

The geography of speciation in narrow endemics of the “*Haenydra*” lineage (Coleoptera, Hydraenidae, *Hydraena*)

Ignacio Ribera^{1,2*}, Agustín Castro³, Juan A. Díaz⁴, Josefina Garrido⁵, Ana Izquierdo¹, Manfred A. Jäch⁶ and Luis F. Valladares⁷

1 Institute of Evolutionary Biology (CSIC-UPF), Barcelona, Spain

2 Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain

3 IES Clara Campoamor, Lucena, Córdoba, Spain

4 Universidad de Santiago, Lugo, Spain

5 Universidad de Vigo, Vigo, Spain

6 Naturhistorisches Museum Wien, Wien, Austria

7 Universidad de León, Spain

*Correspondence: Ignacio Ribera, Institute of Evolutionary Biology (CSIC-UPF), Passeig Marítim de la Barceloneta 37, 08003 Barcelona, Spain. E-mail: ignacio.ribera@ibe.upf-csic.es

ABSTRACT

Aim We compare different scenarios for the origin of the endemic western Mediterranean species of aquatic Coleoptera of the “*Haenydra*” lineage (Hydraenidae, *Hydraena*) through contrasting predictions of the relationship between their phylogenetic and geographic distance. We test whether species originated 1) through successive periods of dispersal and speciation, or through fragmentation of the range by 2) random vicariance or 3) geographical isolation due to a general rarefaction of the population density.

Location Europe.

Methods We build a phylogeny using 3Kb of three mitochondrial and two nuclear gene fragments of ca. 50% of the known species of “*Haenydra*”, including all the western Mediterranean taxa. We estimated divergence times with a molecular clock and compiled the distribution of the species and their geographical centroids. We tested the relationship between geographical and phylogenetic distance using bivariate plots, mantel tests and the comparison of the observed phylogeny with the one minimising geographical distances between species, as measured with an Euclidean Minimum Spanning Tree (EMST).

Results The monophyly of “*Haenydra*” was strongly supported, although its phylogenetic placement within *Hydraena* was not resolved. “*Haenydra*” was estimated to be of Tortonian origin, with most of the current species originating during the Pleistocene. Four well-sampled lineages including most of the Iberian endemics were chosen for the detailed geographical analyses. In two of them (*H. tatii* and *H. emarginata* clades) there was a significant association between geographical and phylogenetic distance, and the reconstructed phylogeny was identical to that obtained through the EMST, demonstrating a strong non-randomness of the geographical distribution of the species. In the other two (*H. iberica* and *H. bitruncata* clades) there was no association between geographical and phylogenetic distance, and the observed phylogeny was not the one minimising geographical distances. In one of the clades this lack of association seems to be due to the secondary recent expansion of one of the species (*H. iberica*), erasing the geographical signal of their distributions.

Main conclusions We show that it is possible to obtain strong evidence of stasis of the geographical ranges of narrow endemic species through the study of their phylogenetic relationships and their current distributions. In at least two of the studied clades current species seem to have originated through the fragmentation of a more widely distributed species, without further range movements. This process of range expansion and fragmentation seems to have occurred repeatedly within the “*Haenydra*” lineage, contributing to the accumulation of narrow endemics in Mediterranean Pleistocene refugia.

Key words Speciation, range expansion, range shifts, Pleistocene refugia, narrow endemics, Iberian Peninsula

INTRODUCTION

One of the potentially most fruitful links between biogeography and evolution is the study of the geography of speciation. Since the early formulations of speciation modes (sympatric, allopatric, peripatric, Mayr, 1963) the fundamental role of geography has been recognised, and there have been many attempts to reconstruct the history of speciation through the distribution of current species (e.g. Lynch, 1989; Barraclough & Vogler, 2000; Fitzpatrick & Turelli, 2006). There is a recognised weakness common to all these studies: species do change their geographical ranges, and it cannot be assumed that the current geographical range of a species is the same as that at the moment of speciation, or that ranges are kept through the cladogenetic process (Gaston, 2003). This prompted many authors to conclude that evolutionary inferences concerning the geography of species in the past will often not be reliable (Chesser & Zink, 1994; Gaston, 1998; Losos & Glor, 2003). However, it would be equally wrong to assume that all species have suffered modifications in their ranges large enough to erase any geographical signal from the past, as in some cases there is strong evidence to support the stasis of geographical ranges, either through the fossil record (e.g. Jablonski, 1987) or with indirect evidence from ecological and phylogenetic data (e.g. Carranza & Wade, 2004; Martínez-Solano et al., 2004). A particularly difficult case are the lineages with abundance of narrowly distributed, mostly allopatric species. The reduced range (often the result of specialised ecological requirements) and non overlapping distribution, sometimes through several cladogenetic events (Fitzpatrick & Turelli, 2006), strongly suggest allopatric speciation, but then the question is if the species have originated, and always persisted, where they are currently found.

A possible way to test the persistence of a geographical signal in the current distribution of a clade of species is through the comparison of the observed phylogenetic and spatial relationships with a random null model (Barraclough & Nee, 2001). In this way, it is possible to contrast alternative scenarios for the temporal origin and phylogenetic relationships of the species through the comparison of the observed and predicted spatial patterns. Using this approach, we test here potential scenarios for the origin of several clades with mostly allopatric, narrowly distributed species in a genus of European water beetles. These scenarios lead to a different relationship between the geographical and phylogenetic distance between the species (Fig. 1), so that by looking at this relationship it should be possible to discriminate among them:

1) Range expansion through successive bouts of dispersal with subsequent speciation. This would be generally equivalent to “island hopping”, or stepping-stone colonisation (MacArthur & Wilson, 1967). The starting situation is a small range to which new areas are added sequentially, to be eventually “lost” due to speciation. The resulting pattern will be a general positive relationship between phylogenetic and geographic distance, with more distant species having the oldest divergences. This relationship will be asymmetrical (triangular in a bivariate plot, Fig. 1a): while there could not be species which are geographically distant but phylogenetically close (unless there is long range dispersal), there could be species that are geographically close but phylogenetically distant (e.g. species resulting from the initial, most ancient splits). The age of the species will generally decrease with respect to the geographical origin of the range expansion. Typical examples would be the colonisation of archipelagos (Gillespie & Roderick, 2002; Keppel et al., 2009), or of new available areas e.g. by progressive deglaciation (Hewitt, 2000).

2) Range expansion with subsequent speciation due to a fragmentation of the initial range. In this case the starting situation is the maximum expand of the range, which becomes fragmented and reduced with time, leading to speciation. Depending on the nature of the barriers fragmenting the initial range two outcomes could be expected:

If the range is fragmented due to random vicariance there should be no correlation between geographical and phylogenetic distance (Fig. 1b). By “random” we mean that the location and the temporal succession of barriers to gene flow, determining the topology of the phylogeny, is independent of the shape of the range of the species or its phylogeographic structure, even if

coherent between different lineages with the same general characteristics living in the same area. Typical examples could be fragmentation of a range by increase of sea level, or tectonic fragmentation of microplates (e.g. Sanmartín, 2003).

Alternatively, the range may be fragmented due to rarefaction of the gene flow when there is a progressive and more or less uniform degradation of the general conditions that allowed the initial range expansion. This would be equivalent to the “refuge speciation” of Moritz et al. (2000) or the vicariance by niche conservatism of Wiens (2004), when the barriers due to suboptimal conditions are not established randomly (as in the previous scenario) but occur through the lines of minimum gene flow, which would correspond to the largest area with the lowest population density (Fig. 1c). In this case, the general relationship between geographical and phylogenetic distance should increase monotonically but not linearly, as it could be expected that the phylogenetic distance is inversely proportional to the square of the linear geographical distance. The age of the species will generally increase with respect to the centroid of the distribution of the clade. Under a perfect scenario, the match between geographical and phylogenetic distances should be optimal, that is, the topology of the observed phylogeny should be the one that minimises the geographical distances between species.

We test these scenarios using a lineage of aquatic beetles with abundance of narrowly distributed, allopatric species (the “*Haenydra*” lineage of the genus *Hydraena*, family Hydraenidae; Hansen, 1998). Our specific aim is to ascertain whether the south-western European species of the group could be said to have persisted in the same areas in which they are currently found since their origin, that is, whether they are local endemics or have suffered range shifts (e.g. as consequence of glacial cycles) large enough to have erased the phylogenetic signal of their current distribution. The three possibilities outlined above differ in the resulting relationship between geographical and phylogenetic distances, although the power of the conclusions varies depending on the observed pattern: if a strict positive relationship is found (according to the third scenario) this can be taken as a strong evidence of a non-random geographical arrangement of the species. However, the existence of deviations from this strict proportionality (either partially, as in Fig. 1a, or completely, as in Fig. 1b) could indicate either a compliance with the predictions derived from the initial hypotheses or the breakdown of the expected pattern due to subsequent changes in the range of some of the species.

MATERIALS AND METHODS

Background on the taxonomy of the group

The genus *Hydraena* Kugelann, with ca. 700 known species and many more to be described, is the most diverse of family Hydraenidae, and probably the most diverse among the aquatic Coleoptera (Jäch & Balke, 2008). Species of *Hydraena* are typically fully aquatic, living in the riparian area of small streams and rivers. Many authors have recognised groups of apparently closely related species based on external morphology, which have in some cases being named as subgenera, or, more recently, as informal species groups (Jäch et al., 2000). One of these traditionally recognised groups is *Haenydra*, described by Rey (1886) as a separate genus, and considered by different authors both as a genus (e.g. Ieniştea, 1968; Rocchi, 2009) or subgenus (e.g. Berthélemy, 1986; Perkins, 1997; Hansen, 1998). In a cladistic analysis of *Hydraena* sensu lato Jäch et al. (2000), even acknowledging their likely monophyly, synonymised *Haenydra* (together with all previously defined subgenera with the exception of *Hydraenopsis* Janssens), as they would render *Hydraena* sensu stricto paraphyletic, and considered it only as an informal species group (the *Hydraena gracilis* group).

Currently there are 86 recognised species of the “*Haenydra*” lineage (Hansen, 1998; Jäch, 2004; Audisio et al., 2009), usually found in clean, fast flowing waters, often in mountain streams.

