Molecular data and ecological niche modelling reveal the evolutionary history of the common and Iberian moles (Talpidae) in Europe

Violaine Nicolas, Jessica Martínez-Vargas, Jean-Pierre Hugot

Violaine NICOLAS. Institut de Systématique, Evolution, Biodiversité, ISYEB UMR 7205 – CNRS, MNHN, UPMC, EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, CP51, 55 rue buffon, Paris, France.

Tel: (+33) 1 40 79 48 52

vnicolas@mnhn.fr

Title: Molecular data and ecological niche modeling reveal the evolutionary history of the common and Iberian blind moles (Talpidae) in Europe

Violaine NICOLAS[1], Jessica MARTÍNEZ-VARGAS[2], Jean-Pierre HUGOT[1]

[1] Institut de Systématique, Evolution, Biodiversité, ISYEB UMR 7205 – CNRS, MNHN, UPMC, EPHE, Muséum National d'Histoire Naturelle, Sorbonne Universités, Paris, France.

[2] Departament de Biologia Animal, de Biologia Vegetal i d'Ecologia, Facultat de Biociències, Universitat Autònoma de Barcelona, E-08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain

Short running title: Phylogeography of moles in western Europe

Nicolas *et al.*

Abstract

According to mitochondrial data, the common mole, *Talpa europaea*, is paraphyletic. This could be explained either by an ancient introgression of mtDNA from the Iberian blind mole *T. occidentalis* to *T. europaea*, or the existence of a differentiated taxonomic entity in northern Spain that needs to be described. In this study we combined mitochondrial (Cytb) and nuclear (HDAC2) data to investigate these two alternative hypotheses. Based on both mitochondrial and nuclear data and an extensive geographical sampling (399 sequenced individuals) we show that the populations of *T. europaea* from Spain and southwestern France (south of the Loire River) are phylogenetically closer to *T. occidentalis* than to *T. europaea.* The Spanish-French lineage has some morphological characters resembling more to *T. occidentalis* (e.g. eyes) and others resembling more to *T. europaea* (external measurements, mesostyle of the first upper molar). It also seems to have several distinctive dental characters, suggesting that it should be recognized as a new species. Within the three lineages we found a marked phylogeographical pattern, with several allopatric or parapatric lineages, dating from the Pleistocene. Our genetic data combined with species distribution models support the presence of several putative glacial refugia during glacial maxima for each species.

Introduction

Moles of the genus *Talpa* are strictly subterranean mammals and are widely distributed throughout the western Palearctic region, from the Iberian Peninsula to Siberia (Hutterer 2005). Nine species were recognized in the last revision of the genus (Hutterer 2005): the Altai mole *T. altaica* Nikolasky, 1883, the Caucasian mole *T. caucasica* Satunin, 1908, the Levant mole *T. levantis* Thomas, 1906, the Père David's mole *T. davidiana* Milne-Edwards, 1884, the blind mole *T. caeca* Savi, 1882, the common mole *T. europaea* Linnaeus, 1758, the Iberian blind mole *T. occidentalis* Cabrera, 1907, the

Roman mole *T. romana* Thomas, 1965, and the Balkan mole *T. stankovici* Martino and Martino, 1931. Based on genetic data, Bannikova *et al*. (2015) recently recognized three additional species: *T. talyschensis* Vereschagin, 1945, *T. ognevi* Stroganov, 1948, and *Talpa ex gr. levantis*. DNA data support a monophyletic origin of the genus (Colangelo *et al.* 2010; Bannikova *et al.* 2015). Interestingly, all *Talpa* species, except *T. europaea*, are endemic to small geographical regions (Mitchell-Jones *et al.* 1999; Hutterer 2005). Until recently, the common mole was thought to be widespread across Europe, from the Ebro River in Spain to the Ob and Irtish Rivers in Russia (Hutterer 2005). Despite the abundance of the common mole in its natural habitats, little data are available on its morphologic or genetic variability. Up to 26 forms were described but only three subspecies of *T. europaea* were retained by Hutterer (2005): *europaea*, *cinerea* and *velessiensis*. Loy & Corti (1996) showed a wide variation, related both to geography and climate, in the morphology of the mandible of *T. europaea* throughout Europe. They also showed distinct morphologies in peripheral populations (Italian, Spanish and British) with respect to the rest of Europe (Loy & Corti 1996). Investigations of the evolutionary history of *T. europaea* through mitochondrial DNA (mtDNA) revealed three differentiated mtDNA lineages: one restricted to Spain, one restricted to Italy and the third one widespread across Europe (Feuda *et al.* 2015). The Spanish lineage forms a monophyletic lineage with *T. occidentalis*, suggesting the paraphyly of *T. europaea*. Feuda *et al*. (2015) proposed two different scenarios to explain this result:

1) An ancient introgression of mtDNA from *T. occidentalis* to *T. europaea* during some phases of the late Pliocene;

2) The existence of a differentiated taxonomic entity in northern Spain that needs to be described.

In this study we analyze both mitochondrial (mtDNA) and nuclear (nDNA) data to tease apart these two hypotheses. If the paraphyly of *T. europaea* is only supported by mtDNA, then an ancient mitochondrial introgression is likely. If it is supported by both mtDNA and nDNA data, then the existence of an undescribed taxonomic entity in northern Spain is likely. We also combined genetic data and ecological niche modeling to make inferences about the evolutionary history of these lineages.

Material and methods

Sampling

We extensively sampled moles in France (270 individuals of *T. europaea*) and Spain (six *T. europaea*, 22 *T. occidentalis*). We also added one and two specimens of *T. europaea* from Netherlands and Switzerland, respectively. These specimens were assembled from recent captures in the field and from museum collections (See Appendix S1 in Supporting Information, and Fig. 1). Tissue samples from fresh material were stored in 96% ethanol.

We took the body measurements (weight, head and body length, tail length, hind foot length) of 222 field captured *T. europaea* (see specimens list in Appendix S1).

Laboratory techniques

We extracted DNA using the NucleoSpin Tissue Core kit (Macherey-Nagel, Hoerdt, France). We sequenced one mtDNA gene, the Cytochrome b (Cytb), and one nDNA gene, the intron 10 of the histone deacetylase 2 (HDAC2). These two genes were previously used to study the phylogeography and evolutionary history of several taxa of Soricomorpha (e.g. Colangelo *et al.* 2010; He *et al.* 2014; Jacquet *et al.* 2014; Bannikova *et al.* 2015; Jacquet *et al.* 2015). The Cytb gene was amplified and sequenced for 298 individuals using polymerase chain reaction (PCR) primers L14723, H15915 (Ducroz *et al.* 2001), TALPF2 (ACYGCATTCATAGGGTACGT), TALPF3 (CAAACCCRCTAAACACACCA), TALPR2 (ACGTACCCTATGAATGCRGT) and TALPR3 (TGGTGTGTTTAGYGGGTTTG; this study). The internal primers designed in this study were used to amplify the DNA of museum specimens, for which DNA was degraded. The HDAC2 gene was amplified and sequenced for 95 individuals using primers HDAC2-EX10U and HDAC2-EX11L (Hassanin *et al.* 2013). The PCR consisted of 35 cycles: 30 s at 94°C, 40 s at 52-58°C (Cytb: 52°C, HDAC2: 58°C), and

90 s at 72°C. The double-stranded PCR products were purified and sequenced at Genoscope (Ivry Sur Seine, France) using the Sanger di-deoxy sequencing run on an ABI 3730XL sequencer (Applied Biosystems, Foster City, California). All sequences were submitted to the GenBank database (Appendix S1, KU189427 to KU189724 for Cytb, KU189332 to KU189426 for HDAC2). We obtained at least partial mtDNA sequences for 298 individuals (276 *T. europaea* and 22 *T. occidentalis*) and nDNA sequences for 95 individuals (76 *T. europaea* and 19 *T. occidentalis*).

Phylogenetic analyses

We included in our Cytb phylogenetic analyses all the newly sequenced specimens (298 individuals) and all specimens of *T. europaea* and *T. occidentalis* available in the GenBank database (98 individuals). We estimated evolutionary relationships among distinct Cytb haplotypes by constructing phylogenetic trees using maximum-likelihood (ML) and Bayesian (BA) inference. We used two sequences of *T. romana* (GenBank FN640563 and FN640564) and two sequences of *T. caeca* (GenBank FN640558 and FN640559) as outgroups, because they are the sister taxa of *T. europaea* and *T. occidentalis* (Colangelo *et al.* 2010; Bannikova *et al.* 2015). The computer program MrModeltest ver. 2 (Nylander 2004) was used to evaluate the fit of 24 nested models of nucleotide substitution to the data. The model (GTR+I+G) chosen by MrModeltest according to the Akaike information criterion was then used in ML and BA analysis. The ML analysis was performed with PHYML ver. 3.0 (Guindon *et al*. 2010) on the PHYML online Web server (Guindon *et al.* 2005). In the PHYML procedure, we used the BIONJ distance-based tree as the starting tree and 100 bootstrap replicates. Bayesian inference was performed with MrBayes ver. 3.2.5 (Ronquist *et al.* 2012) using 20 million generations (burnin = 25%), sampling every 1000 steps. Convergence between chains was checked by manual inspection using the “sump” command and by using the mean SD of split frequencies, which was below 0.01 in our runs, as an indicator of convergence. Only specimens for which complete Cytb sequences (1140 bp, no uncertainty) were available were retained in our phylogenetic analyses (i.e. 360 specimens). To get the best possible picture of the geographical distribution of each

genetic lineage, sequences from an additional 36 individuals (mtDNA sequences shorter than 1140 bp or with uncertainties) were aligned with the long sequences and unambiguously assigned to particular lineages by a neighbor joining analysis (K2P genetic distances) using PAUP ver. 4b10 (Swofford 2000; see Appendix S1 for more details on specimens included in each analysis).

The relationships among Cytb sequences were also investigated by constructing a network using the median-joining (MJ) method available in NETWORK v4.500 (Bandelt *et al.* 1999). This method accounts for the coexistence of ancestral and descendent haplotypes, multifurcations and reticulate relationships (Posada & Crandall 2001), and it is therefore suitable for studying intraspecific relationships. One network was constructed for each lineage identified in the ML and BA analyses.

