Evolutionary history and species delimitations: a case study of the hazel dormouse, *Muscardinus avellanarius*

A Mouton[1], A Mortelliti[2], A Grill[3],M Sara[4], B Kryštufek[5], R Juškaitis[6], A Latinne[17], G Amori[8], E Randi[9, 23], S Büchner[10], B Schulz[11], S Ehlers[12], J Lang[13], P Adamik[14] , G Verbeylen[15], M Dorenbosch[16], R Trout[17] , M Elmeros[18], G Aloise[19], S Mazzoti [20], F Matur[21], F Poitevin [22], J. R. Michaux[1,24]

[1] Institut de Botanique, Bâtiment 22, Université de Liège (Sart Tilman), Boulevard du Rectorat 27, 4000 Liège, Belgium

[2] Department of Wildlife Ecology, Fisheries and Conservation Biology, University of Maine 5755 Nutting Hall, Orono, ME 04469 (USA)

[3] Department for Botany and Biodiversity Research University of Vienna, Rennweg 14, A-1030 Wien

[4] Dipartimento Biologia Ambientale e Biodiversità, Laboratorio di Zoogeografia ed Ecologia Animale,Via Archirafi 18, 90123 Palermo

[5] Slovenian Museum of Natural History, Prešernova 20, SI-1000 Ljubljana, Slovenia

[6] Institute of Ecology of Nature Research Centre, Akademijos 2, LT-08412 Vilnius, Lithuania

[7] EcoHealth Alliance, 460 West 34th Street, New York, NY 10001, USA

[8] CNR, Institute of Ecosystem Studies, Viale dell'Università 32, 00185, Rome, Italy

[9] Laboratorio di Genetica, Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Via Cà Fornacetta 9, 40064 Ozzano Emilia (BO), Italy

[10] Ortsstr.174, Friedersdorf 02829 Markersdorf, Germany

[11] Stiftung Naturschutz Schleswig-Holstein Eschenbrook 4, 24113 Molfsee, Germany

[12] Gutenbergstr, 8 24118 Kiel, Germany

[13] Institute of Animal Ecology and Nature Education, Altes Forsthaus, Hauptstr. 30, Institut für Tierökologie und Naturbildung, 35321 Gonterskirchen, Germany

[14] Department of Zoology, Palacky University, 17, listopadu 50, CZ-771 76 Olomouc Czech Republic

[15] Natuurpunt Studie / Mammal Working Group, Coxiestraat 11, B-2800 Mechelen, Belgium

[16] Natuurbalans, Radboud Universiteit, Toernooiveld 1, Postbus 6508, 6503 GA Nijmegen

[17] Holtside Bungalow, Batts Corner, Dockenfield, Farnham, Surrey GU10 4EX, UK

[18] Department of Bioscience - Wildlife Ecology and Biodiversity, Aarhus University, 8410, Rønde, Denmark

[19] Museo di Storia Naturale della Calabria e Orto Botanico, Via P. Bucci s.n., Rende (Cosenza), Italy

[20] Museo di Storia Naturale, Via Filippo de Pisis, 24, I-44100 Ferrara, Italy

[21] Science and Art Faculty, Department of Biology, Bülent Ecevit University, Zonguldak 67100, Turkey

[22] CEFE UMR 5175, CNRS - Université de Montpellier - Université Paul-Valéry Montpellier - EPHE – Laboratory Biogeography and Vertebrate Ecology, 1919 route de Mende, 34293 Montpellier, France

[23] Aalborg University, Department of Biotechnology, Chemistry and Environmental Engineering, Sohngaardsholmsvej 57, DK-9000 Aalborg, Denmark

[24] CIRAD Animal et Gestion Intégrée des Risques (AGIRs), Campus de Baillarguet, F-34093 Montpellier Cedex 5, France.