They share the absence of parameres in the aedeagus and a similar external morphology, typically narrow and elongated (Fig. 2). They are distributed in the north Mediterranean region from Iberia to Iran (Hansen, 1998; Jäch, 2004). Many species of this lineage have very restricted distributions, often limited to a single valley or mountain system, but there are also some species with very wide geographical ranges, such as e.g. *H. gracilis*, present in the whole Europe from north Iberia to the Urals (Jäch, 2004).

Taxon sampling

We included a comprehensive sampling of the west European species of “*Haenydra*”, including all species occurring in the Iberian Peninsula, plus a representation of species from other Mediterranean areas (Appendix S1). Missing species in some of the species groups could be tentatively placed according to the external morphology and the characteristics of the male genitalia. In total, we included examples of 39 named species of “*Haenydra*”.

The monophyly of *Hydraena* is generally accepted (Perkins, 1989; Jäch et al., 2000; Beutel et al., 2003), but there is no agreement on the internal phylogeny and, in particular, in the phylogenetic position of the “*Haenydra*” lineage. Therefore, trees were rooted in *Adelphydraena*, hypothesised to be the sister group of *Hydraena* (Perkins, 1989), and in any case clearly outside “*Haenydra*”. Within the genus *Hydraena* we sampled as outgroups for “*Haenydra*” a comprehensive selection including most species groups as defined from morphology (Jäch et al., 2000), among them all previously defined subgenera with the sole exception of *Spanglerina* Perkins (with four Neotropical species, Hansen, 1998) (Appendix S1).

For the species of “*Haenydra*” we follow the taxonomy and nomenclature of Jäch (2004), except for two cases, *H. saga* and *H. catalonica*. *Hydraena saga* as currently recognised has a disjunct distribution, in the Pyrenean region in the west, and from the eastern Alps to Bulgaria in the east (Jäch, 2004; Fig. 4). Preliminary molecular and morphological data strongly suggest that the populations in these two areas are not closely related (unpublished observations), and we consider here as “*H. saga* complex” only the Pyrenean populations. Similarly, our phylogenetic results showed that *H. catalonica* was not monophyletic (see below), and we consider separately the populations from the Pyrenees and the Montseny massif, in central Catalonia (Fig. 4).

DNA extraction and sequencing

Specimens were collected alive in the field and directly killed and preserved in 96% ethanol. DNA was extracted from whole specimens by a standard phenol-chloroform extraction or the DNeasy Tissue Kit (Qiagen GmbH, Hilden, Germany). Vouchers and DNA samples are kept in the collections of the Museo Nacional de Ciencias Naturales (MNCN, Madrid) and the Institute of Evolutionary Biology (IBE, Barcelona) (Appendix S1). DNA extraction was non-destructive, to preserve voucher specimens for subsequent morphometric and morphological study. Typically only males were sequenced, and the male genitalia (used for the identification of the species) dissected and mounted previous to the extraction to ensure a correct identification.

We sequenced five gene fragments, two mitochondrial (5' end of cytochrome c oxidase subunit 1, *cox1*, and 5' end of large ribosomal unit plus the Leucine transfer plus the 3' end of NADH dehydrogenase subunit 1, *rrnL+trnL+nad1*) and two nuclear (small ribosomal unit, *SSU*, large ribosomal unit, *LSU*) (see Appendix S2 for the primers used). For each fragment both forward and reverse sequences were obtained. In some specimens the *cox1* fragment was amplified using internal primers to obtain two fragments of around 400 bp each (Appendix S2). Sequences were assembled and edited with Sequencher 4.7 (Gene Codes, Inc., Ann Arbor, MI). New sequences have been deposited in GenBank with Acc. Nos. XX-XX (Appendix S1). Protein coding genes were

not length variable, and the ribosomal genes were aligned with the online version of MAFFT v.6 using the G-INS-i algorithm and default parameters (Kato & Toh, 2008).

Phylogenetic analyses

Bayesian analyses were conducted on a combined data matrix with MrBayes 3.1.2 (Huelsenbeck & Ronquist, 2001), using five partitions corresponding to the five sequenced fragments (the *rrnL+trnL* fragment was considered a single partition) and a GTR+I+G model independently estimated for each partition. MrBayes ran for 15×10^6 generations using default values, saving trees every 500. “Burn-in” values were established after visual examination of a plot of the standard deviation of the split frequencies between two simultaneous runs.

We also used Maximum Likelihood as implemented in the on-line version of RAxML (which includes an estimation of bootstrap node support, Stamatakis *et al.*, 2008), using GTR+G as the evolutionary model and the same five gene partitions used in MrBayes.

Estimation of divergence times

To estimate the relative age of divergence of the lineages we used the Bayesian relaxed phylogenetic approach implemented in BEAST v1.4.7 (Drummond & Rambaut, 2007), which allows variation in substitution rates among branches. We implemented a GTR+I+G model of DNA substitution with four rate categories using the mitochondrial data set only and pruning specimens with missing gene fragments. We used an uncorrelated lognormal relaxed molecular clock model to estimate substitution rates and the Yule process of speciation as the tree prior. Well supported nodes in the analyses of the combined sequence (mitochondrial and nuclear) were constrained to ensure that the BEAST analyses obtained the same topology. We ran two independent analyses for each group sampling every 1000 generations, and used TRACER version 1.4 to determine convergence, measure the effective sample size of each parameter and calculate the mean and 95% highest posterior density interval for divergence times. Results of the two runs were combined with LogCombiner v1.4.7 and the consensus tree compiled with TreeAnnotator v1.4.7 (Drummond & Rambaut, 2007).

The analyses were run for 30×10^6 generations, with the initial 10% discarded as burn-in. Because of the absence of fossil record to calibrate the trees we used as a prior a rate of 2.0% of pairwise divergence per MY, established for a closely related family (Leiodidae) for a combination of mitochondrial markers including those used here (Ribera *et al.*, 2010b). We set as a prior rate a normal distribution with average rate of 0.01 substitutions/site/MY, with a standard deviation of 0.001.

Geographical analyses

We compiled contour maps of the distribution of the species of the “*Haenydra*” lineage included in the studied clades (see Results) from published and unpublished sources (Jäch, 2004; Sánchez-Fernández *et al.*, 2008; checklist of the species of the Italian fauna, v. 2.0, www.faunaitalia.it) (Fig. 4). Species' range centroids (center of mass of the polygon representing the distribution of a species) and distances between centroids were calculated using ArcGIS 9.2 (Environmental Systems Research Institute Inc., Redlands, CA) (Table 1). To check the association between phylogenetic and geographical distance between the centroids we used three approaches:

1) Bivariate plots of the linear distance between the centroids of the species ranges and the branch lengths of the ultrametric trees (i.e. the estimated age of divergence between species, “phylogenetic distance” on what follows) (Fig. 1).

2) Mantel tests for the general association between the matrices of geographic and phylogenetic distances. Multiple Mantel tests were done using ZT v. 1.1 (Bonet & Van de Peer, 2002), with 10,000 randomisations to generate a null distribution and assess the significance of the results.

3) An optimisation procedure to assess the match between the observed geographical distribution and the topology obtained from the phylogeny. We first compute the Euclidean minimum spanning tree (EMST), i.e. the minimum spanning tree of a set of n points in the plane (the centroids of the distributions), where the weight of the edge between each pair of points is the linear distance between those two points. The EMST connecting n vertices will have $n(n-1)/2$ edges, which are computed through a standard minimum spanning tree algorithm (see e.g. Graham & Hell, 1985), a trivial task for graphs of less than six nodes. The result is a graph connecting all points minimising the weight of the edges, i.e. the distances among centroids. In the scenario outlined in Fig. 1c, the temporal sequence of cladogenetic events will be determined by the length of the edges connecting the centroids: the first split will be between the taxa at both extremes of the longest edge, the second will split the second longest, and so on until the two closest species are separated.

To assess the probability that the observed topology could be identical to that obtained with this optimization procedure we obtained all possible unrooted bifurcating topologies of each of the studied clades in Paup v.2 (Swofford, 2002), and considered them as a null distribution against which the probability of each individual topology was estimated. Note that the use of the EMST determines not only the final topology but the relative order of all the cladogenetic events. We did not consider the relative order in the cases in which the cladogenesis occurs in two different branches of the tree, as this does not affect the final topology (see results below).

RESULTS

Phylogeny of *Hydraena*

The final matrix included 94 taxa and 2831 aligned characters. Part of the *rrnL+trnL+nad1* fragment was missing for two species (Appendix S1), and for two of the repeated specimens of *H. gracilis* only *cox1* was sequenced. The nuclear markers (*SSU* and *LSU*) were sequenced only for a representation of the species of “*Haenydra*” due to the general low variability within this lineage, with many identical sequences between closely related species (Appendix S1).

The runs of MrBayes converged to split frequencies lower than 0.01 at 11×10^6 generations, leaving a total of $4 \times 2 \times 10^6$ generations for the sampling of the tree space (=16,000 trees). The monophyly of *Hydraena* was not well supported (Fig. 2), as well as the basal relationships among the five major clades: 1) the subgenus *Hydraenopsis* (as defined in Jäch et al., 2000); 2) the South African *H. monikae*; 3) the “*Phothydraena*” lineage (*H. testacea* species group sensu Jäch et al., 2000); 4) the *H. rugosa* and *H. circulata* species groups (which were sisters, with good support in the analyses with Bayesian probabilities but not in the ML), and 5) the main lineage within *Hydraena* sensu stricto (including the “*Haenydra*” lineage), which was well supported in both analyses (Fig. 2; see Appendix S1 for the composition of the species groups).

Within the main lineage of *Hydraena* s.str., the *H. palustris* group was sister to the rest (in agreement with Jäch et al., 2000), which was divided in three well supported main clades: 1) *H. bisulcata* and its allies, 2) a clade broadly including the *H. riparia*, *H. minutissima*, *H. rufipes* and *H. holdhausi* groups, and 3) the “*Haenydra*” lineage (Fig. 2). The relationships between these three main clades were not well resolved, with MrBayes favouring a sister relationship between “*Haenydra*” and the *H. bisulcata* group, and RAxML favouring a sister relationship with the *H.*

riparia group (sensu lato), albeit in both cases with low support. In all cases the monophyly of the “*Haenydra*” lineage was strongly supported (Bayesian posterior probability, Bpp = 1.0; ML bootstrap, MLb = 100%; Fig. 2).

