retained very large population sizes over long periods of time. From population genetic theory (Avise 2000), monophyly is expected for neutral markers after 4N_e generations (N_e being the effective population size). In this context, it has to be considered that N_e in hermaphrodites depends on whether or not they are simultaneously hermaphrodite, to what extent they may self-fertilise and to what extent they show multiple paternities. A large N_e has been reported in, for example, Cepaea (Murray 1964), but for T. hispidus, no data are available, although large population sizes are probable for this widely distributed species, which frequently has connected habitats along rivulets and rivers (Duda et al. 2010). Under this hypothesis, however,
it is surprising that the haplotypes are arranged in quite separated clades with a geographic pattern rather than a bush-like bundle of haplotypes. This could be explained by the survival and divergence of the highly distinct mtDNA clades in isolation over long periods, especially in the cold phases of the Pleistocene. This is supported by ample fossil record of *T. bispidus* throughout the Pleistocene, which shows a wide distribution, as indicated, for example, by findings in France, Austria, Hungary, Serbia, Croatia and the Czech Republic exist (Binder 1977; Frank et al. 2011; Ložek 1964; Marković et al. 2005; Mohnár et al. 2010; Rousseau et al. 1992; Šúmegi et al. 2011). The scenario of long-lasting isolation does, of course, not exclude the possibility of sporadic contact with gene flow. At least for several clades, the present co-occurrence is evident from our data. High intraspecific mtDNA variability which was explained by refugial isolation and secondary contact has been found also in other land snail species, for example, in *Arianta arbustorum* (Haase et al. 2003, 2013).

**Phylogeographic considerations**

There is a geographic pattern in the comprehensive genetic tree (Fig. 2). This might reflect the phylogeographic history of the taxa investigated. Besides *T. biconicus* and *T. oreinos*, which split from the basal nodes, the main group in the tree, which contains the *T. bispidus* complex, is further subdivided into two clades: (i) one with a more eastern/northern distribution comprising clades 1–7 and 9, *T. coelomphala*, *T. villosulus* and *T. striolatus* and (ii) a western one consisting of clade 8, *T. clandestinus* and *T. villosus*. A similar separation can also be found in Pfenninger et al. (2005), which described a group with more western clades (*T. picardi*, D, E, F, G, H and I) and a group with predominantly eastern clades (A, B and C). This suggests two old radiations starting from unknown western, respectively, eastern regions. The distant position of *T. biconicus* and *T. oreinos* reinforces further investigations on other related taxa, for example, endemics with small distribution ranges such as *T. montanus* and *T. caelatus* or even the species of the related genus *Petasina*, which are integrated into *Trochulus* by some authors such as Procków (2009) and Welter-Schultes (2012).

Concerning glacial refugia, the data still have to be considered as preliminary. For most clades of the *T. bispidus* complex, the present distribution is not assessed yet, and thus, considerations where potential refugia might have been located remain speculative. Anyhow, the *T. bispidus* complex appears to have spent the glacial periods in several different refugia.

The distribution of clade 2A (Fig. 1) suggests that it persisted even throughout glacial periods in the area east of the ice sheet. This interpretation is supported by its exclusive occurrence in those areas, where it does not coexist with any other clade. Other occurrences of 2A in the far west and east of Austria may be the result of postglacial west- and eastward immigration (Fig. 1). The high variation within-clade 2A (up to 4.4%), even within the formerly glaciated region, implies that the re-colonisation started from a large area, in which the variability had not been drastically reduced by a bottleneck. The ample existence of *T. bispidus* in the fossil record of the Pleistocene loess deposits from Lower and Upper Austria, which also includes the cold phases (Binder 1977; Frank et al. 2011) corroborates the hypothesis that this area was a suitable habitat for *T. bispidus* even throughout the glacial periods.

Clade 2B shows a disjunct distribution with a few occurrences in northern Austria as well as in the south-eastern Alps. This finding is interesting given the low genetic distance between the southern and eastern sample. So far, it is unknown whether this clade has a wider distribution.