ABSTRACT

Robust identification of species entities and evolutionary units is essential to implement appropriate conservation strategies for endangered species. However, definitions of species or evolutionary units are numerous and sometimes controversial, which might lead to biased conclusions, with serious consequences for the management of endangered species. The hazel dormouse, an arboreal rodent of conservation concern throughout Europe is an ideal model species to investigate the relevance of species identification for conservation purposes. This species is a member of the Gliridae family, which is protected in Europe and seriously threatened in the northern part of its range. We assessed the extent of genetic subdivision in the hazel dormouse by sequencing one mitochondrial gene (cyt*b*) and two nuclear genes (BFIBR, APOB) and genotyping 10 autosomal microsatellites. These data were analysed using a combination of phylogenetic analyses and species delimitation methods. Multilocus analyses revealed

the presence of two genetically distinct (approximately 11% cyt b genetic divergence, no nuclear alleles shared) lineages for the hazel dormouse in Europe, which presumably diverged during the Late Miocene. The phylogenetic patterns suggests that *M. avellanarius* populations could be split into two cryptic species respectively distributed in western and central-eastern Europe and Anatolia. However, the comparison of several species definitions and methods estimated the number of species between 1 and 10. Our results revealed the difficulty in choosing and applying an appropriate criterion and markers to identify species and highlight the fact that consensus guidelines are essential for species delimitation in the future. In addition, this study contributes to a better knowledge about the evolutionary history of the species.

Key words: *Muscardinus avellanarius*, species delimitation, evolutionary history

Corresponding authors: Alice Mouton, amouton@ulg.ac.be, +32 (0) 43662130

INTRODUCTION

Molecular techniques are a powerful tool to assess species boundaries and to unravel the within-species population structure. Multiple and genetically distinct populations must be preserved to ensure long-term species survival and ecosystem functioning (Luck *et al.*, 2003). To be effective, management and monitoring programs should thus be focused on the identification of appropriate taxonomic and population units to ensure biological diversity conservation. Many European and international directives and organizations (EU Habitats Directive, Bern Convention, IUCN red list) use taxonomic (species) distinctions as a basis for legal protection and management. Unfortunately, the definition of taxonomic units is seriously jeopardized by the lack of a consensus definition on what is a species or an evolutionary unit (Frankham, 2010). Currently,

more than 26, sometimes contradictory, species concepts may be found in the literature (for review, see Guia and Saitoh, 2006; De Queiroz, 2007; Hausdorf, 2011). The use of different definitions can lead to diverse conclusions concerning the number of species (De Queiroz, 2007) and may have critical consequences on conservation plans (Agapow *et al.*, 2004; Isaac *et al.*, 2004; Zachos *et al.*, 2013). The evolutionary significant unit (ESU) is another important widely used conservation concept. In conservation genetics, many studies use the definition proposed by Moritz (1994), which defined an ESU as "populations that are reciprocally monophyletic for mtDNA alleles and demonstrating significant divergence of allele frequencies at nuclear loci". In practice however, several criteria and definitions are used to delineate an ESU, each stressing different theoretically important factors (see review in Guia and Saitoh, 2006). Similar to the situation regarding the species concept, consensus on what an ESU actually is therefore yet to be reached.

To contribute to the general discussion concerning the best species and evolutionary unit concepts to use, particularly for conservation purposes, we studied the hazel dormouse (*Muscardinus avellanarius*) as a model species. This small mammal is strictly protected in Europe (Habitat Directive Annex IV, Bern Convention Annex III) and threatened by habitat loss and fragmentation of forest habitat (Mortelliti *et al.*, 2008, 2010). This species is the focus of several recent conservation plans, including the restoration of habitat corridors, breeding programs or species reintroductions, especially in the northwestern parts of its range (e.g. reintroductions in England and Wales; Interreg IV A- BioGrenzKorr Syddanmark-Schleswig-K.E.R.N; Interreg IV-Habitat Euregio MR). It is therefore essential to gain further insight into the genetic structure of the hazel dormouse in Europe. Previous phylogeographical studies based on mtDNA only revealed a complex genetic structure for this species, including two highly

divergent and allopatric genetic lineages in Europe which are further subdivided into five genetically and geographically well delimitated sublineages (Mouton *et al.*, 2012a,b). Lineage 1 is spread throughout continental western Europe and Italy, while Lineage 2 is found in central Europe, the Balkan Peninsula and Turkey (Figure 1). However, being based on a single mitochondrial locus (cytochrome *b*) and on a limited number of samples (n=120), these conclusions may not be representative of the actual species tree. In this context, it is essential to gain greater insight into the genetic structure of the hazel dormouse in Europe through a multilocus approach.