Phylogeny of the “*Haenydra*” lineage

There were three well supported main lineages within “*Haenydra*”, together with some species with an isolated position (*H. carniolica*, *H. schuleri* and *H. subintegra*, Fig. 2). Of these three main lineages, one (*H. iberica* lineage) included four Iberian endemics, with uncertain phylogenetic relationships within “*Haenydra*”. The *H. gracilis* lineage included the Iberian endemic *H. exasperata* as sister to a group of central and eastern European species, and the Iberian endemics *H. saga* complex, *H. emarginata* and *H. hispanica* as sister to two species from the Alps (Figs 2, 4). The third, the *H. dentipes* lineage, included two clades with narrow western Mediterranean endemics, one with *H. catalonica* and *H. bitruncata*, and the second with five Iberian species, the *H. tatii* group (sensu Fresneda et al., 1994) plus the morphologically isolated *H. monstrosipes* and *H. zezerensis* (Figs 2, 4).

Two of the species were found to be paraphyletic: in the *Hydraena gracilis* complex (sensu Jäch, 1995), the north Iberian populations were sister to specimens from Britain to Turkey, including *H. anatolica* (Jäch, 1995); and specimens of *H. catalonica* from the Montseny massif (central Catalonia) were sister to the Pyrenean *H. catalonica* plus *H. bitruncata*. For the geographical analyses *H. catalonica* was split in its two geographical lineages, the populations from the Pyrenees (“*H. catalonica* p”) and the populations from the Montseny (“*H. catalonica* m”) (see Methods).

Estimation of divergence times

According to the results of the Beast runs, and using a calibration of 0.01 substitutions/site/MY, the origin of the “*Haenydra*” lineage was estimated to be at ca. 8.5 MY (Tortonian), with a wide confidence interval (Fig. 3). The three main lineages (*H. iberica*, *H. gracilis* and *H. dentipes* lineages) originated ca. 6 MY ago (Messinian), and most species, including all Iberian endemics, were estimated to be less than 2.6 MY old, i.e. of Pleistocene origin (Fig. 3). There were relatively deep divergences within some of the species with wider distributions, such as *H. polita* (0.8 MY between specimens from the Pyrenees and south Germany), or *H. heterogyna* (0.8 MY between specimens from central Italy and the French Alpes Maritimes, Fig. 3; Appendix S1).

Geographical analyses

For the geographical analyses we selected four well supported clades of “*Haenydra*” including most of the south-western narrowly distributed species, and for which the sampling (according to morphology) could be considered complete or with at most a single missing species (see Fig. 4 for the distribution of the species in each of the clades).

1) The *H. iberica* clade, including *H. iberica*, *H. madronensis*, *H. lusitana* and *H. altamirensis*, all of them Iberian endemics (Fig. 4a). Its sister group was not well established (Fig. 2), but in any case it would include species fully overlapping with some of those included in the *H. iberica* clade. The sampling was complete, with no other known species likely to be included.

This clade included only four species, below the minimum necessary for the implementation of Mantel tests in ZT (Bonet & Van de Peer, 2002). All potential sister groups fully overlap with some of the species in the clade, so the geographical distance between the centroids with the

potential sister could be lower than the distance between some members of the clade, erasing the geographical signal. The linear correlation between the geographical and phylogenetic distance of the four species of the clade in the bivariate plot was not significant (Fig. 6a; $R^2 = 0.02$; $p > 0.05$). The three data points corresponding to the species *H. iberica* were outside a hypothetical linear relationship between geographical and phylogenetic distance (Fig. 6a). The centroid of the distribution of this species was closer to *H. lusitana*, and more distant to *H. altamirensis* and *H. madronensis*, than would correspond for their respective phylogenetic distances to obtain a linear relationship, suggesting a possible secondary range expansion of *H. iberica*. To explore this possibility we sequenced the *cox1* fragment of five additional specimens of *H. iberica* through their range (Appendix S1). They all had identical sequence, with the only exception of one base pair for the specimen of south Portugal (voucher MNCN-AI386, Appendix S1), at the southwestern limit of its distribution (Fig. 4a), supporting the hypothesis of a recent expansion.

The observed topology was not in agreement with that obtained from the EMST (Fig. 5a). The topology optimising the geographical distances (EMST) placed *H. altamirensis* as sister to *H. madronensis*, in contrast with the observed relationship (*H. iberica* sister to *H. madronensis*, Fig. 2). In any case, due to the low number of different unrooted trees for four taxa (three), the observed tree could not be said to be different from a random geographical arrangement.

2) The *H. emarginata* clade, including *H. emarginata*, *H. saga* complex, *H. hispanica*, *H. tarvisina* and *H. larissae* (Figs 2, 4b). The first three species are Iberian endemics (with some uncertainty in the case of *H. saga* complex), and the latter two are endemic to the Alps. The sister group of this clade was not well established. The sampling was also complete, with no missing close relative of any of the included species.

Geographical and phylogenetic distances were significantly correlated in this clade, as measured with a Mantel test with 10,000 permutations (Tables 2, 3; $r = 0.9$, $p < 0.05$). The bivariate plot (Fig. 6b) shows a monotonic increase of phylogenetic distance with the distance between centroids without apparent outliers, and in particular without phylogenetically distant species pairs in close geographic proximity. Of all possible unrooted topologies with five taxa (15), the observed one was identical to that determined by the EMST. The observed relative order of two of the nodes in different branches was reversed with respect to the order expected with the EMST (Fig. 5b), although they were estimated to have occurred in close temporal proximity, and the 95% confidence intervals fully overlap each other (Fig. 3). Due to the low number of possible topologies (15) the geographic distribution of the species could only be considered to be marginally different from random ($p < 1/15 = 0.067$). The *H. emarginata* clade was sister to a large group including widespread species such as *H. gracilis* or Iberian endemics such as *H. exasperata*, with a pooled distribution fully overlapping to that of the *H. emarginata* clade. The inclusion of the combined sister group would thus erase the geographical signal and render all tests not-significant.

3) The *H. bitruncata* clade, including *H. catalonica* (split in two groups, the Pyrenean and the populations of the Montseny), *H. bitruncata*, *H. polita* and *H. bensae* (Figs 2, 4c). The first two species have narrow distributions in the NE of the Iberian Peninsula and south France, *H. polita* has a widespread distribution from north Iberia to the eastern Alps, and *H. bensae* is endemic to the Alpes Maritimes (Fig. 4c). The sister of this clade was again not well established. All these species share the absence of a small flagellum in the apical part of the median lobe of the aedeagus, present in the rest of the species of the *H. polita* group (*H. dentipes*, *H. producta*, *H. heterogyna* among those included in the study). According to morphology, *H. bicuspidata*, from the Massif Central in SE France (close to Lyon, Ganglbauer, 1901), should also be included in this clade, as it shares the absence of this flagellum and a very similar body shape. We did the geographical analyses including the studied species only, and with *H. bicuspidata* as sister to *H. catalonica* sensu lato plus *H. bitruncata*.

The Mantel test between geographical and phylogenetic distances with the sampled species of the clade was not significant at the standard level (Tables 2, 3; $r = 0.68$, $p = 0.08$), with the

distances corresponding to *H. polita* (the more widespread species of the group) as outliers from a linear relationship in the bivariate plot (Fig. 6c). The observed topology was not the one optimising geographical distances (Fig. 5c), with two topologies with a better match with the EMST: ((*H. bitruncata*, *H. catalonica* m) *H. catalonica* p), and ((*H. catalonica* m, *H. catalonica* p) *H. bitruncata*) ($p = 3/15 = 0.2$).

When the missing species *H. bicuspidata* was added to the phylogeny in the most likely position according to morphology (as sister to *H. bitruncata* plus *H. catalonica* sensu lato, see above), in the middle of the branch linking these two species with *H. polita*, the Mantel test became significant despite the reduction in the correlation, due to the increase in power ($r = 0.32$; $p < 0.05$). The non-randomness of the observed topology (with the assumption of the position of *H. bicuspidata*) was marginally significant: there were five topologies that match the EMST better than the observed one, i.e. $p = 6/105 = 0.057$.

The sister group of this clade was estimated to be *H. dentipes* plus *H. producta*, although with low support. This fact, added to the likely existence of several missing Alpine species, prevented the inclusion of further taxa in the analyses.

4) The *H. tatii* clade, including five Iberian narrow endemics (*H. tatii*, *H. manfredjaechi*, *H. gaditana*, *H. zezerensis* and *H. monstruosipes*) as sister to some species in the Alps and Italy, *H. devincta*, *H. devillei*, *H. leonhardi* and *H. lapidicola* (Figs 2, 4d). There are three likely missing species in the eastern group: *H. sanfilippoi*, close to *H. lapidicola* (Audisio & de Biase, 1995); *H. bosnica*, close to *H. leonhardi* (Audisio et al., 1996) and *H. hungarica*, also related to *H. bosnica* and *H. leonhardi* (the three share the female gonocoxite with two small, symmetric depressions). All these species have, however, a fully allopatric distribution with the Iberian species, and thus we pooled their distributions and considered them as a single “sister outgroup” taxon. For the analyses the geographical distance was considered to be the shortest of this pooled distribution from the Iberian Peninsula (i.e. the Alpes Maritimes, Table 1). *Hydraena sappho* Janssens, from the small island of Levkás (Greece), has been associated with the Iberian *H. tatii* clade (Audisio et al., 1996). A closer examination of the only known specimen (the holotype, Janssens, 1965) showed that it is most likely related to species from the eastern Mediterranean (M.A. Jäch, manuscript in preparation).

The sister relationship between the species *H. tatii*, *H. manfredjaechi* and *H. gaditana* with *H. zezerensis* plus *H. monstruosipes* was not well supported, although the node was present in all analyses (Maximum Likelihood and Bayesian, both with the full combined sequence and with the mitochondrial data only). According to our results the sister group of the *H. tatii* clade was *H. truncata* (although with low support, Fig. 2), which has a widespread European distribution including the NW of the Iberian Peninsula.

The Mantel test between the geographic and phylogenetic distances was highly significant, both when only the five Iberian species were included ($r = 0.90$, $p < 0.05$), or when the pooled Alpine species were included as a non-overlapping outgroup to the Iberian species (Tables 2, 3; $r = 0.83$, $p < 0.01$). The bivariate plot did not show any obvious outliers at either side of the diagonal of the plot (Fig. 6d). The observed topology was identical to that determined by the EMST, and only two nodes in different branches had a reversed relative order of cladogenesis (Fig. 5d), both when the Iberian endemics were considered alone (five taxa, $p = 1/15 = 0.067$) or when the Alpine sister group was included (six taxa, $p = 1/105 < 0.01$). The two nodes with reversed order were estimated to have occurred in close temporal proximity (1.3 and 1.6MY), with fully overlapping 95% confidence intervals (Fig. 3).