Due to the low number of mutations and the lack of suitable outgroup data, phylogenetic relationships between nDNA HDAC2 sequences were investigated by constructing a network using the MJ method. The phase of each heterozygous haplotype and its reconstruction were carried out using PHASE ver. 2.1.1 (Stephens *et al.* 2001; Stephens & Scheet 2005). 190 sequences (corresponding to 95 specimens) of 668 bp were used in this analysis (Appendix S1).

Molecular diversity

Average percentages of pairwise differences (Kimura 2-parameters, K2P) between Cytb lineages were calculated in PAUP ver. 4b10 (Swofford 2000). Several analyses were carried out to test the levels of genetic variation and genetic differentiation between lineages. The number of haplotypes, number of polymorphic sites, average number of nucleotide differences, nucleotide diversity, and haplotype diversity were calculated for each lineage using DnaSP ver. 5.10 (Librado & Rozas 2009). To compare haplotype richness between lineages while controlling for unequal sample sizes (Leberg 2002) we used a rarefaction analysis (Analytic Rarefaction ver. 1.4; available from Steven Holland, UGA Stratigraphy Lab website; http://www.uga.edu/~strata/software/). We randomly subsampled haplotypes (size of subsample = N of the smaller sample) with replacement 10,000 times to estimate the number of haplotypes that would occur in the smaller sample. We compared the distribution of the subsamples with the number of

haplotypes found in the smaller sample. P-values were calculated based on the number of times in 10,000 subsamples that as many or more haplotypes were found in the larger sample compared to the smaller.

Time of divergence

We used Bayes factor (BF) to select the molecular clock prior (i.e. strict clock versus log-normal relaxed clock). Prior and calibration points follow Feuda *et al*. (2015). The marginal likelihood of the two hypotheses was estimated using stepping-stone sampling as implemented in MrBayes ver. 3.2.5 using 25 million generations. The BF value of 7776 provided strong evidence for the strict clock model (Kass & Raftery 1995), in agreement with the results of Feuda *et al*. (2015). Thus we obtained an approximate estimation of the divergence among the main lineages using the formula $T = d/2\mu$; where T is the time of divergence, d is the genetic distance between groups and μ is the substitution rate in millions of years. We calculated the uncorrected distance and used a substitution rate for the Cytb gene of 0.01407 per site per lineage per million years (Feuda *et al.* 2015).

Species distribution modeling (SDM)

To evaluate current and past habitat suitability for the three main lineages of moles detected in our sample (*T. europaea*, *T. occidentalis* and the Spanish-French lineage) we produced distribution models under two different climatic conditions (current and Last Glacial Maximum [LGM]) using a distribution modeling approach. For *T. europaea* and the Spanish-French lineage, we retrieved occurrence data from Feuda *et al*. (2015). In the study of Feuda *et al*. (2015) these two lineages were considered together and called "*T. europaea*". In our study we consider them as distinct, and attributed to *T. europaea* all occurrence data north and east of the Loire River, and to the Spanish-French lineage all occurrence data south and west of the Loire River (see Supporting Information, Appendix S2). For *T. occidentalis*, the majority of occurrence data were derived from

the online database « Global Biodiversity Information Facility » (http://www.gbif.org) and from the literature (Miller 1912, Appendix S2). To correct sampling bias, a subsample of records regularly distributed in the geographical space was selected using ENMTools ver. 1.3 (Warren *et al.* 2010). This method is the most efficient at correcting sampling bias (Fourcade *et al.* 2014). The resolution of the reference grid was set to 2.5 arc min; i.e. the same as that used for the environmental layers.

As environmental layers, we used available climatic data from the Worldclim database (Hijmans *et al.* 2005). Among the 19 BioClim climatic variables, for our analyses we retained 11 variables that are considered ecologically important for moles and for which pairwise correlations were mostly smaller than 0.75 (Feuda *et al.* 2015): annual mean temperature (BIO1), mean diurnal range (BIO2), isothermality (BIO3), annual range in temperature (BIO7), mean temperature of the driest quarter of the year (BIO9), mean temperature of the warmest quarter of the year (BIO10), precipitation of the driest month (BIO14), precipitation seasonality (BIO15), precipitation of the wettest quarter of the year (BIO16), precipitation of the warmest quarter of the year (BIO18), and precipitation of the coldest quarter of the year (BIO19). These climatic variables are ecologically important for moles because temperature and precipitation affect soil hardness (through drought, freezing and flooding), which conditions the ability of moles to dig, and the availability of feeding resources represented by soil invertebrates (Gorman & Stone 1990). Climatic variables were downloaded for present conditions and for the LGM. LGM climate data were drawn from General Circulation Model (GCM) simulations from two climate models: the Community Climate System Model (CCSM ver. 3) (Collins *et al.* 2006) and the Model for Interdisciplinary Research on Climate (MIROC ver. 3.2) (Hasumi & Emori 2004). These climate models differ in temperature and precipitation patterns. LGM climate as simulated by CCSM3 is colder and dryer than that of MIROC over Europe. For a discussion of the uncertainties associated with the climatic data see Schorr *et al.* (2012) and Varela *et al.* (2015). The use of these two different climate models enabled us to assess and account for modeling uncertainty due to LGM climate data. To predict the potential distribution of the species in current conditions and in the LGM we used Maxent ver. 3.3.3 (Phillips *et al.* 2006), which outputs a model with relative occurrence probability of a species within the grid cells of the study area. To ensure consistency of model predictions among repeated

runs, we performed a 50-fold cross-validation with random seed. To determine whether the predictions for current conditions generated by Maxent were better than random predictions we used the area under the receiver operating characteristic curve (AUC), a commonly used measurement for comparison of model performance (Elith *et al.* 2006). The AUC ranges from 0 to 1, with greater scores indicating better discrimination ability; an AUC greater than 0.5 indicates that the model discriminates better than random.

Results

mtDNA

Our Cytb phylogenetic analyses show that *T. europaea* includes three main lineages (European, Italian and Spanish-French), and is paraphyletic because the Spanish-French lineage is more closely related to *T. occidentalis* than to *T. europaea* (Fig. 2). The genetic divergence between the Spanish-French and the other three lineages are: 8.51% ± 0.27 (European), 8.39% ± 0.33 (Italian) and 7.47% ± 0.39 (*T. occidentalis)*, respectively. The time of divergence between the Spanish-French lineage and *T. occidentalis* is estimated at 2.47 ± 0.12 Million years ago (Mya). Between the European+Italian lineage and the Spanish-French+*T. occidentalis* lineage it is estimated at 2.82 ± 0.10 Mya.

The Spanish-French lineage includes all specimens from Spain and all specimens from southwestern France (see Table S1 for detailed geographical localities), with the exception of two specimens from Mosset (Pyrénées-Orientales; Fig. 1). In France this lineage was not found northward and eastward of the Loire River.

The two other *T. europaea* lineages (European and Italian) are monophyletic (Fig. 2). The Italian lineage includes all Italian specimens, with the exception of one specimen from Bolzano, and one specimen from the southern part of Switzerland (Meride, at the border with Italy; Fig. 1). The European lineage includes all specimens from Bosnia-Herzegovina, Denmark, Germany, Greece, Hungary, Netherlands, Russia, Switzerland,

Sweden, Ukraine, United Kingdom, Turkey, and all the French specimens captured northward and eastward of the Loire River, plus two specimens from the Pyrenees. The genetic divergence between the Italian and the European lineage is 1.91% ± 0.33, which corresponds to a divergence time of 0.667 ± 0.115 Mya.

Nucleotide diversity and the average number of nucleotide differences in the European lineage are half as high as in the three other lineages; they are maximal in *T. occidentalis* (Tab. 1). Taking into account sample size, the haplotype richness is significantly lower ($P<0.05$) in *T. occidentalis* than in the three *T. europaea* lineages. Haplotype richness is also significantly lower in the Spanish-French lineage than in the European lineage.

Within the European lineage, the two haplotypes from Bosnia-Herzegovina cluster together in the Network analysis (light-blue sub-lineage in Fig. 3A; see Appendices S1 and S3 for haplotype numbers and geographical origin of each haplotype). The same is true for all but three haplotypes from Ukraine (7 specimens; brown sub-lineage in Fig. 3A). The haplotype from Hungary (1 specimen; black haplotype in Fig. 3A) differs by 5 mutations from all other haplotypes. All haplotypes from France, western Switzerland, Germany, Denmark, Sweden and Russia cluster together (red sub-lineage in Fig. 3A), except for 4 haplotypes from southeastern France which form a distinct lineage (grey sub-lineage in Fig. 3A). The specimen from Italy is genetically divergent from all other haplotypes (in dark-blue in Fig. 3A). All divergence events between these sub-lineages occurred between 0.283 ± 0.039 and 0.712 ± 0.058 Mya.

Within the Italian lineage, geographical clustering of haplotypes is evident. All specimens from central and eastern Italy cluster together (red sub-lineage in Fig. 3B). Haplotypes from Sondalo (7 specimens, 5 haplotypes) and Modena (2 specimens, one haplotype) are close to one another (blue sub-lineage in Fig. 3B). The two haplotypes from Como (1 specimen per haplotype) are grouped together (black sub-lineage in Fig. 3B). Divergence events between these sub-lineages occurred between 0.330 ± 0.069 and 0.580 ± 0.089 Mya.

Within the Spanish-French lineage, four parapatric and/or sympatric sub-lineages can be identified: one in northwestern Spain (red sub-lineage in Fig. 3C), one in northern Spain

and southern France up to the Gironde estuary (blue sub-lineage in Fig. 3C), one in northeastern Spain and in France up to the Loire River (black sub-lineage in Fig. 3C), and one in France southwestward of the Loire River and northward of the Gironde estuary (grey sub-lineage in Fig. 3C). All divergence events between these sub-lineages occurred between 0.349 ± 0.050 and 1.014 ± 0.056 Mya.