This study is based on the largest sample of tissues ever collected (n =216) covering a substantial part of the range of the species (Fig 1). These samples were analyzed on the basis of one mitochondrial and two nuclear DNA genes as well as 10 polymorphic autosomal microsatellites. We examined the patterns of genetic variation of the hazel dormouse in order to: i) gain further insight into the evolutionary history of the species, ii) to discuss how species and evolutionary unit concepts may be applied to a particular biological model but also generally for threatened species.

MATERIAL AND METHODS

a) Sampling and DNA extraction

In this study, we used a total of 216 *M. avellanarius* samples collected throughout the species range (Figure 1). The samples were obtained by the authors and other field collaborators (see Acknowledgments) and from collections of the Hungarian Museum of Natural History, the Göteborgs Naturhistoriska Museum, the Naturhistorisches Museum in Vienna, the Natural History Museum of Ferrara and the National History Museum of Denmark (see Acknowledgements). Total genomic DNA was extracted from hairs, buccal swabs, tissues or needles (used for the implementation of the passive implanted

transponder (PIT) tag) using the QIAmp DNA Micro kit and the DNeasy Tissue kit (Qiagen Inc., Valencia, CA, USA) according to the manufacturer's instructions. All samples were handled using sterile disposable scalpels. DNA isolation from the museum samples was performed in a separate dedicated ancient DNA laboratory at the University of Liège using a QIAamp DNA Micro kit (Qiagen). These 216 samples included 120 specimens from a previous study (Mouton *et al*., 2012b) and 96 new ones.

Fig. 1

b) Mitochondrial and nuclear DNA amplification

One mitochondrial marker, the cytochrome *b* gene (cyt*b*) was used in this study in addition to two nuclear genes: the intron 7 region of the β-fibrinogen gene (BFIBR) and the gene coding for apolipoprotein B (APOB). Part of cyt*b* sequences (120 individuals) was already available from a previous study (Mouton et al. 2012b) and was associated with the newly amplified sequences. The final dataset included 216 mitochondrial gene sequences and 130 nuclear gene sequences (alleles) from a subset of 65 individuals representative of the main sampling localities (Table 1). Primer sets used to amplify the cyt*b*, BFIBR, APOB genes are listed in the supplementary Table 1. Amplifications were carried out following the protocol described in Mouton *et al*. (2012). Due to amplification difficulties with some samples (museum samples and needles), 6 further internal specific primers were designed for the cyt*b* sequences. These samples were amplified in 12 µl of Multiplex PCR MasterMix (Qiagen), 1 µl of 10 µM of each primer and deionized water for a total of 20 µl. Cycling conditions followed the Qiagen protocol and included an initial step at 95°C for 15 min, followed by 40 denaturation cycles at 94°C for 30 s, annealing at 58-60°C for 90 s, and extension at 72°C for 30 min. Three

151 independent blanks were carried out for each PCR run: i) an extraction blank to monitor
152 exogenous contamination during extraction, ii) a PCR blank to control PCR products, iii)
153 a PCR blank that remained opened during PCR to monitor aerosols during PCR
154 preparation. Purification and cycle-sequencing reactions (forward and reverse) were
155 performed by the Genoscope (Evry, France) using on an ABI 3730 automatic sequencer.
156
157 Table 1
158 c) Microsatellite genotyping
159 Ten amplified polymorphic loci (five modified from Naim *et al*., 2009: mavE3, mavB5,
160 mavG3, mavG6, mavA5 and five from Mills *et al.,* 2013: Mav021, Mav032, Mav036,
161 Mav051, Mav040) were combined in multiplex sets (Mav021-Mav032- Mav051;Mav036-
162 MAV040;mavG3;mavB5;mavE3;mavG6-mavA5) according to their size and fluorescent
163 label and subsequently amplified via multiplex polymerase chain reactions (PCR) in a
164 Mastercycler Gradient (Eppendorf). The multiplex PCRs contained 5 µl of Multiplex PCR
165 MasterMix (Qiagen), 0.2 µM of each primer and deionized water with a final volume of
166 10 µl. The cycling conditions included an initial step at 95°C for 15 min, followed by 35
167 denaturation cycles at 94°C for 30 s, annealing at 60°C for 90 s, and extension at 72°C for
168 30 min. 2 µl of PCR product were mixed with 0.3 µl of LIZ GS500 (Applied Biosystems)
169 and 12 µl of Hi-Di formamide and loaded onto an ABI 3130 Genetic Analyzer at the
170 University of Brussels. The DNA fragments were analyzed using GeneMapper v.4.1
171 software (Applied Biosystems).
172 d) Analyses of mitochondrial and nuclear genes
173 Sequences were aligned with BIOEDIT 7.2.0 (Hall, 1999) using the ClustalW algorithm.
174 Haplotypes were identified using ARLEQUIN 3.5.1.2 (Excoffier and Lischer, 2010). For