DISCUSSION

Origin and phylogeny of the “*Haenydra*” lineage

The monophyly of the species of the “*Haenydra*” lineage was strongly supported, including *H. schuleri*, with a somewhat deviating morphology. Although the phylogenetic position of the lineage was not resolved, it was in any case nested within the main clade of *Hydraena* s.str., in agreement with the conclusions of Jäch et al. (2000) based on morphology. In Perkins (1997) “*Haenydra*” was considered a “primitive” subgenus sister to the remaining species of *Hydraena*, due to the assumed plesiomorphic exocrine secretion delivery system of the pronotum and the reduced number of striae on the elytra (following Berthélemy, 1986). Our results do not support this view. Jäch et al. (2000) suggested that the *H. armipalpis* group (China) could be the sister to *Haenydra*, with the base of the parameres fused with the median lobe of the aedeagus and a similar general structure of the pronotum and elytra, but no species of this group could be obtained for the molecular study.

Within the wider *Hydraena* the main trends of our phylogeny also agree perfectly with the results of Jäch et al. (2000), with several well-defined lineages including *Hydraenopsis* and other species groups considered to be plesiomorphic (*H. monikae*, *H. rugosa* group, “*Phothydraena*” and *H. circulata* group). Our results confirm the inclusion of “*Hadrenya*” (i.e. the *H. minutissima* group, represented by *H. pygmaea*) within the main lineage of *Hydraena* s.str. (close to the species of the *H. riparia* group), as hypothesized by Jäch et al. (2000), but not close to “*Haenydra*”, as assumed by previous authors (Orchymont, 1925; Perkins, 1997).

According to our calibration, the “*Haenydra*” lineage originated in the Tortonian, ca. 10 MY ago, but the main diversification of the group, and the origin of most of the species, took place during the Pleistocene. This post-Messinian diversification would explain the absence of any species of the group in North Africa, despite the intensive search in the area by numerous entomologists and the obvious dispersal abilities of some species through continuous landmasses. There is a record of *H. exasperata* from Morocco in Orchymont (1935) (a male, deposited in the Institut royal des Sciences naturelles de Belgique, Bruxelles), but, as suggested by the same author, it may be a labelling mistake or a specimen carried over from the previous collecting sites in south Spain. The only species present in western European islands are *H. gracilis* in Britain and Ireland, likely to have been connected to mainland Europe during periods of lower sea level, the last one during early Holocene after the last glacial maximum (Lambeck & Chappell, 2001), and three species in Corsica and Sardinia (Audisio et al., 2009). The latter species are often hypothesised to be the result of vicariance after the tectonic separation of the Corsican microplate during the Oligocene (see Audisio et al., 2009 for an overview of possible scenarios). According to our estimations based on the same vicariant split in a related family (Leiodidae, Ribera et al., 2010b), the Corsican *H. evanescens* has an estimated age of 5.2 MY, i.e. the end of the Messinian crisis. This opens the possibility of a colonisation of Corsica and Sardinia through land connections during the Messinian, with vicariance of the island populations after the opening of the Straits of Gibraltar with the re-filling of the Mediterranean (García-Castellanos et al., 2009). Some other Sardinian endemics have been estimated to be of a similar age (Messinian), in particular some cave salamanders of the subgenus *Speleomantes* (Carranza et al., 2008), or the painted frog *Discoglossus sardus* (Zangari et al., 2006). The dispersal of these taxa would have required a land corridor that was most likely also suitable for specimens of *Haenydra*.

All the Iberian endemic species were estimated to be of Pleistocene origin, in agreement with some groups of aquatic beetles (e.g. family Dytiscidae, Ribera & Vogler, 2004), but in sharp contrast with others, such as the two Iberian endemic species of *Enicocerus* Stephens (Hydraenidae, Ribera et al., 2010a) or the western Mediterranean species of Hydrochidae (Hidalgo-Galiana & Ribera, in prep.), all of them of Tortonian or Messinian origin. The most widespread and common species of the lineage, *H. gracilis*, has a recent origin nested within a clade with predominantly narrow endemics. Differences between specimens through the range (Latvia, Britain, Bulgaria) were minimal, strongly suggesting a recent, late Pleistocene expansion of the range. On the contrary, other widespread species of the group (*H. polita*, *H. truncata*) have deep divergences between specimens at different parts of their ranges (estimated to be ca. 1 MY), suggesting the existence of frequent cryptic diversity within the lineage. This is also the case of the species found

to be paraphyletic (including *H. gracilis* sensu lato), or known to be a composite of polyphyletic lineages (as the current concept of *H. saga*).

Geographical analyses

The difficulty in determining past geographical ranges in the absence of fossil record is a major impediment for the use of geographical data in studies of speciation and diversification (Gaston, 2003; Losos & Glor, 2003). In some cases the accumulation of indirect evidence (genetic, ecological, geographic) adds to the credibility of a given scenario, but it is often not possible to contrast its likelihood against alternative possibilities. The use of null models as a reference for the comparison of the observed pattern allows a more rigorous assessment (Barraclough & Nee, 2001). We did not test the geography of speciation in the classic sense (Mayr, 1963), as this seems not to be possible using the only data of the current distributions. In our scenario, speciation ultimately may occur as a result of isolation through rarefaction of the range, but whether this could be considered allopatric or peripatric would depend on the precise distribution of the species at the moment in which speciation is considered to have taken place, which is likely not possible to establish without detailed genetic data (Butlin et al., 2008). Our aim was to understand the processes that could lead to speciation and divergence and their relation to geography, rather than a characterisation of the detailed speciation mechanisms (Fitzpatrick et al., 2009), and in this sense our evolutionary scenario is admittedly an oversimplification.

The finding that for a given clade there is a strong correlation between geographical and phylogenetic distance is a clear evidence of the non-randomness of the spatial distribution of the species. This correlation may be the result of a process in which the probability of speciation is inversely proportional to the distance to the nearest neighbours, as could happen in a progressive rarefaction of a species range due to changes in general climatic or ecological conditions (Fig. 1c). In the traditional models of isolation by distance (Wright, 1943; Kimura, 1953) geographical distance is also proportional to genetic distance, but this proportionality is maintained through the continuous presence of gene flow, more likely to occur between neighbour individuals (or populations). Given a widespread species showing isolation by distance through its range, if the general conditions deteriorate so that the gene flow diminishes progressively it could be expected that the first interruption would occur among the groups of populations separated by the longest distance, followed by the next longest and so on successively, resulting in a topology matching an EMST among the final species.

We have not assessed the degree of overlap between species (as in Lynch, 1989 or Barraclough & Vogler, 2000), as our results only depend on the relative position of the centroid of the distribution of each species in relation to the others, not on possible range expansions or contractions around this centroid. The relative position of the centroids seems to be less evolutionary labile than the extension of the range, which will usually change faster than the rate of speciation (Coyne & Orr, 2004). Our scenarios do not assume either that the range of the ancestral species is the sum of that of the descendants, only that the relative position of the centroid of their distribution is intermediate between the two. Of the two statistical tests we applied to the geographical data, the Mantel test is the least stringent, as it could be still significant when there is a geographical “swap” between two close species, or when (due e.g. to a geographically biased range expansion) the current centroid changes its relative position with respect to the nearest neighbours. On the contrary, the optimization test through the comparison of the topologies is more restrictive, in that any change of the relative position of even nearest neighbour species would result in a suboptimal topology. This is clearly exemplified in the *H. bitruncata* clade, with a significant correlation between geographical and phylogenetic distance as measured with a Mantel test when *H. bicuspidata* is included in its most likely position, but with several topologies with a better match to the geographical optimum and thus not significantly different from a random arrangement.

A limitation of our approach is that while a positive result is a clear indication that species have kept their relative geographical positions, in the cases in which we do not find a significant correlation, or the topology does not optimise geographical distances, it is not possible to affirm that there has been range movement. As seen in Fig. 1, other modes of speciation (e.g. vicariance by random breaks) will result in this lack of correlation even if the species remain in place. In the two clades with the stronger evidence of non-random geographical distribution (*H. tatii* and *H. emarginata* clades) the relationship between phylogenetic and geographical distances increased monotonically, without outliers that could suggest the presence of species phylogenetically distant but geographically close – as expected under a speciation by stepping-stone scenario (Fig. 1a). The relationship was not linear, also as expected (see Introduction and Fig. 1c), but in the clade with the highest number of species (*H. tatii* clade) it became linear with the transformation $P^3 = aG$ (data not shown), strongly suggesting that the phylogenetic distance did indeed decrease with the square of the geographical distance ($P = bG^{-2}$). In both cases the most ancient split was also between the two more distant groups of species, those in the Iberian Peninsula and the Alps, as would be expected with a fragmentation by rarefaction after a range expansion, but not with a speciation subsequent to stepping-stone dispersal.

Even assuming the main diversification mechanism of the “*Haenydra*” lineage (and possible other lineages with abundance of narrow endemics) is the succession of cycles of expansion of some species with subsequent fragmentation (Fig. 1c), the geographical signal, as measured here, will persist only until the next expansion of a species of the clade. This could be the case in the *H. iberica* clade, in which both the genetic uniformity and the deviation from the linear correlation of *H. iberica* strongly suggest a recent expansion from its original range, which could be estimated to be in central Iberia if the geographical distances are interpolated from a linear relationship with phylogenetic distances in Fig. 5a. The most widespread species of the “*Haenydra*” lineage, *H. gracilis*, was found to be of recent origin (ca. 0.5MY), and the few data available shows that the central and northern European form is very homogeneous through its range, suggesting a recent expansion. Although we did not analyse this clade in detail due to the likely high number of closely related missing species, the recognition of distinct taxa in the periphery of its current distribution (Jäch, 1995) suggests that this could be an example of a species complex in the early stages of range fragmentation. Some of these “peripheral” taxa are however island endemics (e.g. *H. elisabethae* in the island of Thassos, or *H. nike* in Samothrace, Jäch, 1995). In these cases, the isolation would be produced by hard, “random” vicariant barriers and there does not need to be a correlation between geographical and phylogenetic distance.