Within the *T. occidentalis* lineage, three sub-lineages can be identified: one grouping the two specimens from Portugal and northwestern Spain (red sub-lineage in Fig. 3D), one grouping the specimens from southern Spain (grey sub-lineage in Fig. 3D) and one grouping the specimens from northern Spain (black sub-lineage in Fig. 3D). The time of divergence between the northern Spain sub-lineage and the two others sub-lineages is 0.825 ± 0.093 Mya. Instead, it is of 1.169 ± 0.016 Mya between southern and western sub-lineages.

nDNA

Eight distinct haplotypes and 18 polymorphic sites were identified in the 190 HDAC2 sequences. In the MJ network of HDAC2 sequences (Fig. 4), the 38 sequences of *T. occidentalis* cluster together in one haplotype. Within *T. europaea* two groups of haplotypes can be identified: one grouping all sequences attributed to the European lineage in the mtDNA analyses (60 sequences), and one grouping all sequences attributed to the Spanish-French lineage in the mtDNA analyses (92 sequences). Unfortunately no specimen of the Italian lineage could be included in this analysis, since only Cytb sequences are available in Genbank and we had no tissue sample from this geographical origin. In agreement with mtDNA results, the genetic divergence is greater between the Spanish-French lineage and the European lineage (more than 12 mutations between haplotypes of these two lineages) than between the Spanish-French lineage and the *T. occidentalis* lineage (1 to 3 mutations). Haplotypes of the European lineage differ by more than 11 mutations from haplotypes of the *T. occidentalis* lineage. Three distinct haplotypes differing by less than two mutations were identified in the European lineage, and four distinct haplotypes differing by less than three mutations

were identified in the Spanish-French lineage.

Species distribution modeling

To study the relationships between genetic diversity and possible glacial refugia, we built a species distribution model (Fig. 5) for the three main lineages observed in our phylogenetic analyses (*T. europaea*, the Spanish-French lineage and *T. occidentalis*) based on the known-presence localities of these taxa. For the three taxa, the AUC value was higher than 0.987, a value considered to correspond to a useful predictive model (Phillips & Dudík 2008).

For *T. europaea*, the model revealed high habitat suitability for this species across most of Europe (except southwestern France and the Alps Mountains) under present-day bioclimatic conditions, in agreement with its current known distribution. The projection of the model to the conditions of the LGM returned lower habitat suitability in Europe with respect to the present-day. According to the MIROC data, suitable habitat occurred in France, with two disconnected patches, and to a lesser extent in the Italian Peninsula, the Balkans and near the Black Sea. According to the (colder) CCSM climate data, few suitable areas for *T. europaea* can be identified, and they occurred in western France, the Italian Peninsula and near the Black Sea.

For the Spanish-French lineage, the model revealed high habitat suitability across northern Spain and most of France (except the northeast) under present-day bioclimatic conditions. This area is larger than the current known distribution of this lineage, as it is not found northward and eastward of the Loire River. According to the MIROC data, during the LGM suitable habitat for this taxon occurred in most on this area; while the CCSM data reveals very small suitable habitat patches (in western and southeastern France).

For *T. occidentalis*, the model revealed high habitat suitability for this species across most of the Iberian Peninsula under present-day bioclimatic conditions, in agreement with the current known distribution of the species. The MIROC model predicted very mild climatic conditions for the LGM in this part of the world, and thus the distribution

predicted for *T. occidentalis* during the LGM is very similar to the present distribution. The relatively high abundance of many cold-adapted species (reindeer, woolly rhinoceros and mammoth) in northern Spain (provinces of Asturias, Cantabria, Bizkaia, Gipuzkoa, Navarra and Girona) during the LGM indicates the presence of steppe environments through this period (Alvarez-Lao & Garcia 2010). This result is not congruent with the climatic conditions predicted by the MIROC model for the Iberian Peninsula. According to the CCSM model, suitable habitat for *T. occidentalis* during the LGM was patchily localized in Catalonia, Andalusia, Cantabria and Portugal.

Discussion

Taxonomic implications

Our mtDNA results are congruent with those of Feuda *et al*. (2015) and show that the species *T. europaea*, as currently recognized, is paraphyletic. Feuda *et al*. (2015) proposed two different scenarios to explain this result:

1) An ancient introgression of mtDNA from *T. occidentalis* to *T. europaea* during some phases of the late Pliocene,
2) The existence of a differentiated taxonomic entity in northern Spain that needs to be described.

Our nDNA data are not congruent with the scenario of the ancient introgression of mtDNA since the paraphyly of *T. europaea* is also supported by the HDAC2 nDNA gene. Thus, the populations of *T. europaea* from Spain and southwestern France should either be considered to belong to the species *T. occidentalis* or to a new taxonomic entity that needs to be described. To decide between these two options it is useful to compare the morphology of these two populations: they can be considered as distinct taxonomic entities if diagnostic characters are found between them.

The taxonomy of the genus *Talpa* remains puzzling due to the uniformity of the species'

external morphology linked to their convergent adaptation to underground life (Niethammer & Krapp 1990; Kryštufek & Vohralík 2001). Karyological techniques can be useful to distinguish some species, but this is not the case of *T. europaea* and *T. occidentalis* (Capanna 1981; Jiménez *et al.* 1984; Loy *et al.* 2005; Colangelo *et al.* 2010). Allozymic and mtDNA data analyses proved to be efficient tools to identify species (Filippucci *et al.* 1987; Loy *et al.* 2001; Tryfonopoulos *et al.* 2009; Colangelo *et al.* 2010; Bannikova *et al.* 2015). According to Colangelo *et al.* (2010), the Cytb K2P genetic distance between *Talpa* species varies from 8.6% to 15.6%. These values are similar to those found between the Spanish-French lineage and the two other *T. europaea* lineages (> 8.4%), and close to that found between the Spanish-French lineage and *T. occidentalis* (7.47%), suggesting a species status of the Spanish-French lineage. However, using mtDNA genetic distances to assign taxonomic rank (Johns & Avise 1998; Baker & Bradley 2006) is problematic, and is not without its detractors (Ferguson 2002). Morphologic and/or morphometric analyses have been the most widely used techniques for the identification of mole species and the clarification of their taxonomic status (Corti & Loy 1987; Kryštufek 1987, 1994, 2001; Loy *et al.* 2001; Kryštufek & Benda 2002). *Talpa europaea* is distinguishable from *T. occidentalis* by having open eyes. In *T. occidentalis* the eyelids are fused together and completely covered by membranes (Petrov 1971; Niethammer & Krapp 1990; Witte 1997; Aulagnier *et al.* 2008). All specimens from the Spanish-French lineage have their eyelids fused together, as in *T. occidentalis*. *Talpa europaea* tends to have larger body size and condylobasal length (weight = 36-130 g, head and body = 100-165 mm, tail = 20-51 mm, foot = 16-25 mm, condylobasal length = 33-37 mm) than *T. occidentalis* (weight = 30-70 g, head and body = 90-135 mm, tail = 16-35 mm, foot = 14-18 mm, condylobasal length = 29-32 mm) (Miller 1912; Jiménez *et al.* 1984; Niethammer & Krapp 1990; Aulagnier *et al.* 2008). As previous studies included within *T. europaea* specimens from both the true *T. europaea* lineage and specimens from the Spanish-French lineage, we compared the values obtained for these two lineages based on 222 genetically identified specimens (133 specimens belonging to the *T. europaea* lineage and 89 specimens belonging to the Spanish-French lineage; Appendix S1). Weight, head and body length, and foot length were significantly greater in the Spanish-French lineage (weight = 89 ± 17 g, head and body = 149 ± 7 mm, foot = 21.5 ± 1.5 mm) than

in the *T. europaea* lineage (weight = 76 ± 12 g, head and body = 144 ± 8 mm, foot = 20.9 ± 1.6 mm; $P < 0.02$ for the three measurement; Student t tests). Measurements were not significantly different between the two lineages for tail length (27 ± 3 mm and 26 ± 4 mm for the Spanish-French lineage and *T. europaea* lineage, respectively; $P = 0.758$).

The last character commonly used to differentiate *T. europaea* from *T. occidentalis* is the mesostyle of the upper first molar (M1), which is simple in *T. europaea* and double in *T. occidentalis* (Miller 1912; Capanna 1981; Jiménez *et al.* 1984; Niethammer & Krapp 1990; Cleef-Roders & Hoek Ostende 2001; Aulagnier *et al.* 2008). All specimens from the Spanish-French lineage have a simple M1 mesostyle, alike *T. europaea* (Fig. 6). As this character is dependent on wear, we only analyzed a subset of specimens with unworn or little-worn molars (*T. europaea*: n=66; Spanish-French lineage: n=57; *T. occidentalis*: n=51). We also found that the mesostyle condition of the upper second molar (M2) and the upper third molar (M3) generally differs between the Spanish-French lineage and the two other species. In *T. occidentalis* and *T. europaea*, the mesostyle of these two molars is divided into two cusps (Figs. 6A and 6B). The two cusps have subequal size, they are aligned on a plane that extends parallel to the parastyle and metastyle. Some specimens from the Spanish-French lineage seem to have a simple mesostyle in M2 (n = 13/57; Fig. 6C). Other specimens show an additional minute cusp (n = 44/57; Fig. 6D). Unlike in *T. europaea* and *T. occidentalis*, this cusp is much smaller than the main cusp of the mesostyle and is located in the crest that connects the mesostyle to the metacone of M2, i.e. in a more lingual position than the mesostyle itself. In the M3 of some specimens from the Spanish-French lineage, the mesostyle is composed of a main anterior cusp and a slightly smaller posterior cusp (n = 35/57; Fig. 6C). In other specimens this posterior cusp is not clearly discernible because its posterior border is fused to the crest that runs from mesostyle to the metacone of M3 (n = 22/57; Fig. 6D). Despite inter-individual variability, the mesostyle condition of M2 and M3 of the specimens from the Spanish-French lineage differs from that found in *T. europaea* and *T. occidentalis*.

To conclude, the Spanish-French lineage has a somewhat intermediate morphology between *T. europaea* and *T. occidentalis*, with some characters resembling more to *T. occidentalis* (e.g. eyes) and others more to *T. europaea* (e.g. external measurements and

mesostyle of M1). It also seems to have several distinctive characters (e.g. mesostyle of M2 and M3).These results, combined with our genetic data, suggest that the Spanish-French lineage should be recognized as a new species. Additional in-depth morphological analyses and comparison with type specimens are necessary to formally describe it.