the nuclear genes, heterozygous states were identified as strong double peaks of similar height in both forward and reverse strands, or when the particular base corresponding to the dominant peak alternated on the two chromatograms (Hare and Palumbi, 1999). Nuclear haplotype reconstruction was then conducted using the Bayesian algorithms provided by PHASE 2.1 in DnaSP V. 5.0 (Librado and Rozas, 2009). Two runs were conducted for $1x10^3$ iterations with the default values.

- Phylogenetic analyses

Phylogenetic reconstructions were performed using the maximum likelihood (ML) and Bayesian inference (BI) approaches. Analyses were run independently on mitochondrial and nuclear loci (APOB/BFIBR) and then on the combined dataset (cyt*b*/BFIBR/APOB) (Supplementary Table 1). The nucleotide substitution model that best fitted the dataset was identified with the web application FINDMODEL (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html), developed from MODELTEST (Posada and Crandall, 1998). Three other Gliridae sequences (two *Eliomys quercinus*, one *Glis glis,* GenBank accession number FR848958-FR848957-AJ225031 were chosen as outgroups).

RAxML (Stamatakis, 2006) and MrBayes (Ronquist *et al.*, 2012) allow for data partitioning, thus increasing the accuracy and ability to account for gene specific rates and nucleotide heterogeneity. The ML tree for the cytb, nuclear and combined datasets were constructed using the RAxML software package implemented on a web server "RAxML Blackbox" (Stamatakis *et al,* 2008). The GTR+G substitution model was applied in the analyses. The robustness of the tree was assessed using the rapid bootstrap procedure with 1000 replications implemented in RAxML. The Bayesian phylogeny reconstruction was implemented in MRBAYES 3.2. Metropolis - coupled Markov chain Monte Carlo (MCMC) sampling was performed with 5 chain runs for 5×10^6 generations

200 with one tree sampled every 1000 generations. Bayesian posterior probabilities were
201 picked from the 50% majority rule consensus of trees sampled every 1000 generations,
202 while discarding trees obtained before the chains reached stationary distribution ('burn
203 in', empirically determined by checking the likelihood values).

204 • Genetic diversity and population differentiation

205 Haplotype (h) and nucleotide (π) diversities of the main lineages identified by
206 phylogenetic analyses were estimated for the three loci independently using ARLEQUIN
207 3.5.1.2. Tables of nuclear allele frequency were computed with GENEPOP 4.2.2
208 (Rousset, 2008) and frequency differences were tested for each nuclear locus and across
209 all nuclear loci for all pairs of lineages with GENEPOP 4.2.2.

210 • Divergence time estimates

211 Divergence dates were estimated using Bayesian inferences implemented in BEAST
212 1.7.4 (Drummond *et al.*, 2012) on the cyt*b* dataset. We used two calibration constraints.
213 The first one was based on paleontological estimates and corresponds to the divergence
214 time between *Eliomys quercinus* and *Eliomys melanurus* (FR848958-FR848957,
215 FR848955, FR848956) at 7 ± 0.9 Mya; (Montgelard *et al.*, 2003). The second calibration
216 was based on the estimated split between the Gliridae family and the Sciuridae family.
217 According to Montgelard *et al.* (2002, 2003) and Nunome *et al.* (2007), the Gliridae
218 family arose around 50 Mya. Three *Dryomys* sequences (GI 1694645, xxxx,xxxx) and two
219 additional *Glis* sequences (GI 226486489, GI226486475) were added to our dataset to
220 calibrate the tree. We applied an exponential prior on the tmrca (time of the most recent
221 ancestor) of all taxa, which required specification of only the offset and mean. The model
222 of nucleotide substitution that best fitted the dataset was estimated with the web
223 application FINDMODEL, developed from MODELTEST (Posada and Crandall, 1998).
224 Analyses were performed under the GTR+G+I, an uncorrelated lognormal molecular

clock, and a Bayesian skyline coalescent tree model. These priors were selected because they better fitted the data than any other molecular clock and population models according to the Bayes factor calculated to compare the models. Two independent runs with MCMC length of 50.10^6 were performed with sampling every 5000 generations. Convergence of the chains to the stationary distribution was checked using TRACER 1.5 (Rambaut and Drummond, 2009). All BEAST computations were performed on the computational resource Bioportal at the University of Oslo (http://www.bioportal.uio.no).