The *H. tatii* clade shows the strongest evidence for a non-random distribution among the four tested. Its origin was estimated to be at ca. 3.7 MY, and the five Iberian endemics have a common origin at ca. 3.3 MY, i.e. by late Pliocene. This is coincident in time with an acute cooling period that may have facilitated the expansion of the ancestral species, prior to the origin of the Mediterranean climate at ca. 3.1-3.2 MY, with a strong seasonality and an increase in aridity (Suc, 1984; Mijarra et al., 2009). Subsequent cladogenetic events within the *H. tatii* clade would have taken place during the Pleistocene glacial cycles within the Iberian Peninsula, and without changes in the geographical location of the resulting species – or at least changes not large enough to erase the geographical signal in their current distribution.

Although with lower support, the *H. bitruncata* clade also showed some evidence of geographical structure when *H. bicuspidata* was included in its most likely phylogenetic position, with a significant overall correlation between geographical and phylogenetic distances. The origin of this clade was estimated to be at the Pliocene-Pleistocene transition, again a cold period (Lisiecki & Raymo, 2007) that could have made possible the expansion of species typical of cold mountain streams. The distribution of this clade, between NE Iberia (north of the Ebro valley) and the Alps falls outside the traditionally recognised Pleistocene refugia (the southern peninsulas), but still show signs of conservation of the geographical ranges.

The most common view of the effect of glacial cycles are that they induce regular expansions during favourable times (either glacials or interglacials depending on the autecology of

the species), followed by range contraction to refugia when conditions turn adverse (Dynesius & Jansson, 2000; Bennet & Provan 2008; Stewart et al. 2010). These repeated cycles produce different degrees and patterns of phylogeographic structure (Hewitt, 2000). However, it seems that for some lineages, among them *Haenydra*, the range expansions are infrequent and affecting only some taxa. The period of contraction of the range results in the generation of multiple, isolated “residual” species. The process would thus not be cyclical in the sense that conditions return to the same original state, but accumulative: each expansion produces a set of new species that do not contribute to the next cycle, and overlap with the species resulting from the previous ones. The concept of “refugia” (as defined by Stewart et al., 2010) will apply not to species, as they would not suffer cyclical periods of contraction-expansion, but to the lineage: successive glaciations would eradicate populations (or species) in the glaciated areas, allowing the survival (and accumulation) of the species only in the refugia. If the species are able to expand their ranges only occasionally, as seems to be the case of “*Haenydra*”, either they remain in the refugia as narrow endemics, or, when they expand, they do not mix with the populations that are left as they would already be different species.

Concluding remarks

With our approach we have shown that under some circumstances it is possible to obtain strong evidence of stasis of the geographical ranges of narrow endemic species through the study of their phylogenetic relationships and their current distributions. At least one of the studied clades, with five narrow, fully allopatric Iberian species, seems likely to be formed by true endemics, originated in the areas in which they are currently found through the fragmentation during the late Pliocene and the Pleistocene of a more widely distributed species. This “speciation within refugia” supports the increasing evidence of the complexity of the evolutionary processes that took place in the Mediterranean peninsulas during the Pleistocene glacial cycles, which were far from being simple “repositories” of accumulated genetic diversity (Gómez & Lunt, 2007).

ACKNOWLEDGEMENTS

We thank all colleagues mentioned in Appendix S1 for supplying material for study, A. Faille and A. Hidalgo for help in the lab, W. Zelenka for the habitus of Fig. 2, P. Abellán for geographical data, and A. Cieslak, J. Castresana and P. Abellán for comments. This work was funded through the EU program Synthesys (projects AT-TAF-217, 1613, 2201 and 2391 to IR, JAD, LFV and JG respectively) and projects CGL2004-00028 and CGL2007-61665 to IR.

REFERENCES

- Audisio, P. & de Biase, A. (1995) Materials towards a biogeographical analysis of Italian *Hydraena* and *Haenydra* (Coleoptera, Hydraenidae). *Fragmenta Entomologica*, **27**, 163-189.
- Audisio, P., de Biase, A. & Jäch, M.A. (1996) Greek species of the genus *Hydraena* (subgenus *Haenydra* Rey) (Coleoptera, Hydraenidae). *Aquatic Insects*, **18**, 65-90.
- Audisio, P., Trizzino, M., de Biase, A., Mancini, E. & Antonini, G. (2009) A new species of *Hydraena* (Coleoptera: Hydraenidae) of the *H. evanescens* complex from Sardinia. *Zootaxa*, **2318**, 281-289.
- Barracough, T.G. & Nee, S. (2001) Phylogenetics and speciation. *Trends in Ecology and Evolution*, **16**, 391-399.
- Barracough, T.G. & Vogler, A.P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist*, **155**, 419-434.

- Bennett, K.D. & Provan, J. (2008) What do we mean by 'refugia'? *Quaternary science reviews*, **27**, 2449-2455.
- Berthélemy, C. (1986) Remarks on the genus *Hydraena* and revision of the subgenus *Phothydraena* (Coleoptera: Hydraenidae). *Annales de Limnologie*, **22**, 181-193.
- Beutel, R.G., Anton, E. & Jäch, M.A. (2003) On the evolution of adult head structures and the phylogeny of Hydraenidae (Coleoptera, Staphyliniformia). *Journal of Zoological Systematics and Evolutionary Research*, **41**, 256-275.
- Bonnet, E. & Van de Peer, Y. (2002) zt: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **7**, 101-112.
- Butlin, R.K., Galindo, J. & Grahame, J.W. (2008) Sympatric, parapatric or allopatric: The most important way to classify speciation? *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, **363**, 2997-3007.
- Carranza, S., Romano, A., Arnold, E.N. & Sotgiu, G. (2008) Biogeography and evolution of European cave salamanders, *Hydromantes* (Urodela: Plethodontidae), inferred from mtDNA sequences. *Journal of Biogeography*, **35**, 724-738.
- Carranza, S. & Wade, E. (2004) Taxonomic revision of Algero-Tunisian *Pleurodeles* (Caudata: Salamandridae) using molecular and morphological data. Revalidation of the taxon *Pleurodeles nebulosus* (Guichenot, 1850). *Zootaxa*, **488**, 1-24.
- Chesser, R.T. & Zink, R.M. (1994) Modes of speciation in birds - a test of Lynch's method. *Evolution*, **48**, 490-497.
- Coyne, J.A. & Orr, H.A. (2004) Speciation. Sinauer Associates, Sunderland, 545 pp.
- Dynesius, M. & Jansson, R. (2000) Evolutionary consequences of changes in species' geographical distributions driven by Milankovitch climate oscillations. *Proceedings of the National Academy of Sciences of the USA*, **97**, 9115-9120.
- Drummond, A. & Rambaut, A. (2007) Beast: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Fitzpatrick, B.M., Fordyce, J.A. & Gavrillets, S. (2009) Pattern, process and geographic modes of speciation. *Journal of Evolutionary Biology*, **22**, 2342-2347.
- Fitzpatrick, B.M. & Turelli, M. (2006) The geography of mammalian speciation: Mixed signals from phylogenies and range maps. *Evolution*, **60**, 601-615.
- Fresneda, J., Lagar, A., Sáinz-Cantero, C.E. & Delgado, J.A. (1994) Sobre las especies ibéricas del grupo de *Hydraena tatii* (Coleoptera, Hydraenidae). *Zoologica Baetica*, **5**, 27-39.
- Ganglbauer, L. (1901) Beiträge zur Kenntniss der paläarktischen Hydrophiliden. *Verhandlungen der Zoologisch-Botanischen Gesellschaft in Wien*, **51**, 312-332.
- García-Castellanos, D., Estrada, F., Jiménez-Munt, I., Gorini, C., Fernández, M., Vergés, J. & De Vicente, R. (2009) Catastrophic flood of the Mediterranean after the Messinian salinity crisis. *Nature*, **462**, 778-781.
- Gaston, K.J. (1998) Species-range size distributions: Products of speciation, extinction and transformation. *Philosophical Transactions of the Royal Society, London, B, Biological Sciences*, **353**, 219-230.
- Gaston, K. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford, 280 pp.
- Gillespie, R.G. & Roderick, G.K. (2002) Arthropods on islands: Colonization, speciation, and conservation. *Annual Review of Entomology*, **47**, 595-632.
- Gómez, A. & Lunt, D.H. (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In: S. Weiss & N. Ferrand (Eds), *Phylogeography of southern European refugia*. Springer, Dordrecht, pp. 155-188.
- Graham, R.L. & Hell, P. (1985) On the history of the Minimum Spanning Tree problem. *Annals of the History of Computing*, **7**, 43-57.
- Hansen, M. (1998) *World catalogue of insects. Volume 1. Hydraenidae (Coleoptera)*. Apollo Books, Stenstrup, 168 pp.
- Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907-913.