Geographical distribution of T. europaea *and the Spanish-French lineage*

Based on our genetic data, the Spanish-French lineage and the true *T. europaea* are allopatric and each is distributed in one of the sides of the Loire River (with the exception of two specimens of *T. europaea* captured in the geographic range of the Spanish-French lineage; Fig. 1). A similar pattern is observed in the semi-fossorial rodent *Microtus arvalis;* two lineages occurring in western Europe on opposite sides of the Loire River (Tougard *et al.* 2008). When we performed our SDM analyses we did not include within the geographic range of *T. europaea* the two specimens captured in the geographic range of the Spanish-French lineage (specimens from Mosset in the Pyrenees). However, our results show that the southern part of France, from the border of the Alps Mountains to the eastern part of the Pyrenees Mountains, is suitable for this species (Fig. 5). Thus, it is not surprising to find both species in Mosset. According to our SDM approach, *T. europaea* should also be captured around the Gironde estuary. Despite extensive sampling, we only captured the Spanish-French lineage in this area. For the Spanish-French lineage, the climatic model revealed high habitat suitability across northern Spain and most of France (except the northeast) under present-day bioclimatic conditions. This area is larger than the current known distribution of the species. Thus, factors other than climatic ones probably explain the geographic distribution of these two species. Our SDM do not take into account species interactions (e.g. competition, predation, parasitism, mutualism) and this is important since realized versus fundamental niches often differ (Pearman *et al.* 2008; Sinclair *et al*. 2010; Giannini *et al*. 2013). Mutual competitive exclusion between the two mole species may explain their geographical distribution. In other subterranean mammals, like pocket gopher, interspecific differences in body size and digging strategy have been shown to

confer competitive dominance of one species over another depending on soil characteristics (Marcy *et al.* 2013). In regions where divergent soil types co-occur, the ranges of different pocket gopher species can overlap (Thaeler 1968). A similar result was found in the Japanese moles, where the dominant mole species, *Mogura wogura*, is progressively expanding its range northwards, displacing the small inferior species *M. imaizumii.* Soil hardness was shown to affect the geographical distribution of these species and to allow their co-existence only under specific circumstances (Abe 2001). We showed significant body size differences between the two mole species. It would be interesting to further investigate how these differences in body size, and potentially other morphological differences, affect the digging behavior of these moles and their competitive dominance. More data on soil characteristics are also needed (percent soil clay, bulk density, shrink-swell capacity, soil hardness). This would allow us to test if interspecific differences in body size and digging strategy explain the mostly allopatric distribution of these two species, and their coexistence in the locality of Mosset (Pyrenees). It would also be interesting to test if body size vary geographically within each species according to habitat characteristics and/or competitive interactions between the two species, as previously shown for other mole species (Abe 1996; Loy & Capanna 1998). The Loire River is unlikely sufficient to explain the allopatric distribution of *T. europaea* and the Spanish-French lineage. Moles dig tunnels which may lie anywhere from 5 to 150 cm below the surface and they are known to be able to swim for 30-50 minutes, period during which they can cover distances of over a kilometer (Gorman & Stone 1990). Moreover, according to our genetic analyses, other big rivers like the Seine, Garonne or Rhin do not constitute effective barriers to geographical dispersal of moles. This hypothesis needs to be further tested by adequate geographical sampling. The complex Plio-Pleistocene history of this river (marine transgressions, creation of a system of stepped fluvial terraces) could also have contributed to maintain the separation of the two species (Cyprien *et al.* 2004; Nehlig 2010; Proust *et al.* 2010).

Phylogeographic history

The possible historical factors responsible for the genetic break between *T. occidentalis*

and the three *T. europaea* lineages were already discussed by Feuda *et al.* (2015). Because of our divergence time estimates are congruent with those of Feuda *et al.* (2015), we will not discuss these factors further. Due to a better numerical and geographical sampling, our study allows us for the first time to shed light on what evolutionary processes led to the current pattern of distribution of genetic variation in each of these lineages.

Subterranean mammals have low vagility, and they live in a fairly constant environment that is characterized by the absence of light and small fluctuations of temperature and humidity (Lacey *et al.* 2000). However, climatic oscillations leading to extreme drought or freezing, which dramatically increase hardness of soils and lower the availability of feeding resources represented by the soil fauna, can have a significant impact on moles (Gorman & Stone 1990). Thus, the Quaternary Period, which was dominated by Ice Ages and involved repeated global cooling and increasing of the Arctic and Antarctic ice sheets (Hewitt 2004), is expected to have left traces on the genetic variability of mole species. In agreement with this hypothesis we found a marked phylogeographical pattern within the four mole lineages studied, with several allopatric or parapatric lineages. Divergence events between sub-lineages vary but were all estimated to have occurred less than 1 Mya and more than 0.283 Mya. It seems likely that the populations started to diverge during earlier phases of the glacial periods and not during the last glaciation. As genetic lineages possess a history that extends over past glacial conditions, we should consider refugia as special areas that have persisted through several paleogeographic and climatic events (Médail & Diadema 2009). In the absence of available climatic data for previous glacial periods, we used the LGM conditions as a proxy of the paleoclimatic conditions of glacial periods. However, it should be stressed that the intensity of glaciations was not always the same during the last 1 My, and that even for the LGM several uncertainties are associated with the climate data (Schorr *et al.* 2012). In our data, uncertainty resulting from climate model is evident from the LGM species distributions obtained with the CCSM and MIROC models.

In *T. europaea* we found two main lineages: a European and an Italian one. The same phylogeographic structure was observed in several other mammal species (e.g. Grill *et al.* 2009) and is linked to the fact that the Italian Peninsula was one of the main refugial areas in southern Europe during the climatic oscillations of the Pleistocene (Hewitt

2004). In the semi-fossorial rodent *Microtus arvalis*, with a geographic range similar to that of the common mole, several genetic lineages that predated the LGM widely (0.475-0.086 Mya) were also found (Tougard *et al.* 2008). In both species, a distinct lineage was observed in the Italian Peninsula compared to western Europe, probably due to the Alpine barrier. However, we found one haplotype related to the European lineage in northern Italy, suggesting that genetic exchange between western Europe and northern Italy lineages was possible during interglacial periods, as previously found in *Myodes glareolus* (Colangelo *et al.* 2012). While in *T. europaea* we found one widespread lineage from France to Russia, in *Microtus arvalis* Tougard *et al*. (2008) found several lineages in the same area. According to our paleoecological model, central and eastern Europe were largely unsuitable for the common mole during glacial maxima. Extremely low temperatures for extended periods likely favored the establishment of permafrost, where soil hardness and scarcity of edaphic resources could have represented the main limiting factors for moles (Feuda *et al*. 2015). Our genetic data (red sub-lineage in Fig. 3A and distribution of haplotypes in Appendices S1 and S3) suggest that most of central and eastern Europe was probably recently colonized by the moles that would have survived in France during glacial maxima. However, southeastern localities from Ukraine, Bosnia-Herzegovina, Hungary and northern Italy showed a certain degree of differentiation (Fig. 3A). According to the paleoecological model, during the LGM suitable areas were still available in northern Italy, the Balkan Peninsula and near the Black Sea, which would have allowed several populations to survive in allopatry is this area. However, these populations would not have geographically spread over Europe when favorable climatic conditions returned. *Talpa europaea* specimens from the southeastern part of France form a distinct sub-lineage (grey sub-lineage in Fig. 3A) compared to the other French specimens (red sub-lineage in Fig. 3A), suggesting that at least two distinct refuge areas would have persisted in France. Our distribution model projected to the LGM-MIROC data conditions also suggests two disconnected areas with high probabilities of potential presence: one in the southeastern and central-eastern parts of France and one in southwestern and central-western parts of France. The second population would have then recolonized most of central and eastern Europe, while the first one would have remained restricted to a small area.

In the Italian lineage we found high haplotype diversity and three sub-lineages. The high diversity observed in northern and central-eastern Italy, i.e. in the Padano-Venetian plain, is congruent with other studies (Canestrelli *et al.* 2012) and reinforces the hypothesis that this area could have acted as a long-term refugium for several species over multiple Pleistocene glaciations. Our paleoecological model shows good habitat suitability in the Padano-Venetian plain, and also reveals that habitat suitability was not homogenous in this part of Italy during the LGM, which may have led to the observed divergence. Additional studies with a better geographical sampling are necessary to precisely delimit the geographical distribution of these three sub-lineages.

In the Spanish-French lineage we found 4 sub-lineages, three of them being allopatric or parapatric, suggesting a history of population allopatric differentiation in multiple refugia during the Pleistocene. Unfortunately, our species niche modeling approach does not allow us to identify the localization of these possible refugia. Three distinct mtDNA sub-lineages were identified at the Spanish-French border. This area is known as an important biogeographic region, where both inter- and intra-specific lineages are known to come into contact in a number of vertebrate taxa (Mila *et al.* 2013). This pattern is due to the presence of the Pyrenees Mountains, which could have acted either as a major geographic barrier and/or as a region where taxa could have survived glaciations and diverged in various refugia thanks to mild Mediterranean and oceanic climatic influences (Feliner 2011; Charrier *et al.* 2014).

Within *T. occidentalis* we found three distinct sub-lineages with a probable allopatric distribution, suggesting a history of population allopatric differentiation in multiple refugia during the Pleistocene. This pattern has been described for several vertebrate taxa within this region and is linked to the geological and ecological complexity of this area (Martínez-Solano *et al.* 2006; Gómez & Lunt 2007; Miraldo *et al.* 2011). The geographical distribution of the three sub-lineages fits our niche projection onto the CCSM paleoclimatic scenario, indicating that suitable LGM climatic conditions for *T. occidentalis* could have existed in Catalonia, Andalusia, Cantabria and Portugal. These areas could correspond to several cryptic refugia for the Iberian blind mole. A better geographic sampling in the Iberian Peninsula is necessary to further test this hypothesis.

Acknowledgements

Molecular analyses were supported by the ATM ‘Génomique et collections’ and the Service de Systématique Moléculaire of the MNHN (UMS 2700, Paris, France). We are grateful to the members of the « Réseau Taup-green » (France) for their help in mole capture and particularly: Jérôme Dormion, Frédéric Cairey-Remonnay, Michel Derosier, Yves Dupont, Jean-Michel Georgeon, Jean-Luc Laclamette, Eric Lafon, Gwendoline Moncel. For the moles captured in France thank also to: Jean-François Barbot, François Burgot, Bernard Dessandre, Bertrand Deputte, Roger Fons, Michèle Giroud, Gabriel Pagès, Daniel Petit, Bernard Pradier, Léon Quemener, Gilou Nicolas and Alexis Ribas. For the moles captured in Spain thanks to: Carlos Feliu and Marcos Miñarro. We are grateful to Steven Holland for allowing us to use his software Analytic Rarefaction ver. 1.4 (http://www.uga.edu/~strata/software/). We thank two anonymous reviewers for their helpful comment on the first version of this paper.