- Species delimitation

The Generalized Mixed Yule Coalescent (GMYC) method (Pons *et al.*, 2006; Fujisawa and Barraclough, 2013) is a likelihood method for delimiting independently evolving species. This method compares two models: a) a null model, which assumes a single coalescent process for the entire tree, and b) an alternative Generalized Mixed Yule Coalescent (GMYC), which identifies the transition points from a Yule (species) to a coalescent (population) process. A likelihood ratio test (LRT) was used to evaluate whether the null model was to be rejected or not. If the GMYC model fits the data significantly better than the null model, the threshold T allows estimation of the number of species present in the dataset (Parnmen *et al.*, 2012). The GMYC method requires an ultrametric tree without identical sequences to avoid zero length terminal branches that hamper the likelihood estimation (Fujisawa and Barraclough, 2013). Analyses of the mitochondrial haplotypes were performed using BEAST computations under the same conditions as described above. GMYC analyses were then performed using the R package SPLIT (http://r-forge.r-project.org/projects/splits/).

The Poisson tree processes (PTP) is a new model that can delimit species using non-ultrametric phylogenies (Zhang *et al.*, 2013). The fundamental assumption of this method is that the number of substitutions is significantly higher between species than

within species (Zhang *et al.*, 2013). The test was implemented on the new PTP web server http://species.h-its.org/ptp/ using the phylogenetic trees (cyt*b*, nuclear and combined) obtained in the previous analyses.

The level of cyt*b* net genetic distance between clusters was also used to delimit species differentiation according to the Genetic Species Concept (Baker and Bradley, 2006). Bradley and Baker (2001) concluded that a cyt*b* genetic distance < 2% would equal the intraspecific variation, while values between 2% and 11% would require further study concerning the specific status and values over 11% would be indicative of species recognition. The net genetic distance between lineages was computed using MEGA version 5.2 (Tamura *et al*., 2011) under the Kimura two parameter model (K2P model) for the cyt*b* dataset (to allow comparison with the study of Bradley and Baker (2001) concerning the genetic species concept).

e) Microsatellite analysis

MICROCHECKER version 2.2.3 was used to identify any possible systematic genotyping errors. A linkage disequilibrium (LD) test for each pair of microsatellite loci and conformity to Hardy-Weinberg equilibrium (HWE) was performed using GENEPOP 4.2.2 (Rousset 2008). We calculated pairwise Fst and Rst values to measure the genetic differentiation. Rst is a pairwise population genetic distance that is analogous to Fst, but that takes into account differences in the number of repeats between microsatellite alleles (allele size). We applied the test suggested by (Hardy *et al.*, 2003) and implemented in SPAGeDi version 1.2 (Hardy and Vekemans, 2002) to choose the most suitable estimators. This test indicates whether or not allele sizes provide more information on population differentiation. We compared the observed Rst values with the distribution of Rst obtained after 10 000 allele size permutations (pRST). Rst would

be expected to be significantly higher than the mean permuted value (pRST) when the migration rate is lower than the mutation rate (Hardy *et al.*, 2003). A non-significant result (Rst not significantly different from pRst) would suggest that the allele size is not informative for population differentiation. Significant tests on Rst values are expected if populations had diverged for a sufficiently long time and/or if populations exchanged migrants at a rate similar or inferior to the mutation rate (Hardy *et al.*, 2003). The allelic richness (AR) was calculated by using the rarefaction procedure implemented in FSTAT 2.9.3.2 (Goudet 2001). We used GENETIX v4.05.2 (Belkhir *et al*., 1996-2004) for factorial correspondence analysis (FCA) on the microsatellite data. This approach makes no prior assumptions about the population structure model and HW and linkage equilibrium are not assumed (Allendorf and Luikart, 2007).