For areas in western Austria covered by glaciers during the Last Glacial Maximum, a postglacial re-colonisation must be assumed. The occurrence of five quite differentiated clades within this region suggests that they immigrated from different refugia. As *T. bispidus* is apparently an euryoecious species (Procków et al. 2013), dispersal could have taken place quite fast. The locations of those presumed refugia – except of clade 2A, for which a colonisation from the east is plausible – remain speculative. The same applies to the clades found in Sweden, the type locality of *T. bispidus*, which was covered by an ice shield during the last glacial. In our study, we detected two highly distinct lineages there.

Despite the fact that *T. striolatus* is an euryoecious species like *T. bispidus* (Kerney et al. 1983; Procków 2009; Duda et al. 2010), it shows quite low intraspecific variation in the mtDNA (maximum 3.7%) and slightly smaller shell morphological variability (Duda et al. revised). The three subspecies investigated (*striolatus*, *jacovensis* and *danubialis*) are not clearly separated in the COI tree (Fig. S3); there is, however, a geographic pattern indicating that the populations of *T. striolatus* investigated in this study may have been distributed over a wide range during the last glacial. Given the limited sampling in the present study, final conclusions about the phylogeography of this species should include samples from the whole distribution area.

We assume a completely different situation for the *T. oreinos* taxa. Their potential dispersal ability is quite poor due to their specific ecological niche of patchy caricetum-firmae meadows and cool alpine rock areas nearly free of vegetation (Duda et al. 2010). As most parts of their current distribution area remained ice-free during the Last Glacial Maximum (Van Husen 1997), we conclude that they survived at least the last glaciation within this region at the north-eastern margin of the Eastern Alps. Alpine glacial refugia in the same regions were proposed earlier.
(Schönswetter et al. 2002; Rabitsch et al. 2009). The ranges of the two subspecies reflect two separated glacial refugia. Similar cases are documented for several endemic invertebrates and vascular plants in the north-eastern Alps (Rabitsch et al. 2009).

**Consequences for DNA barcoding**

Not all species of the genus *Trochulus* can be easily recognised morphologically; some are distinguished solely by non-distinct characters (e.g. wider umbilicus) and others by anatomical features (Schileyko 1978; Prockow 2009). Moreover, juveniles are hardly determinable, and hence, a DNA barcoding approach could be helpful for identifying individuals collected in the field (e.g. for a biodiversity inventory). Nonetheless, as evident from the data – at least at the current state of knowledge – any barcoding attempt concerning the *T. bispidus* complex appears futile. If a species is not monophyletic in the mtDNA gene tree, it cannot be assigned with a COI barcode, even if the thresholds are set very high. Before species delimitation in the *T. bispidus* complex is accomplished on the basis of more data, trying to define a DNA barcode for *T. bispidus* does not make sense. Difficulties in molecular species delimitation on the basis of DNA data, and hence also for DNA barcoding, have been reported frequently for snails (Davison et al. 2009; Sauer & Hausdorf 2012) reflecting the problems of cryptic species on the one hand and high mtDNA divergence within species on the other hand.

This problem is less relevant for the other *Trochulus* species investigated in this study. Firstly, both *T. oreinos* subspecies can be assigned unequivocally based on their barcodes (maximum 1.4% distance within subspecies and 13.7% between subspecies). Secondly, nearly the whole distribution range of these two taxa was investigated. Concerning *T. striolatus*, all individuals (*n* = 51) morphologically assigned to this species are monophyletic and display low intraspecific variation (maximum 3.7%) over a quite large distribution range. The nearest relative in our tree (*T. villosus*) is separated by 9.5%.

With respect to *T. biconicus, T. villus*, *T. clandestinus* and *T. villosulus*, the COI data of the present investigation are in accordance with those of Pfenninger et al. (2005). Each of these species is monophyletic, and in a pooled data set, the maximum intraspecific distances are low: *T. biconicus* (2%), *T. villus* (2.7%), *T. clandestinus* (0.3%) and *T. villosulus* (2.1%). Therefore, these species should be easily determinable by DNA barcoding.