References

Abe, H. (1996). Habitat factors affecting the geographic size variation of Japanese moles. *Mammal Study*, 21, 71-87.

Abe, H. (2001). Soil hardness, a factor affecting the range expansion of *Mogera wogura* in Japan. *Mammal Study*, 26, 45–52.

Alvarez-Lao, D. J. & Garcia, N. (2010). Chronological distribution of Pleistocene cold-adapted large mammal faunas in the Iberian Peninsula. *Quaternary International*, 212, 120–128.

Aulagnier, S., Haffner, P., Mitchell-Jones, A. J., Moutou, F. & Zima, J. (2008). *Guide des mammifères d'Europe, d'Afrique du Nord et du Moyen-Orient*. Paris: Delachaux et Niestlé.

Baker, R. J. & Bradley, R. D. (2006). Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, 87, 643-662.

Bandelt, H. J., Forster, P. & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48.

Bannikova, A. A., Zemlemerova, E. D., Colangelo, P., Sözen, M., Sevindik, M., Kidov, A. A., Dzuev, R. I., Kryštufek, B. & Lebedev, V. S. (2015). An underground burst of diversity –

a new look at the phylogeny and taxonomy of the genus *Talpa* Linnaeus, 1758 (Mammalia: Talpidae) as revealed by nuclear and mitochondrial genes. *Zoological Journal of the Linnean Society*, 175, 930-948.

Canestrelli, D., Salvi, D., Maura, M., Bologna, M. A. & Nascetti, G. (2012). One species, three Pleistocene evolutionary histories: phylogeography of the Italian crested newt, *Triturus carnifex*. *PloS one*, 7, e41754.

Capanna, E. (1981). Karyotype and cranial morphology of *Talpa romana* (Thomas) from terra typica. *Mammalia*, 45, 71–82.

Charrier, O., Dupont, P., Pornon, A. & Escaravage, N. (2014). Microsatellite marker analysis reveals the complex phylogeographic history of *Rhododendron ferrugineum* (Ericaceae) in the Pyrenees. *PloS One*, 9, e92976.

Cleef-Roders, v. J. T. & Hoek Ostende, v. d. L. W. (2001). Dental morphology of *Talpa europaea* and *Talpa occidentalis* (Mammalia: Insectivora) with a discussion of fossil *Talpa* in the Pleistocene of Europe. *Zoologische Mededelingen*, 75, 51-68.

Colangelo, P., Bannikova, A. A., Krystufek, B., Lebedev, V. S., Annesi, F., Capanna, E. & Loy, A. (2010). Molecular systematics and evolutionary biogeography of the genus *Talpa* (Soricomorpha: Talpidae). *Molecular Phylogenetics and Evolution*, 55, 372-380.

Colangelo, P., Aloise, G., Franchini, P., Annesi, F. & Amori, G. (2012). Mitochondrial DNA reveals hidden diversity and an ancestral lineage of the bank vole in the Italian peninsula. *Journal of Zoology*, 287, 42-52.

Collins, W. D., Bitz, C. M., Blackmon, M. L., Bonan, G. B., Bretherton, C. S., Carton, J. A., Chang, P., Doney, S. C., Hack, J. J., Henderson, T. B., Kiehl, J. T., Large, W. G., McKenna, D. S., Santer, B. D. & Smith, R. D. (2006). The Community Climate System Model version 3 (CCSM3). *Journal of Climate*, 19, 2122-2143.

Corti, M. & Loy, A. (1987). Morphometric divergence in southern European moles (Insectivora, Talpidae). *Bollettino di Zoologia*, 54, 187-191.

Cyprien, A.-L., Visset, L. & Carcaud, N. (2004). Evolution of vegetation landscapes during the Holocene in the central and downstream Loire basin (Western France). *Vegetation History and Archaeobotany*, 13, 181-196.

Ducroz, J. F., Volobouev, V. & Granjon, L. (2001). An assessment of the systematics of arvicanthine rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications. *Journal of Mammalian Evolution*, 8, 173-206.

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J. M., Peterson, T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. & Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151.

Feliner, G. N. (2011). Southern European glacial refugia: A tale of tales. *Taxon*, 60, 365-372.

Ferguson, J. W. (2002). On the use of genetic divergence for identifying species. *Biological Journal of the Linnean Society*, 75, 509-516.

Feuda, R., Bannikova, A. A., Zemlemerova, E. D., Febbraro, M., Loy, A., Hutterer, R., Aloise, G., Zykov, A. E., Annesi, F. & Colangelo, P. (2015). Tracing the evolutionary history of the mole, *Talpa europaea*, through mitochondrial DNA phylogeography and species distribution modelling. *Biological Journal of the Linnean Society*, 114, 495-512.

Filippucci, M. G., Nascetti, G., Capanna, E. & Bullini, L. (1987). Allozyme variation and systematic of European moles of the genus *Talpa* (Mammalia, Insectivora). *Mammalia*, 68, 487–499.

Fourcade, Y., Engler, J. O., Rodder, D. & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS One*, 9, e97122.

Giannini, T. C., Chapman, D. S., Saraiva, A. M., Alves-dos-Santos, I. & Biesmeijer, J. C. (2013). Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants. *Ecography*, 36, 649-656.

Gómez, A. & Lunt, D. H. (2007). Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. In N. Weiss & N. Ferrand (Eds) *Phylogeography of southern European refugia* (pp. 155–188). Dordrecht: Springer.

Gorman, M. L. & Stone, R. D. (1990). *The natural history of moles*. London: Christopher Helm Publications.

Grill, A., Amori, G., Aloise, G., Lisi, I., Tosi, G., Wauters, L. A. & Randi, E. (2009). Molecular phylogeography of European *Sciurus vulgaris*: refuge within refugia? *Molecular Ecology*, 18, 2687-2699.

Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307-321.

Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. (2005). PHYML Online- a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Research*, 33, W557-W559.

Hassanin, A., An, J., Ropiquet, A., Nguyen, T. T. & Couloux, A. (2013). Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: Application to the tribe Bovini (Cetartiodactyla, Bovidae). *Molecular Phylogenetics and Evolution*, 66, 766-775.

Hasumi, H. & Emori, S. 2004. K-1 Coupled GCM (MIROC) Description. Tokyo: Center for Climate System Research, University of Tokyo.

He, K., Shinohara, A., Jiang, X. L. & Campbel, K. L. (2014). Multilocus phylogeny of talpine moles (Talpini, Talpidae, Eulipotyphla) and its implications for systematics. *Molecular Phylogenetics and Evolution,* 70, 513-521.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B*, 359, 183-195.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.

Hutterer, R. (2005). Order Soricomorpha. In D. E. Wilson & D. M. Reeder (Eds) *Mammal species of the world: A taxonomic and geographic reference, third ed.* (pp. 220–311). Baltimore: Johns Hopkins University Press.

Jacquet, F., Nicolas, V., Colyn, M., Kadjo, B., Hutterer, R., Akpatou, B., Cruaud, C. & Denys, C. (2014). Forest refugia and riverine barriers promote diversification in the West African pygmy shrew (*Crocidura obscurior* complex, Soricomorpha). *Zoologica Scripta*, 43, 131-148.

Jacquet, F., Denys, C., Verheyen, E., Bryja, J., Hutterer, R., Kerbis Peterhans, J.C., Stanley, W.T., Goodman, S.M., Couloux, A., Colyn, M. & Nicolas, V. (2015). Phylogeography and evolutionary history of the *Crocidura olivieri* complex (Mammalia, Soricomorpha): from a forest origin to broad ecological expansion across Africa. *BMC Evolutionary Biology*, 15, 71.

Jiménez, R., Burgos, M. & Diaz De La Guardia, R. (1984). Karyotype and chromosome banding in the mole (*Talpa occidentalis*) from the south-east of the Iberian Peninsula. Implication on its taxonomic position. *Caryologia: International Journal of Cytology*, 37, 253-258.

Johns, G. C. & Avise, J. C. (1998). A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene. *Molecular Biology and Evolution*, 15, 1481-1490.

Kass, R. & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.

Kryštufek, B. (1987). Skull variability of *Talpa romana stankovici* from Macedonia. *Acta Theriologica*, 32, 463–474.

Kryštufek, B. (1994). The taxonomy of blind moles (*Talpa caeca* and *T. stankovici*, Insectivora, Mammalia) from south-eastern Europe. *Bonner Zoologische Beiträge*, 45, 1–16.

Kryštufek, B. (2001). Skull analysis of small blind moles from Turkey and Iran. *Folia Zoolgica*, 50, 19–25.

Kryštufek, B. & Vohralík, V. (2001). Mammals of Turkey and Cyprus. Introduction, checklist, Insectivora. Koper: Science and Research Centre of the Republic of Slovenia.

Kryštufek, B. & Benda, P. (2002). The Caucasian mole *Talpa caucasica* – a new mammal for Iran. *Mammalian Biology*, 67, 113-116.

Lacey, E. A., Patton, J. L. & Cameron, G. N. (2000). *Life underground: the biology of subterranean rodents*. Chicago, IL: University of Chicago Press.

Leberg, P. L. (2002). Estimating allelic richness: Effects of sample size and bottlenecks. *Molecular Ecology*, 11, 2445-2449.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.

Loy, A. & Corti, M. (1996). Distribution of *Talpa europea* (Mammalia, Insectivora, Talpidae) in Europe: A biogeographic hypothesis based on morphometric data. *Italian Journal of Zoology*, 63, 277-284.

Loy, A. & Capanna, E. (1998). A parapatric contact area between two species of moles (genus *Talpa*): character displacement investigated through the geometric morphometric of skull. *Acta Zoologica Academiae Scientiarum Hungaricae,* 44, 151–164.

Loy, A., Capula, M., Palombi, A. & Capanna, E. (2001). Genetic and morphometric evidence of introgression between two species of moles (Insectivora: *Talpa europaea* and *Talpa romana*) in central Italy. *Journal of Zoology*, 254, 229-238.