- [Hidalgo-Galiana, A. & Ribera, I. Late Miocene diversification of the genus *Hydrochus* (Coleoptera, Hydrochidae) in the west Mediterranean area. *Journal of Biogeography*, under review (submitted 15 January 2010).]
- Huelsenbeck, J.P. & Ronquist, F. (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754-755.
- Ieniştea, M.A. (1968) Die Hydraeniden Rumäniens (Coleoptera, Hydraenidae). *Travaux du Muséum National d'histoire naturelle Grigore Antipa*, **8**, 759-795.
- Jablonski, D. (1987) Heritability at the species level: Analysis of geographic ranges of Cretaceous mollusks. *Science*, **238**, 360-363.
- Jäch, M.A. (1995) The *Hydraena* (*Haenydra*) *gracilis* Germar species complex (Insecta: Coleoptera: Hydraenidae). *Annalen des Naturhistorischen Museums in Wien, Series B*, **97**, 177-190.
- Jäch, M.A. (2004) Hydraenidae. In: I. Löbl & A. Smetana (Eds), *Catalogue of Palaearctic Coleoptera. Volume 2: Hydrophiloidea - Histeroidea - Staphylinoidea*. Apollo Books, Stenstrup, pp. 102-122.
- Jäch, M.A. & Balke, M. (2008) Global diversity of water beetles (Coleoptera) in freshwater. *Hydrobiologia*, **595**, 419-442.
- Jäch, M.A., Beutel, R.G., Díaz, J.A. & Kodada, J. (2000) Subgeneric classification, description of head structures, and world check list of *Hydraena* Kugelann (Insecta: Coleoptera: Hydraenidae). *Annalen des Naturhistorischen Museums in Wien, Series B*, **102**, 177-258.
- Janssens, E.L. (1965) Les *Hydraena* de l'Égée. *Mémoires de l'Académie Royale de Belgique, Classe des Sciences*, **16(4)**, 1-126.
- Katoh, K. & Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286-298.
- Keppel, G., Lowe, A.J. & Possingham, H.P. (2009) Changing perspectives on the biogeography of the tropical south pacific: Influences of dispersal, vicariance and extinction. *Journal of Biogeography*, **36**, 1035-1054.
- Kimura M (1953) "Stepping-stone" model of population. *Annual Report of the National Institute of Genetics, Japan*, **3**, 62-63.
- Lambeck, K. & Chappell, J. (2001) Sea level change through the last glacial cycle. *Science*, **292**, 679-686.
- Lisiecki, L.E. & Raymo, M.E. (2007) Plio-Pleistocene climate evolution: trends and transitions in glacial cycle dynamics. *Quaternary Science Reviews*, **26**, 56-69.
- Losos, J. & Glor, R.E. (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology and Evolution*, **18**, 220-227.
- Lynch, J.D. (1989) The gauge of speciation: on the frequencies of modes of speciation. In: D. Otte & J.A. Endler (Eds), *Speciation and its consequences*. Sinauer Associates, Sunderland, MA, pp. 527-553.
- Macarthur, R. & Wilson, E.O. (1967) *The theory of island biogeography*. Princeton University Press, Princeton, NJ, 203 pp.
- Martínez-Solano, I., Gonçalves, H., Arntzen, J. & García-París, M. (2004) Phylogenetic relationships and biogeography of midwife toads (Discoglossidae: *Alytes*). *Journal of Biogeography*, **31**, 603-618.
- Mayr, E. (1963) *Animal species and Evolution*. Harvard University Press, Cambridge, 811 pp.
- Mijarra, J.M.P., Barron, E., Manzaneque, F.G. & Morla, C. (2009) Floristic changes in the Iberian Peninsula and Balearic islands (south-west Europe) during the Cenozoic. *Journal of Biogeography*, **36**, 2025-2043.
- Moritz, C., Patton, J.L., Schneider, C.J. & Smith, T.B. (2000) Diversification of rainforest faunas: an integrated molecular approach. *Annual Review of Ecology and Systematics*, **31**, 533-563.
- Orchymont, A. d' (1925) Contribution à l'étude des hydrophilides. iv. *Bulletin et Annales de la Société Royale d'Entomologie de Belgique*, **65**, 298-305.

- Orchymont, A. d' (1935) Inventa entomologica itineris hispanici et maroccani quod a. 1926 fecerunt Harald et Hakan Lindberg. XXII. Palpicornia. *Commentationes Biologicae, Helsingfors*, **5**, 1-22.
- Perkins, P.D. (1989) *Adelphydraena*, new genus, and two new species from Venezuela, and remarks on phylogenetic relationships within the subtribe Hydraenina (Coleoptera: Hydraenidae). *Proceedings of the Biological Society of Washington*, **102**, 447-457.
- Perkins, P.D. (1997) Life on the effective bubble: Exocrine secretion delivery systems (ESDS) and the evolution and classification of beetles in the family Hydraenidae (Insecta: Coleoptera). *Annals of the Carnegie Museum*, **66**, 89-207.
- Rey, C. (1886) Histoire naturelle des Coléoptères de France (suite). *Annales de la Société linnéenne de Lyon*, **32**(1885), 1-187, pl. 1-2.
- Ribera, I., Castro, A. & Hernando, C. (2010a) *Ochthebius (Enicocerus) aguilerai* sp.n. from central Spain, with a molecular phylogeny of the Western Palaearctic species of *Enicocerus* (Coleoptera, Hydraenidae). *Zootaxa*, **2351**, 1-13.
- Ribera, I., Fresneda, J., Bucur, R., Izquierdo, A., Vogler, A.P., Salgado, J.M. & Cieslak, A. (2010b) Ancient origin of a western Mediterranean radiation of subterranean beetles. *BMC Evolutionary Biology*, **10**, 29.
- Ribera, I. & Vogler, A.P. (2004) Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, **13**, 179-193.
- Rocchi, S. (2009) Coleotteri acquatici endemici presenti nell'appennino settentrionale (coleoptera: Dytiscidae, hydraenidae, scirtidae, elmidae). *Onychium*, **7**, 48-52.
- Sánchez-Fernández, D., Bilton, D.T., Abellán, P., Ribera, I., Velasco, J. & Millán, A. (2008) Are the endemic water beetles of the Iberian Peninsula and the Balearic islands effectively protected? *Biological Conservation*, **141**, 1612-1627.
- Sanmartin, I. (2003) Dispersal vs. vicariance in the Mediterranean: Historical biogeography of the Palearctic Pachydeminae (Coleoptera, Scarabaeoidea). *Journal of Biogeography*, **30**, 1883-1897.
- Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology*, **57**, 758-771.
- Stewart, J.R., Lister, A.M., Barnes, I. & Dalen, L. (2010) Refugia revisited: individualistic responses in space and time. *Proceedings of the Royal Society B*, **277**, 661-671.
- Suc, J.P. (1984) Origin and evolution of the Mediterranean vegetation and climate in Europe. *Nature*, **307**, 429-432.
- Swofford, D.L. (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (* and Other Methods) Version 4*. Sinauer, Sunderland, MA.
- Wiens, J.J. (2004) Speciation and ecology revisited: phylogenetic niche conservatism and the origin of species. *Evolution*, **58**, 193-197.
- Wright, S. (1943) Isolation by distance. *Genetics*, **28**, 114-138.
- Zangari, F., Cimmaruta, R. & Nascetti, G. (2006) Genetic relationships of the western Mediterranean painted frogs based on allozymes and mitochondrial markers: evolutionary and taxonomic inferences (Amphibia, Anura, Discoglossidae). *Biological Journal of the Linnean Society*, **87**, 515-536.

Biosketch

This work is part of an ongoing collaboration among the authors to study the systematics and biogeography of the Western Mediterranean Hydraenidae, with a special focus on the Iberian Peninsula.

Tables

Table 1. Centroids of the distribution of the species in the clades used for the geographical analyses (in decimal coordinates). See Fig. 4 for the distribution of the species.

clade	species	X	Y
<i>H. iberica</i>	<i>H. altamirensis</i>	-4.868	39.504
	<i>H. iberica</i>	-6.474	41.221
	<i>H. lusitana</i>	-7.614	41.277
	<i>H. madronensis</i>	-4.316	38.391
<i>H. emarginata</i>	<i>H. emarginata</i>	-2.746	42.894
	<i>H. hispanica</i>	-7.157	41.588
	<i>H. larissae</i>	9.395	46.041
	<i>H. saga complex</i>	-0.015	42.630
	<i>H. tarvisina</i>	11.588	46.003
<i>H. bitruncata</i>	<i>H. bensaie</i>	6.000	44.175
	<i>H. bicuspidata</i>	4.430	45.350
	<i>H. bitruncata</i>	2.314	42.253
	<i>H. catalonica p</i>	1.262	42.415
	<i>H. catalonica m</i>	2.384	41.815
	<i>H. polita</i>	6.071	45.869
<i>H. tatii</i>	<i>H. gaditana</i>	-5.293	36.679
	<i>H. manfredjaechi</i>	-2.541	38.228
	<i>H. monstruosipes</i>	-6.858	42.771
	<i>H. tatii</i>	-3.757	37.132
	<i>H. zezerensis</i>	-7.584	40.385

Table 2. Matrices of geographical linear distance between the centroids of the species (in km). See text and Fig. 2 for the definition of the “sisters outgroup” in the *H. tatii* clade. In the *H. bitruncata* clade, “*H. catalonica p*” refers to populations from the Pyrenees, and “*H. catalonica m*” from the Montseny.

<i>H. iberica</i> clade	altamirensis	iberica	lusitana		
iberica	236.3				
lusitana	308.1	96.0			
madronensis	132.9	367.0	431.0		
<i>H. emarginata</i> clade	emarginata	hispanica	larissae	saga complex	
hispanica	390.1				
larissae	1025.4	1414.6			
saga complex	225.5	600.5	839.8		
tarvisina	1190.0	1579.9	170.1	998.0	
<i>H. bitruncata</i> clade	bensae	bicuspidata	bitruncata	catalonica p	catalonica m
bicuspidata	175.0				
bitruncata	367.2	380.0			
catalonica p	430.9	415.0	88.9		
catalonica m	393.6	450.0	48.9	114.8	
polita	188.0	100.0	500.5	541.9	537.5
<i>H. tatii</i> clade	gaditana	manfredjaechi	monstruosipes	tatii	zezerensis
manfredjaechi	295.8				
monstruosipes	690.5	627.5			
tatii	145.5	160.6	682.9		
zezerensis	460.2	501.3	270.7	495.7	
[sister outgroup]	1300.0	1000.0	1100.0	1100.0	1300.0

Table 3. Matrices of phylogenetic distances (i.e. age estimates, in MY) between the species. See text and Fig. 2 for the definition of the “sister outgroups” in the *H. tatii* clade. In the *H. bitruncata* clade, “*H. catalonica p*” refers to populations from the Pyrenees, and “*H. catalonica m*” from the Montseny. In the *H. bitruncata* clade, the age estimate of *H. bicuspidata* (not included in the study) is the middle point of the branch between the two nodes in which is hypothesized to be placed based on morphological evidence (see text).