**Conclusions**

The most striking outcome of the present study is the very well-supported paralogy revealed in the mtDNA phylogeny of the *T. bispidus* complex. None of the nine mtDNA clades shows a morphological differentiation; however, it remains questionable if any of them might represent cryptic species. Therefore, gene flow has to be tested in populations where representatives of multiple clades can be found. The presence of morphologically well-defined species within the complex, together with possible cryptic species, suggests that budding speciation in the genus *Trochulus* occurred frequently. The *T. bispidus* complex is a very prominent example of a species in which, despite the intense investigation, species delimitation remains unclear. This makes further analysis even more interesting because in the process of speciation, the mechanisms (such as hybridisation) that provoke such a pattern are still not sufficiently understood.

**Acknowledgements**

We thank the Austrian Science Foundation for financial support of this work (FWF Project No. 19592-B17). The ‘Friends of the Museum of Natural History Vienna’ provided financial support for travel expenses. Zoltan Feher, Ted von Proschwitz, Ira Richling and Ulrich Schnepf provided important samples from regions outside the Eastern Alps. For help during collection trips, special thanks go to Helmut Baminger, Barbara Daubl, Agnes Bisenberger, Philipp Haselwanter, Christoph Hörweg, Paul Mildner, Laura Zopp, Wilhelm Pinsker, Anatoly Schileyko, Reiko Slapnik, Erich Weigand and Sabine Zwierschitz. Furthermore, we thank Laura Zopp for archiving the specimens and maintaining the database. We are very grateful to Wilhelm Pinsker for many valuable discussions and critical comments on the article. The file with the LGM glacial line was generously provided by Peter Schönswetter and was first published in (Schönswetter et al. 2002).

**References**


Budding speciation in the genus *Trochulus* • L. Kruckenauer et al.


L. Kruckenbauer et al. • Budding speciation in the genus Trochulus
Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Fig. S1.** Neighbour-joining tree of the COI sequences from all individuals providing an overview of the different clades, subclades and species. The height of the triangles represents the number of individuals found in each clade, and the depth of the triangle reflects the divergence of haplotypes in each clade. Black dots indicate maximum support. Bootstrap support above 75% is indicated. Note that the tree contains only little information about the relationships between the clades. The colour code is the same as in Fig. 1. Details are given in Figs S2–S4.

**Fig. S2.** Partial tree of the neighbour-joining tree of the COI sequences (Fig. S1) illustrating the clade 2 in detail. Each individual is defined by an individual Id, the origin of geographic region and the Id of the sampling sites. For relevant nodes, bootstrap values are indicated.

**Fig. S3.** Partial tree of the neighbour-joining tree of the COI sequences (Fig. S1) illustrating the various clades and subclades in detail: Clades 3 to 7 as well as T. coelomphala, T. striolatus, and T. villosulus. Each individual is defined by an individual Id, the origin of geographic region and the Id of the sampling sites. For relevant nodes, bootstrap values are indicated. The haplogroups (hg1–hg5) and subspecies of T. striolatus are indicated.

**Fig. S4.** Partial tree of the neighbour-joining tree of the COI sequences (Fig. S1) illustrating the various clades and subclades in detail: Clades 1, 8, and 9 as well as the species T. clandestinus, T. oreinos, T. bicornis, T. villosus, Monacha cantiana, Plicuteria lubomirski, and Isognomostoma isognomostomos. Each individual is defined by an individual Id, the origin of geographic region and the Id of the sampling sites. For relevant nodes, bootstrap values are indicated.

**Table S1.** Summarising information for the individuals investigated in this study together with GenBank accession numbers and coordinates of sampling sites. Individual Id (indId), Geographic origin (region), sampling site ID (spID), Country (cn). Abbreviations of clades as in Fig. 1 and Is_iso (Isognomostoma isognomostomos); Mo_can (Monacha cantiana); Pl_lub (Plicuteria lubomirski); Clade numbers correspond to those in Figs. 1, 2 and 4.

**Table S2.** Mean and maximum genetic p-distances (in %) between COI sequences within species and clades of T. hispidus (upper two lines) and mean p-distances between species and clades. Abbreviations of clades as in Fig. 1 and Pl_lub (Plicuteria lubomirski); Mo_can (Monacha cantiana); Is_iso (Isognomostoma isognomostomos). Clade numbers correspond to those in Figs. 1, 2 and 4.