Loy, A., Colangelo, P., Annesi, F. & Capanna, E. (2005). Origin and evolution of Western European moles (genus *Talpa*, Insectivora): a multidisciplinary approach. *Mammal Study*, 30, S13-S17.

Marcy, A. E., Fendorf, S., Patton, J. L. & Hadly, E. A. (2013). Morphological adaptations for digging and climate-impacted soil properties define Pocket Gopher (*Thomomys* spp.) distributions. *Plos One*, 8, e64935.

Martínez-Solano, I., Teixeira, J., Buckley, D. & García-París, M. (2006). Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. *Molecular Ecology*, 15, 3375-3388.

Médail, F. & Diadema, K. (2009). Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, 36, 1333-1345.

Mila, B., Surget-Groba, Y., Heulin, B., Gosa, A. & Fitze, P. S. (2013). Multilocus phylogeography of the common lizard *Zootoca vivipara* at the Ibero-Pyrenean suture zone reveals lowland barriers and high-elevation introgression. *BMC Evolutionary Biology*, 13, 192.

Miller, G. S. (1912). *Catalogue of the mammals of Western Europe (Europe exclusive of Russia) in the collection of the British Museum*. London: Trustees of the British Museum.

Miraldo, A., Hewitt, G. M., Paulo, O. S. & Emerson, B. C. (2011). Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, 11, 170.

Mitchell-Jones, A. J., Amori, G., Bogdanowicz, W., Krystufek, B., Reijnders, P. J. H., Spitzenberger, F., Stubbe, M., Thissen, J. B. M., Vohralik, V. & Zima, J. (1999). *The atlas of European mammals*. London: T & AD Poyser Natural History.

Nehlig, P. (2010). Géologie du bassin de la Loire. Geosciences 12, 10-23.

Niethammer, J. & Krapp, F. (1990). *Handbuch der Säugetiere Europas, 3/I: Insektenfresser, Primaten*. Wiesbaden: Aula-Verlag.

Nylander, J. A. (2004). MrModeltest v2. Uppsala, Sweden: Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Pearman, P. B., Guisan, A., Broennimann, O. & Randin, C. F. (2008). Niche dynamics in space and time. *Trends in Ecology & Evolution*, 23, 149-158.

Petrov, B. M. (1971). Taxonomy and distribution of moles (genus *Talpa*, Mammalia) in Macedonia. *Acta Musei Macedonici Scientiarum Naturalium*, 12, 117-138.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-239.

Phillips, S. J. & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161-175.

Posada, D. & Crandall, K.A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, 16, 37-45.

Proust, J.-N., Renault, M., Guennoc, P. & Thinon, I. (2010). Sedimentary architecture of the Loire River drowned valleys of the French Atlantic shelf. *Bulletin de la Société Géologique de France*, 181, 129-149.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61, 539-542.

Schorr, G., Holstein, N., Pearman, P. B., Guisan, A. & Kadereit, J. W. (2012). Integrating species distribution models (SDMs) and phylogeography for two species of Alpine *Primula*. *Ecology and Evolution*, 2, 1260-1277.

Sinclair, S. J., White, M. D. & Newell, G. R. (2010). How useful are species distribution models for managing biodiversity under future climates? *Ecology and Society* 15, 8.

Swofford, D., (2000). PAUP: phylogenetic analysis using parsimony. Washington DC: Smithsonian Institution.

Stephens, M., Smith, N. J. & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, 68, 978-989.

Stephens, M. & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. *American Journal of Human Genetics*, 76, 449-462.

Thaeler, C. S. (1968). An analysis of the distribution of pocket gopher species in northeastern California (Genus *Thomomys*). Berkeley: University of California Press.

Tougard, C., Renvoise, E., Petitjean, A. & Quere, J. P. (2008). New insight into the colonization processes of common voles: inferences from molecular and fossil evidence. *PloS One*, 3, e3532.

Tryfonopoulos, G. A., Thanou, E. G., Fraguedakis-Tsolis, S. E. & Chondropoulos, B. P. (2009). New data on the distribution and genetic structure of Greek moles of the genus *Talpa* (Mammalia, Talpidae). *Journal of Zoological Systematics and Evolutionary Research*, 48, 188-193.

Varela, S., Lima-Ribeiro, M. S. & Terribile, L. C. (2015). A short guide to the climatic variables of the last glacial maximum for biogeographers. *PLoS One,* 10, e0129037.

Warren, D. L., Glor, R. E. & Turelli, M. (2010). ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, 33, 607-611.

Witte, G. R. (1997). *Der Maulwurf Talpa europaea*. Magdeburg, Germany: Westarp Wissenschaften.

Tables

Table 1: Diversity estimates for the main lineages identified in the Cytb phylogenetic analyses. N = number of specimens; h = number of distinct haplotypes; S = number of polymorphic sites; Hd = haplotype diversity; Pi = nucleotide diversity; k = average number of nucleotide differences.

	N	h	S	Hd	Pi	k
European	195	98	151	0.932 ± 0.015	0.00371 ± 0.00035	4.225
Italian	17	12	30	0.956 ± 0.033	0.00755 ± 0.00102	8.603
Spanish-French	118	43	90	0.908 ± 0.021	0.00830 ± 0.00069	9.460
T. occidentalis	22	9	38	0.892 ± 0.041	0.00724 ± 0.00222	8.255

Figure legends

Fig. 1. Map of sampling points showing the distribution of the principal phylogenetic lineages identified on the basis of mtDNA Cytb analyses: *T. occidentalis* (orange); *T. europaea* (pink: European lineage; purple: Italian lineage; green: Spanish-French lineage). Country names abbreviations: BH = Bosnia-Herzegovina, DE = Denmark, FR = France, GR = Greece, HU = Hungary, IT = Italy, NE = Netherlands, PO = Portugal, RU = Russia, SI = Switzerland, SP = Spain, SW = Sweden, UK = United Kingdom, UR = Ukraine. The Loire River is in blue and the Pyrenees Mountains are in light grey.

Fig. 2. Phylogeny recovered by the maximum likelihood (ML) analysis (GTR + I + G model). Numbers on nodes represent ML bootstrap support values and Bayesian posterior probabilities, respectively.

Fig. 3. Median-joining network of the three main lineages of *T. europaea* (A- European; B- Italian; C- Spanish-French) and *T. occidentalis* (D) recovered in the mtDNA Cytb phylogenetic analyses. Branch lengths are proportional to the number of substitutions. Circles represent haplotypes, with size proportional to relative frequencies. In each lineage several distinct sub-lineages are identified by different colors, and the geographical distribution of these sub-lineages is provided in the maps.

Fig. 4. Median-joining network of *T. europaea* and *T. occidentalis* nDNA HDAC2 sequences.Branch lengths are proportional to the number of substitutions. Circles represent haplotypes, with size proportional to relative frequencies. Colors refer to distinct mtDNA Cytb lineages: pink = European lineage, green = Spanish-French lineage, orange = T. *occidentalis* lineage.

Fig. 5. Species distribution modeling of the three main lineages of *Talpa* recovered in our phylogenetic analyses as estimated by Maxent for present-day conditions (current) and for the Last Glacial Maximum (LGM) based on the Model for Interdisciplinary Research on Climate (MIROC) and on the Community Climate System Model (CCSM) paleoclimatic models. Warmer colors show areas with higher probability of presence.

Fig. 6. Mesostyles of the three upper molars (from the left to right: M1, M2, M3) on their buccal sides of the three main lineages of *Talpa* recovered in our phylogenetic analyses (A- *T. occidentalis*; B- *T. europaea*; C and D- Spanish-French lineage). Two

sets of molars are displayed for the Spanish-French lineage to account for the inter-individual variability regarding the mesostyle condition of M2 and M3. Arrowhead in M2 of Fig. 6D points out the additional minute cusp. PS = parastyle, MS = metastyle, MC = metacone. Specimens field numbers: Fig.6A: TO11.09.15.01; Fig.6B: YA0197; Fig.6C: YA0346 (M1) - YA0425 (M2) - YA0386 (M3); Fig.6D: YA0444. Scale bars = 1 mm.

Supporting Information

Appendix S1. List of specimens used in this study, with geographic origins, field numbers, haplotype numbers and GenBank accession numbers. In bold: specimens used in previous papers.
For specimen 11165 (Cytb GenBank KF801513), its geographical origin in GenBank and the paper of Feuda *et al*. (2015) was erroneously attributed to Eyzies-de-Tayac (Dordogne, France) while it was captured in Cerisiers (Yonne, France).

Appendix S2: Maps of localities used to train the Maxent species distribution model (SDM) for *T. europaea sensu stricto* (A); the Spanish-French lineage (B) and *T. occidentalis* (C).

Appendix S3: Median-joining network of the four main lineages of *T. europaea* (A-European; B- Italian; C-Spanish-French) and *T. occidentalis* (D) recovered in the mtDNA Cytb phylogenetic analyses. Each haplotype is designated by a number, and the correspondence between haplotypes and specimen numbers are provided in Appendix S1. Length of branches is proportional to the number of substitutions along a given branch, with haplotypes numbers.