<i>H. iberica</i> clade	altamirensis	iberica	lusitana	madronensis	
iberica	1.582				
lusitana	4.517	4.517			
madronensis	2.295	2.295	4.517		
<i>H. emarginata</i> clade	emarginata	hispanica	larissae	saga complex	
hispanica	0.997				
larissae	1.980	1.980			
saga complex	0.637	0.997	1.980		
tarvisina	1.980	1.980	0.744	1.980	
<i>H. bitruncata</i> clade	bensae	bicuspidata	bitruncata	catalonica p	catalonica m
bicuspidata	4.784				
bitruncata	4.784	1.9065			
catalonica p	4.784	1.9065	0.123		
catalonica m	4.784	1.9065	1.755	1.755	
polita	4.784	3.813	3.813	3.813	3.813
<i>H. tatii</i> clade	gaditana	manfredjaechi	monstruosipes	tatii	zezerensis
manfredjaechi	3.158				
monstruosipes	6.511	6.511			
tatii	1.040	3.158	6.511		
zezerensis	6.511	6.511	2.497	6.511	
[sister outgroup]	7.490	7.490	7.490	7.490	7.490

Figure legends

Figure 1. Schematic representation of the different hypothesized scenarios for the speciation after a range expansion. In the first column, geographical distribution of the species (A to E); second column, phylogenetic relationships among them; third column, bivariate plot of geographical linear distances (G) and phylogenetic distance (P) (approximate values). a) Speciation by stepping-stone colonization; b) speciation by vicariance due to the random appearance of barriers to gene flow (represented by lines); c) speciation by rarefaction of gene flow between favourable patches. The three scenarios result in different patterns in the relationship between phylogenetic and geographic distance: a) triangular relationship, with no closely related species at large geographical distances; b) no relationship; and c) positive relationship, with neither closely related species at large geographical distances nor distant phylogenetic species at close geographical distances (see text for details). Numbers indicate the temporal sequence of the cladogenetic processes.

Figure 2. Phylogram of the species of *Hydraena* obtained in MrBayes. Numbers in nodes, Bayesian posterior probability / bootstrap support in RAxML; “-” marks nodes not present in the RAxML analyses. With vertical bars, the four clades used in the geographical analyses. Habitus: *H. catalonica*.

Figure 3. Ultrametric tree obtained with Beast, using the mitochondrial data of a selection of the specimens in the “*Haenydra*” lineage and constraining the well-supported nodes of the topology represented in Fig. 2 (black circles). The four clades studied in the geographical analyses are marked in red. Numbers in nodes, age estimate (MY); blue bars, 95% confidence intervals. Vertical bar, Pliocene / Pleistocene transition (2.6 MY).

Figure 4. Maps with the distribution of the species included in the clades used in the geographical analyses, with their centroids (see Table 1 for the coordinates). a) *H. iberica* clade; b) *H. emarginata* clade; c) *H. bitruncata* clade; d) *H. tatii* clade (in the Alpes Maritimes, minimum distance with the species of the Alpine group).

Figure 5. Schematic representation of the Euclidean Minimum Spanning Trees (EMST) among the centroids of the species in the four clades used for the geographical analyses (a, *H. iberica* clade; b, *H. emarginata* clade; c, *H. bitruncata* clade; d, *H. tatii* clade). Circles represent the geographical position of the centroids, as in Fig. 4. Lines between centroids represent cladogenetic events splitting species at both sides of it. Numbers represent the temporal order of the cladogenetic events. In b) and c), the reconstructed phylogeny (taken from Fig. 3) agrees with the EMST except for the temporal order of some nodes in different branches, not affecting the topology (numbers in blue, observed sequence; in red, temporal sequence according to the EMST). In a) and c) the observed phylogeny (in blue) does not agree with the phylogeny reconstructed from the EMST (in red).

Figure 6. Bivariate plots between the geographical (linear) distance between the centroid of the species (G, km) and the phylogenetic distance (P, age estimate in MY). a) *H. iberica* clade, grey circles, distances to *H. iberica*; b) *H. emarginata* clade; c) *H. bitruncata* clade, grey circles, distances to *H. polita*; white circles, distances to *H. bicuspidata*; d) *H. tatii* clade.

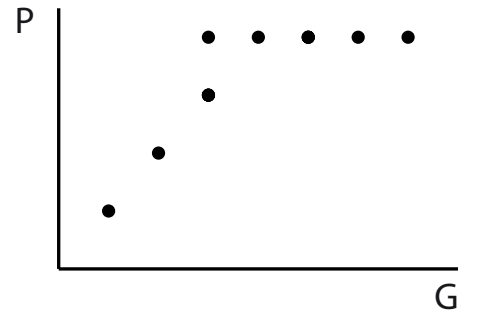
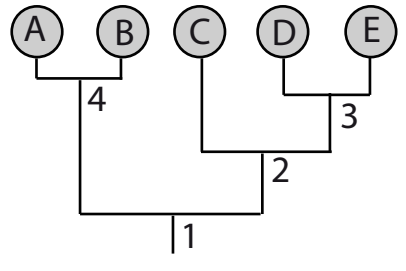
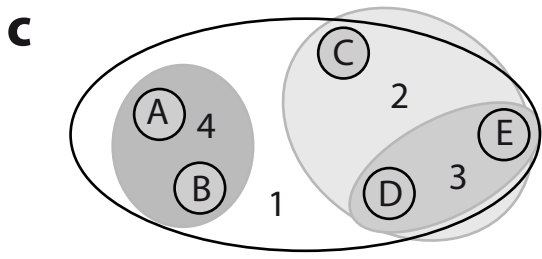
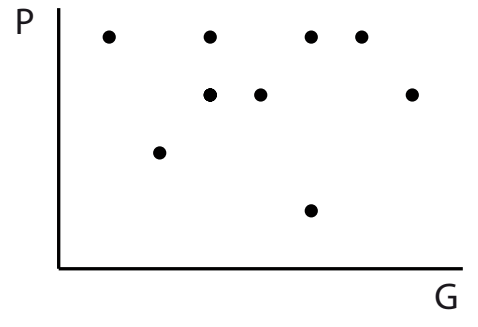
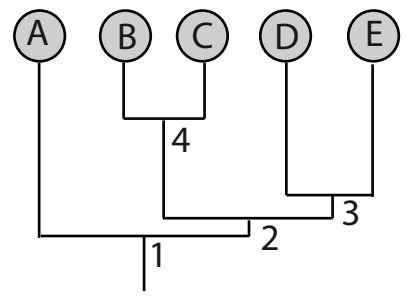
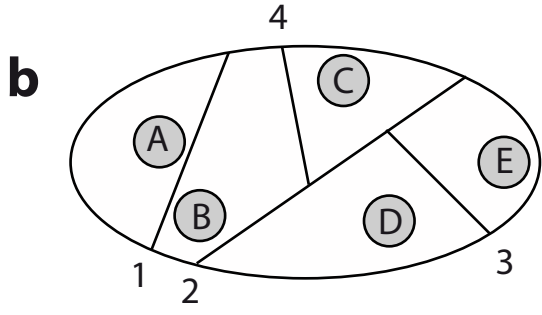
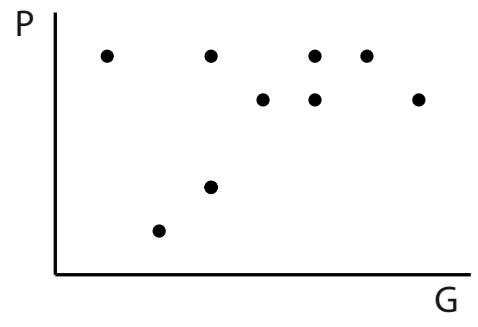
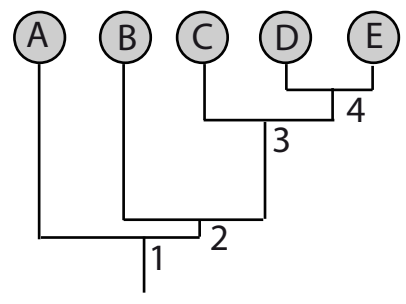
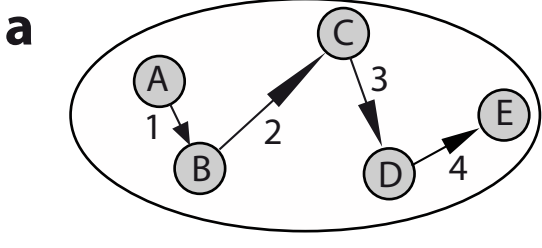
Electronic Appendixes