To identify the likely number of genetically distinct groups within *M. avellanarius*, we then used Bayesian assignment as implemented in Structure version 2.1 (Pritchard *et al*., 2000). Ten iterations were run for each *K* value from 1 to 10 using an admixture model with a burn-in of 5 $x10^5$ and MCMC values of $5x10^6$. The output of the STRUCTURE analyses was extracted in STRUCTURE HARVESTER (Earl and vonHoldt, 2011). The *K* value that best fitted the dataset structure was revealed by the increasing likelihood of the data and was chosen as the smallest *K* value capturing the major data structure (Pritchard and Wen, 2004). The optimal number of clusters was then assessed based on the correction proposed by Evanno et al. (2005). All STRUCTURE computations were performed on the computational resource Bioportal at the University of Oslo (http://www.bioportal.uio.no). A visual output of STRUCTURE was generated using CLUMPAK (Kopelman et al., 2015).

RESULTS

a) Mitochondrial and nuclear DNA

- Sequence variation

A 704 bp fragment was sequenced from the cytochrome *b* (cyt*b*) gene of the mitochondrial DNA (mtDNA) and contained 135 variable sites. A total of 54 haplotypes was identified within the cyt*b* dataset. For the BFIBR and the APOB genes, 680 bp and 849 bp fragments were obtained, respectively. The BFIBR gene contained 23 variable sites whereas the APOB gene contained 35 variable sites. A total of 12 BFIBR alleles and 26 APOB alleles were identified within our dataset. The haplotype/allele distributions within our dataset are summarized in (Table 1).

- Phylogenetic analyses

Trees obtained for the cyt*b* gene (see Figure 2a) by ML and Bayesian analyses gave similar topologies and revealed the presence of two major lineages which were further geographically structured, as previously reported in Mouton *et al.* (2012b). The haplotypes of the first lineage (hereafter Lineage 1; Bayesian Probabilities, BP=100 , Bootstrap Support, BS=99) clustered into two well supported allopatric sublineages: a western sublineage (BP = 98, BS = 81) encompassing individuals from Belgium, France, Switzerland, northern Italy, Luxembourg and western Germany and a central-southern Italian sublineage (BP = 100, BS = 92).

Within the second lineage (hereafter Lineage 2; BP=92 , BS=68), we observed the presence of three sublineages: a highly supported Balkan sublineage (BP =100, BS=98), with individuals from Serbia, Slovenia, Austria, Macedonia, a Turkish sublineage (BP = 100, BS =97) and another weakly supported central-northern sublineage with individuals from eastern-central and northern Germany, Lithuania, Latvia, Poland, Romania, Hungary, England, Sweden, Denmark, Slovakia and the Czech Republic. The nuclear phylogenetic tree (Figure 2b) also recovered the two major Lineages. Within the

Lineage 2, the tree recovered a monophyletic Turkish sublineage (BP=100, BS=98). In contrast, the Balkan sublineage seemed to be structured into several groups with well supported Slovenian (BP=100, BS=100) and Macedonian groups (BP=100, BS=84). The tree also recovered the weakly supported central-northern sublineage. The relationships within the Lineage 1 (BP= 98, BS=90) were less clear in the nuclear phylogenetic tree than in the mitochondrial dataset.

The ML and Bayesian trees (Figure 2c) combining nuclear and mitochondrial datasets (2233 bp) showed the same topology with Lineage 1 (BP=100, BS =100) and Lineage 2 (BP=91, BS=60), which were further divided into five sublineages.

Fig. 2

- Genetic diversity and population differentiation

Haplotype diversities within lineages were quite high (supplementary Table 3), ranging from 0.245 to 0.775 for mitochondrial markers and from 0.071 to 1 for nuclear markers. Nucleotide diversities were low, ranging from 0.006 to 0.014 for mitochondrial markers and from 0.0008 to 0.007 for nuclear markers. The nuclear allele frequency table (Table 2) showed that no alleles are shared between Lineage 1 and Lineage2 and exact tests of genic differentiation computed across all pairs of lineages were significant at each locus (APOB, BFIB) as well as over both nuclear loci (p=0.000). Interestingly, it seems that the variation of nuclear allelic frequencies is also geographically distributed (supplementary Table 4). No alleles are shared within the substructure in the Lineage 2 while within the Lineage 1 the Italian sublineage shared a single BFIBR and a single APOB allele with the western sublineage.