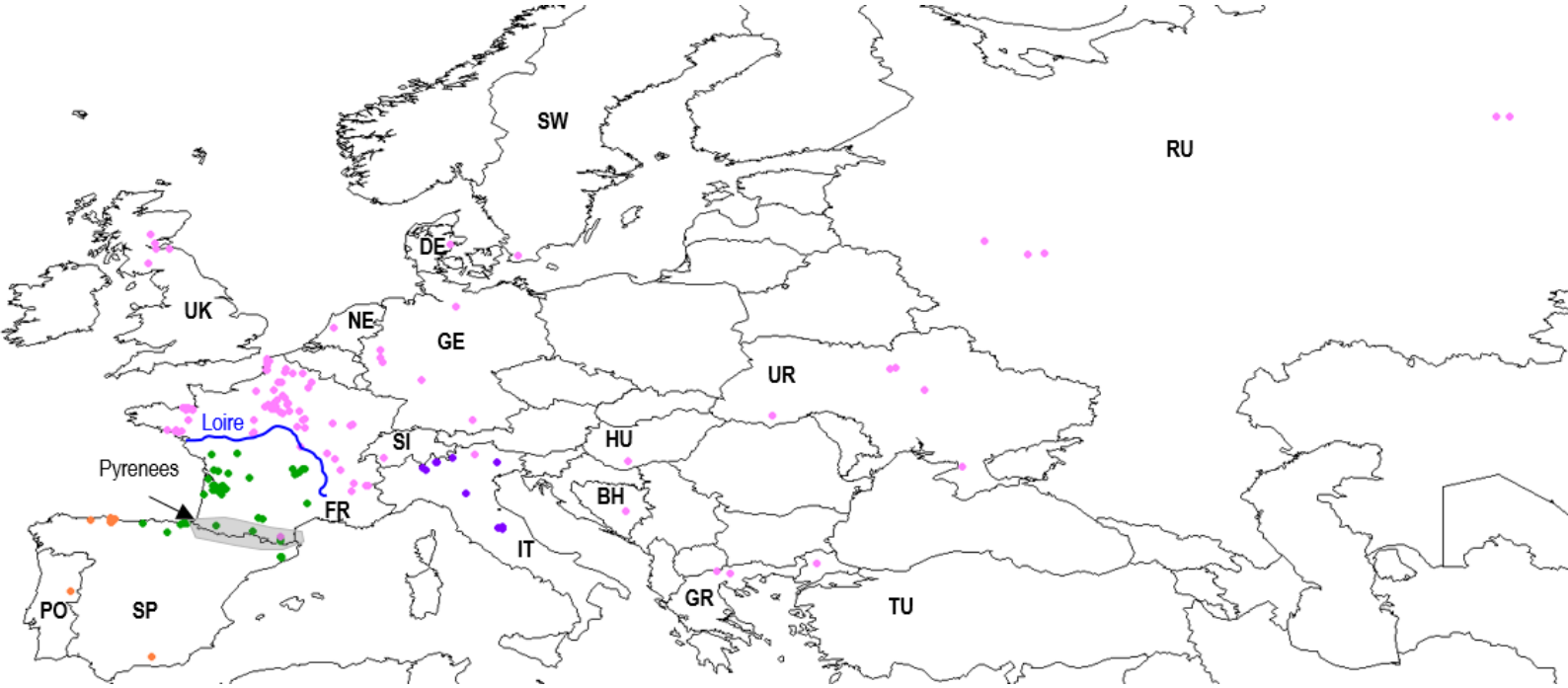

Figure 1

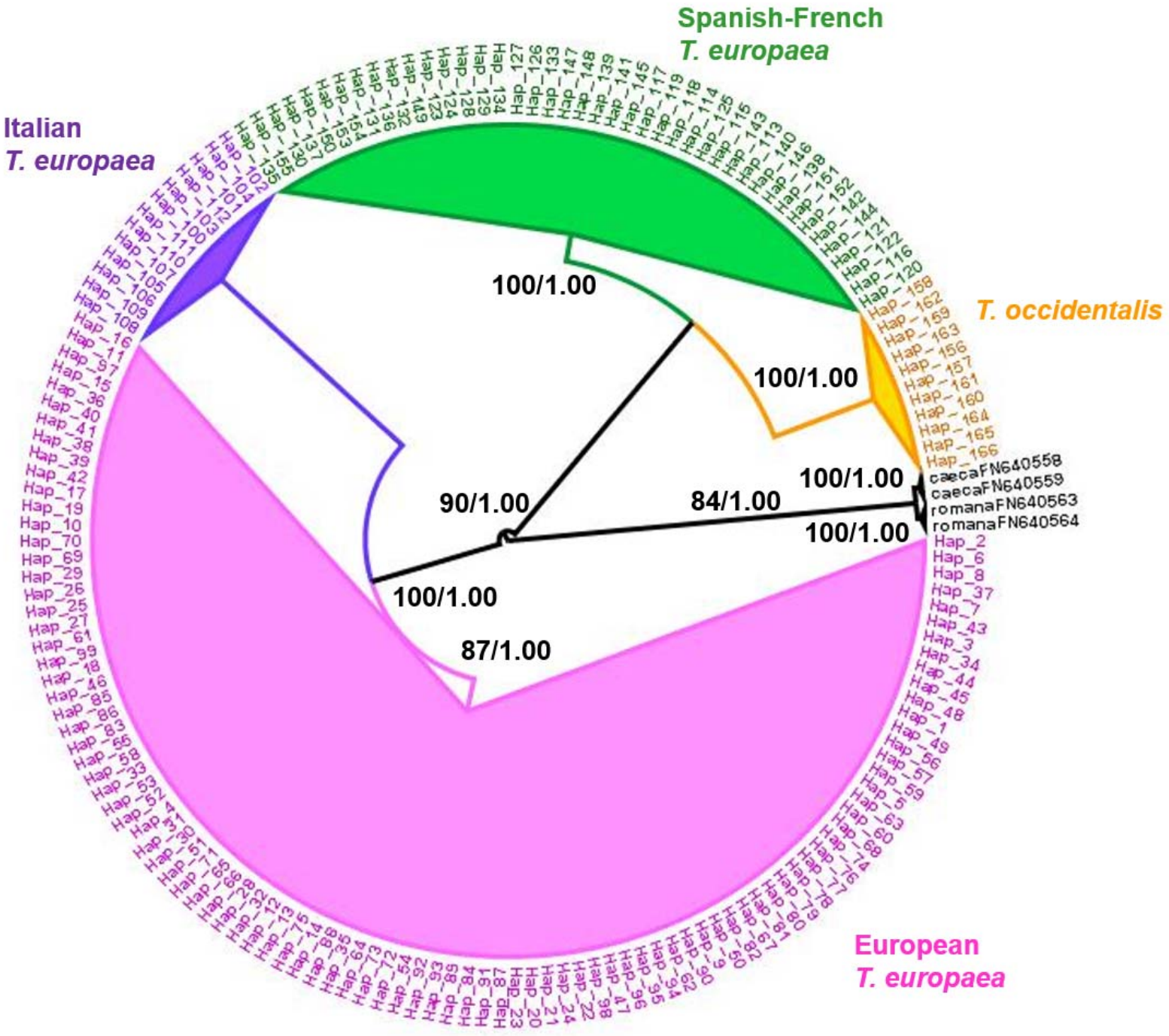
Spanish-French
T. europaea
Italian
T. europaea
T. occidentalis
European
T. europaea
100/1.00
100/1.00
90/1.00
84/1.00
100/1.00
100/1.00
100/1.00
87/1.00
caecaFN640558
caecaFN640559
romanaFN640563
romanaFN640564

Figure 2

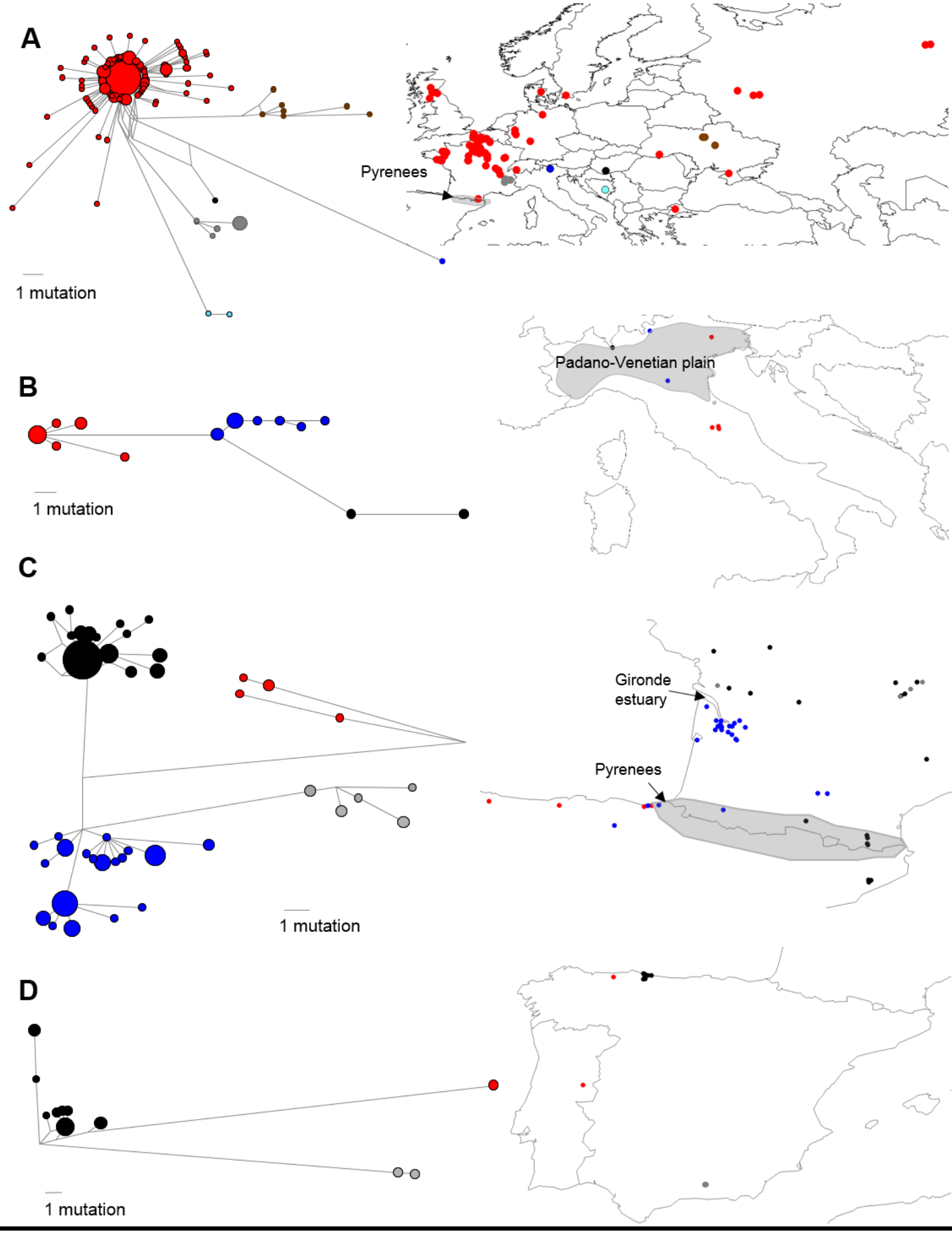
A
Pyrenees
1 mutation
B
Padano-Venetian plain
1 mutation
C
Gironde
estuary
Pyrenees
1 mutation
D
1 mutation