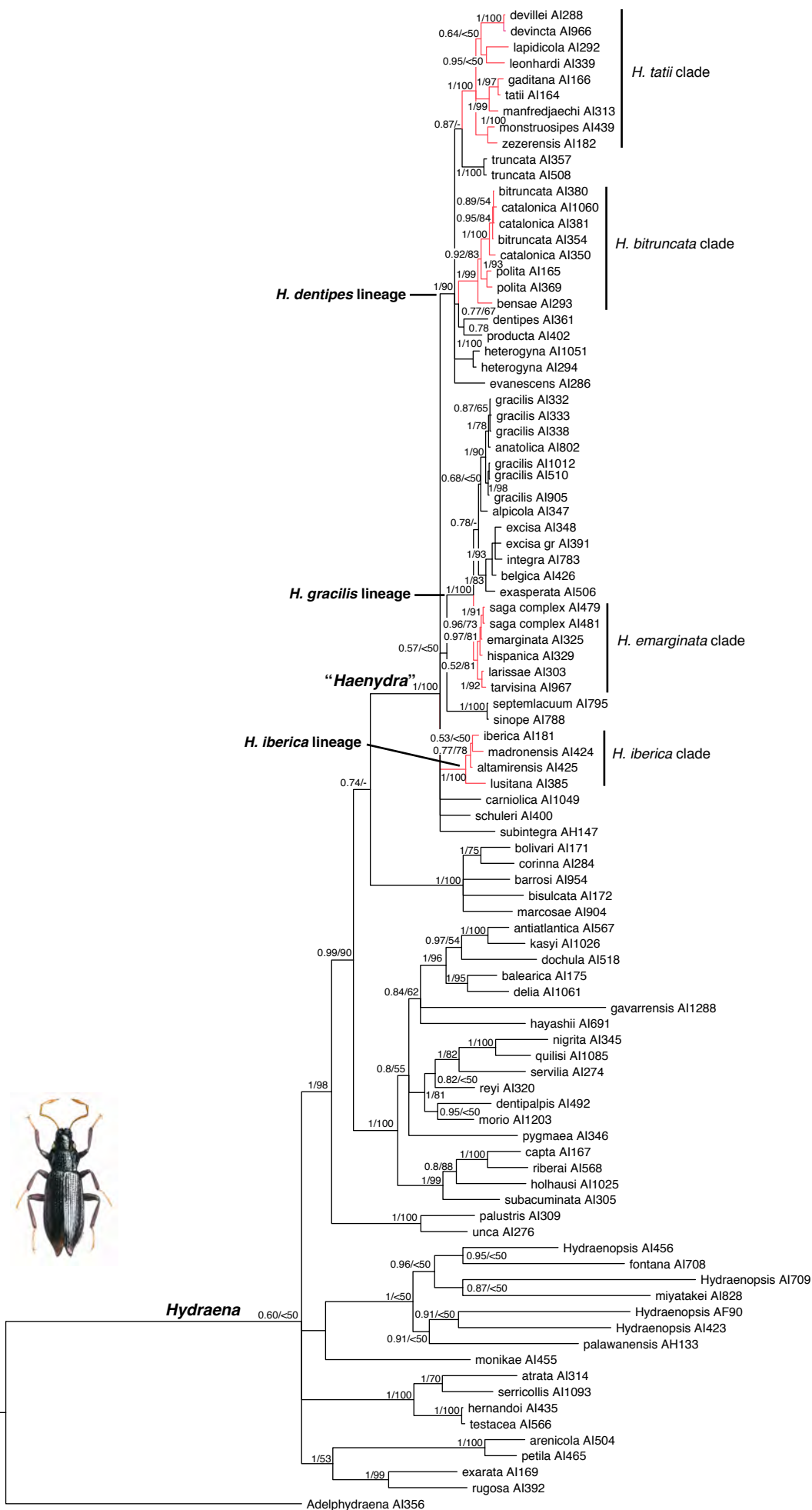
Appendix S1

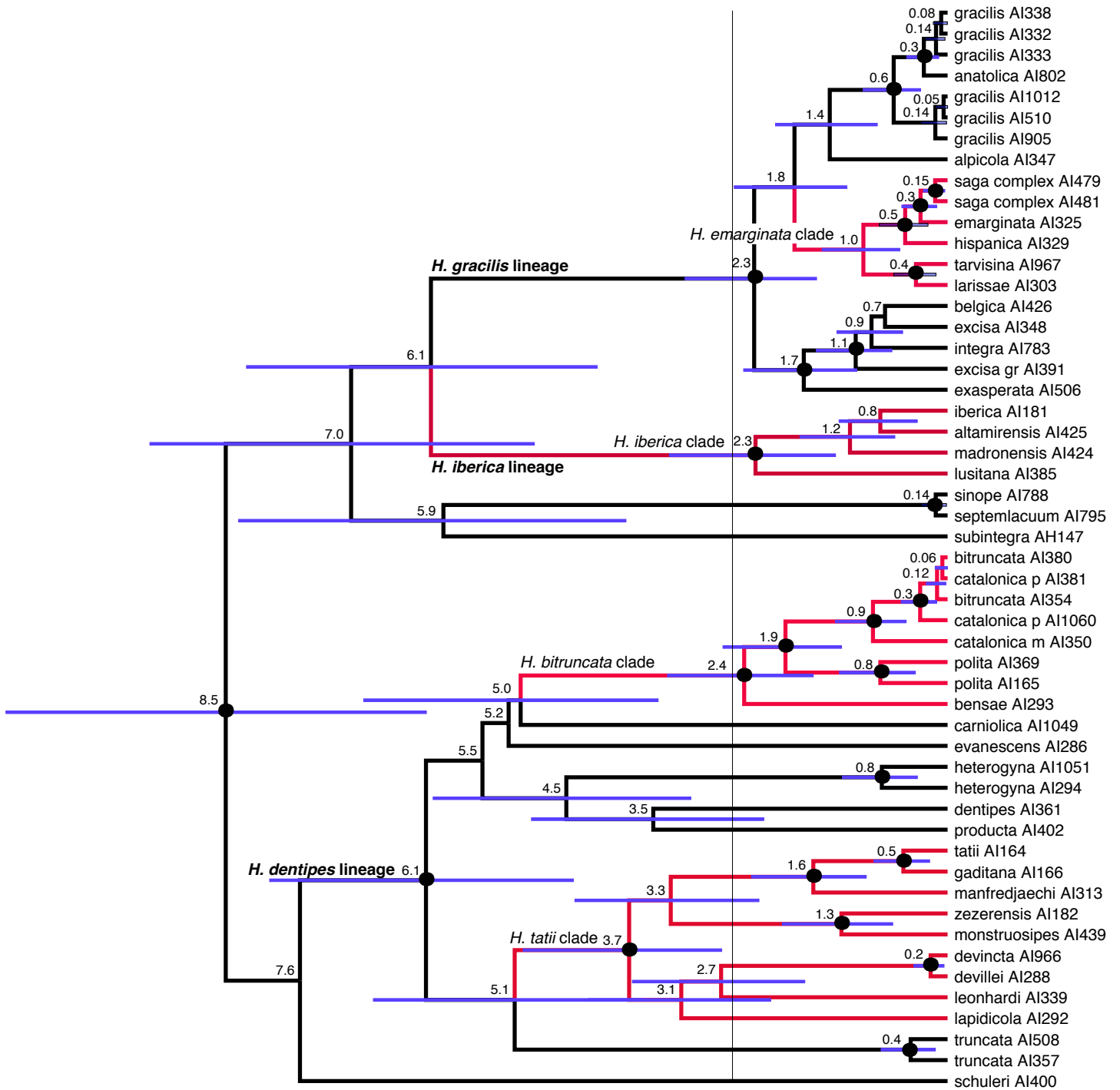
Specimens used in the study, with locality, collector, voucher reference numbers and accession numbers for the sequences. Specimens “a” to “e” of *H. iberica* were sequenced to check for intraspecific variability and not included in the main analyses (see text). Nomenclature follows Hansen (1998) and Jäch (2004).

Appendix S2

List of primers used for sequencing.

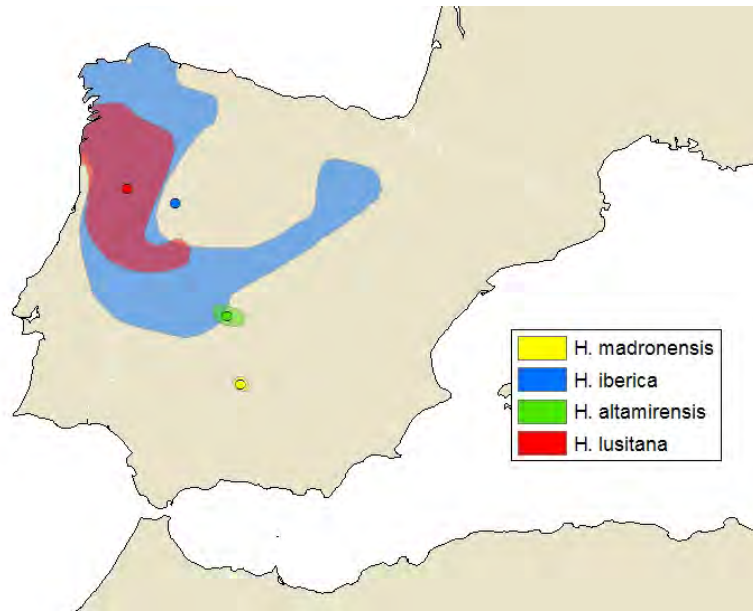




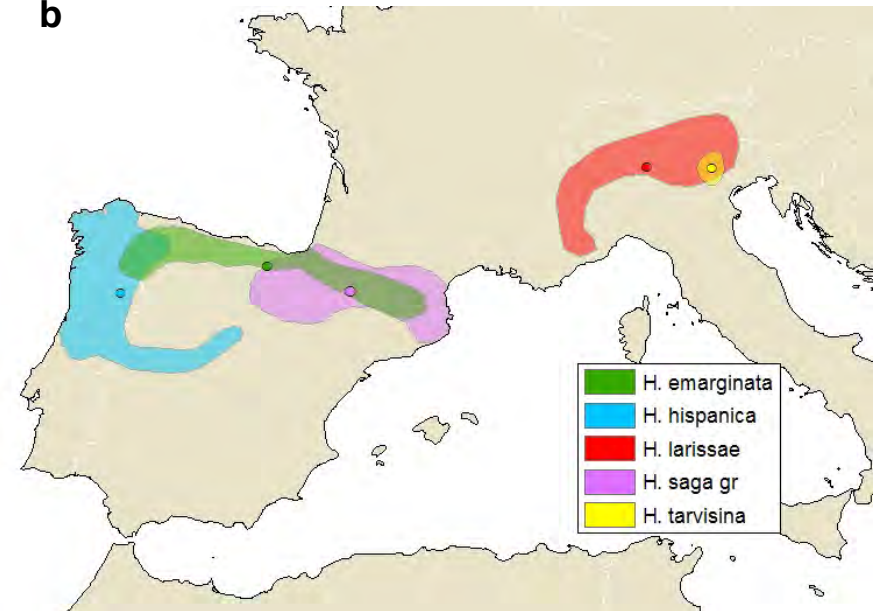


1.0 MY

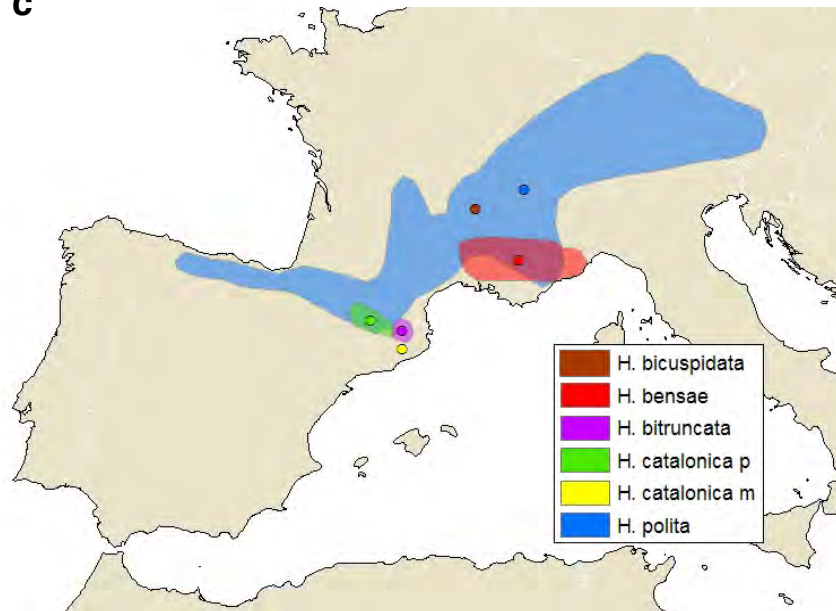
a



b



c



d