Table 2

- Divergence time estimates

Divergence time analyses estimated the split between Lineage 1 and Lineage 2 around 6.55 Mya (4.53-8.79). The split between the western European and the Italian sublineages seemed to have taken place around 2.76 Mya (1.77-3.73). Within Lineage 2, the Balkans, Turkish and central-northern sublineages diverged around 2.49 Mya (1.48-3.43). The divergence time estimates with their confidence intervals are summarized in (Figure 3).

Fig. 3

- Species delimitation

The GMYC model was preferred over the null model of uniform branching rates. The likelihood of the GMYC model was significantly higher than that of the null model of uniform (coalescent) branching rates (LR=33.063, P = 0.00). The model based on the cyt*b* dataset led to an estimate of 10 geographically (confidence interval: 9-15) structured putative species (hereafter referred to as GMYC species) for *M. avellanarius* (Figure 3).

PTP yielded a more conservative delimitation than GMYC, with the three putative species identified for *M.avellanarius* for the cyt*b* dataset corresponding to the western European, Italian sublineages and Lineage 2 (Figure 2a). In contrast, the model revealed a single species for the nuclear datasets (Figure 2b). On the other hand, the PTP method based on the combined dataset (nuclear and mitochondrial) identified five putative species corresponding to the five sublineages identified in the mitochondrial phylogenetic tree (Figure 2c).

The extent of genetic divergence (net K2P distance, cyt*b* dataset) was very high between the major two lineages L1 and L2 (10.3%), and high among sublineages within the same lineage (range 3.4-4.2%; Table 3).

Table 3

b) Microsatellites analyses

- Population structure and genetic diversity

No evidence of any scoring error due to stuttering and no evidence of large allele dropout were detected by Microchecker. A significant heterozygote deficit and inbreeding coefficient (Fis) correlated with a significant departure from HWE and LD were detected for most of pairs of loci. These results are consistent with the existence of a population structure, which could be expected at this broad geographic level. A summary statistic of microsatellites data is available in the supplementary Table 5 (ST5).

The STRUCTURE analysis revealed k= 2 as the most likely estimate of K (Figure 4). These clusters corresponded to Lineages 1 and 2, respectively, observed in the phylogenetic trees (mitochondrial, nuclear and combined datasets). The two major lineages (Lineage 1 and Lineage 2) were also clearly identified as separate genetic groups in the FCA (Figure 4). Permutation tests revealed that the multilocus Rst value was significantly higher than the mean pRst. This implies that Rst should be a better estimator than Fst of population differentiation for this group. The Rst values (0.42) was higher than the Fst values (0.26).

Fig. 4

DISCUSSION

- Molecular markers and evolutionary history of the hazel dormouse

Our detailed genetic analysis (cyt*b* mitochondrial DNA, nuclear genes (APOB, BFIB), combined dataset (cyt*b*, APOB, BFIB) and microsatellites) generated compelling

empirical evidence on the existence of two major genetic lineages for the hazel dormouse in Europe. Due to the maternal inheritance of the cyt*b*, the results might reflect only the matrilineal history and might be strongly biased if there is a sex biased dispersal (Zhang and Hewitt 2003; Ballard and Whitlock 2004). In addition, the mtDNA is also characterized by a hypermutability and evidence of homoplasy has been detected in animal phylogenetic analyses (Nabholz et al., 2008; Galtier et al. 2006). The nuclear genes (APOB, BFIB) did not exhibit the same strong differentiation as the cyt*b*. This result is probably due to their slower evolutionary rate and their higher coalescent time as compared to the mitochondrial DNA (Zink and Barrowclough, 2008). However, a high degree of differentiation is underlined as Lineage 1 and Lineage 2 don't share any nuclear alleles. Microsatellites markers exhibited the same strong differentiation with an important Rst value wich is concordant with population that diverged for a sufficiently long time. However because of their high mutation rate, microsatellite data analyses can become problematic when studying the evolutionary relationships between groups that diverged several millions years ago. Indeed, allele size difference may not be related to divergence and homoplasy has been often observed (Zhang and Hewitt, 2003; Estoup et al. 2002). The molecular markers used in the present study might thus present some limitations but altogether they complement one other in deciphering the evolutionary history of the hazel dormouse.