Figure 3

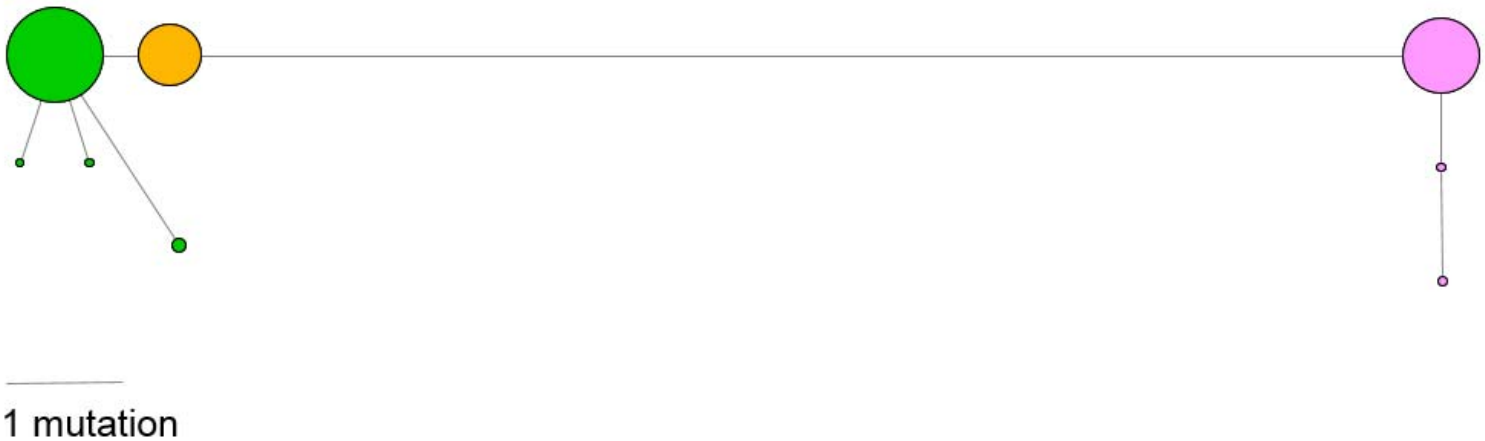

Figure 4

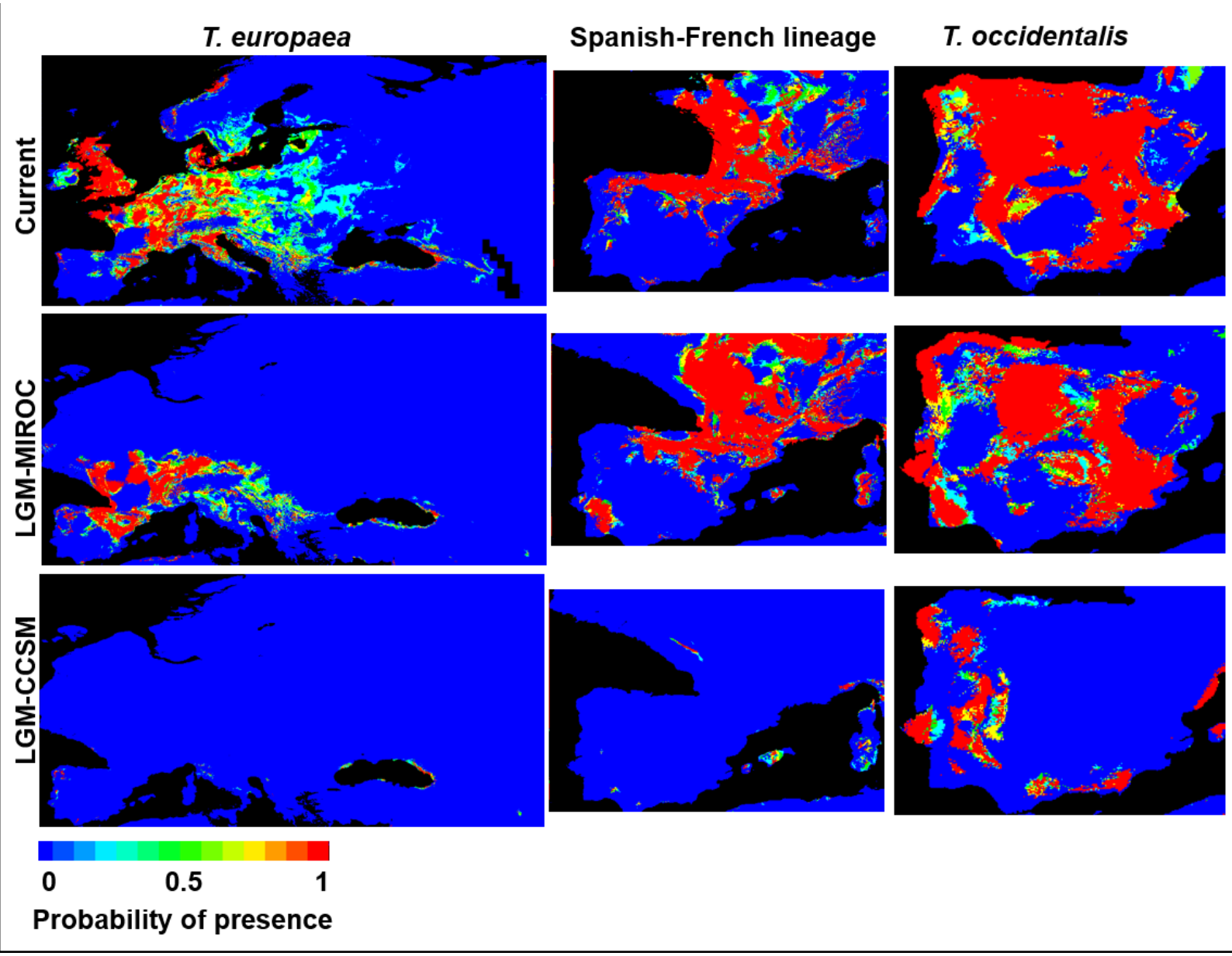
T. europaea
Spanish-French lineage
T. occidentalis
Current
LGM-MIROC
LGM-CCSM
0
0.5
1
Probability of presence

Figure 5

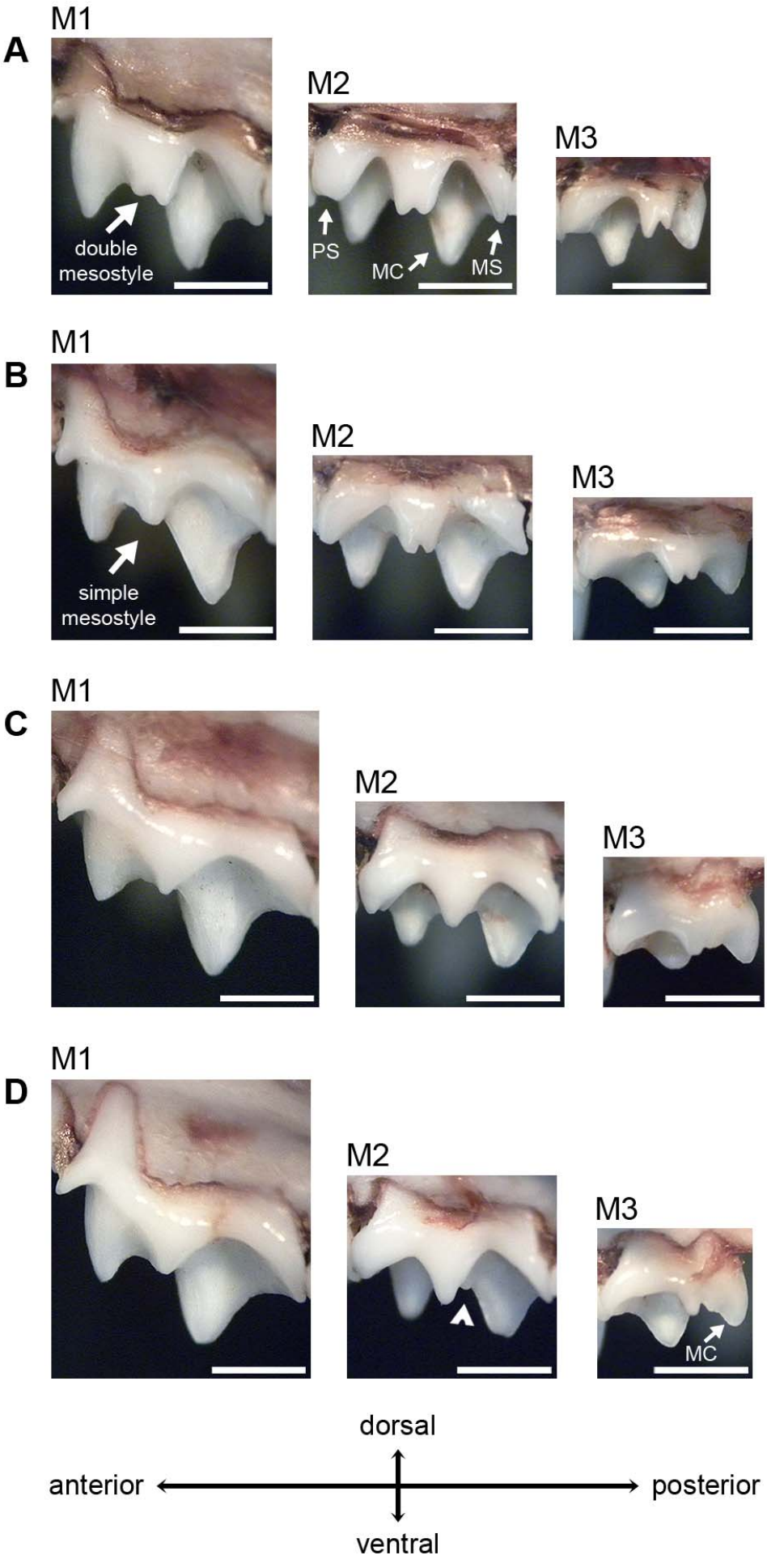
A
M1
double mesostyle
M2
PS
MC
MS
M3
B
M1
simple mesostyle
M2
M3
C
M1
M2
M3
D
M1
M2
M3
MC
dorsal
anterior
posterior
ventral

Figure 6