This study evidenced new insight for the phylogeographic history of the species in Northern Europe. The presence of a cyt*b* widespread haplotype (HC15, Table 1) shared by individuals from Slovakia, Poland, Sweden, Denmark, eastern, central and northern Germany, and England is likely due to a recent expansion of *Muscardinus* in central-northern Europe. It has been suggested that the postglacial migration of *M. avellanarius* to Denmark and northern Germany occurred around 12,000 BP (Aaris-Sørensen 1998)

following the extension of deciduous forest promoted by the warmer climate (Vilhelmsen, 2003). The colonization of England by the hazel dormouse probably originated in Denmark and proceeded via a land bridge (Doggerland), which connected Britain to Europe up to the Scandinavian region during and after the last Ice Age (up to 8000 BP) (Lambeck, 1995; Masters and Flemming, 1983). Evidence of this post-glacial colonization route has been documented for other animals such as the pool frog, *Rana lessonae* (Snell *et al.*, 2005), and different small mammals (bank vole, *Myodes glareolus*, field vole, *Microtus agrestis*, and pygmy shrew, *Sorex minutus*) (Searle *et al.*, 2009).

Our results also strongly suggest that the Late Miocene (Tortonian-Messinian: 11 Mya - 5.33 Mya) was an important period for differentiation of the hazel dormouse in Europe. The beginning of the Miocene (23 Mya) was characterized by high mean global temperatures, which peaked during the Middle Miocene climatic optimum between 17 and 15 Mya (Zachos et *al*., 2001) and was recognized as a flourishing period for the Gliridae family (Nadachowski and Daoud, 1995). This warm phase was followed by a climatic cooling (MCC) during the Middle Miocene, around 15–13.5 Mya related to the development of Antarctic ice-sheets (Legendre et al. 2005; Fortelius et al. 2006; Costeur et al. 2007). These climate changes of the Middle to Late Miocene had major impacts on western European terrestrial mammalian fauna, whose diversity declined, with loss of a significant part of their previous forest-dwelling species (Legendre *et al.*, 2005; Fortelius *et al.*, 2006; Costeur *et al.*, 2007). During the Late Miocene and the beginning of the Pliocene, around 5-7 Mya, the climate continued to cool and a seasonality system appeared which had a substantial impact on European land fauna and flora (Casanovas-Vilar *et al.*, 2010). In addition to being an important period of climate change, the Late Miocene in Europe was characterized by some peculiar paleogeographic features. Tobien (1967) and successive studies on faunal assemblages during the Miocene (eg.

447 Fortelius et al., 1996; Casanovas-Vilar et al. 2005) recognized the presence of two
448 distinct major biogeographical provinces in Europe in the Late Miocene: a first one, with
449 a prevalent woodland character is recorded in Central Europe (Portugal, Spain, Belgium,
450 France, Switzerland, Germany, Poland, Czech Republic, Slovakia, Austria, Hungary) and
451 another one with a steppe and/or savanna character is recorded in the Eastern
452 Mediterranean (Greece, Turkey, Serbia, Montenegro, Romania, Moldavia, Ukraina).
453 These two bioprovinces were separated by an inner sea, the Paratethys, (today the
454 remanants of the Paratethys are the Black and Caspian seas). This paratethys barrier
455 acted as a barrier which isolated western Europe from the exchange of flora or fauna
456 and was periodically diruspted allowing for the migration of animals. During the middle
457 Miocene the two western provinces were not very different and they even may
458 constitute a single province characterized by a high diversity of forest-adapted
459 mammals (Casanovas-Vilar et al. 2010). All together, those climatic and physiographic
460 changes during the Miocene might have triggered the diversification of several taxa
461 (Fortelius et al. 2006). The distribution of several mammals are known to have diverged
462 at that time (Santucci *et al.*, 1998; Ludt *et al.*, 2004; Colangelo *et al.*, 2010). For the
463 Gliridae family, the late Miocene was a period of decline (Nadachowski and Daoud 1995)
464 but also an important period of differentiation. We might hypothesized that the climatic
465 changes combined with the presence of new environmental conditions favoured the
466 separation of the ancestor of the hazel dormouse. Interestingly, the formation of the two
467 biogeographical provinces that are however discordant with the current geographic
468 distribution of Lineage 1 and Lineage 2 occurred at the same period. In addition, the
469 hazel dormice spread likely in Italy when a connection with the European continent was
470 established as suggested by the presence of the fossil genus at that time (Kotsakis 2003;
471 Casanovas-Vilar et al. 2010) and they disappeared from the Iberian peninsula. This

recent expansion in Italy might explain why we don't observe a strong differentiation for Lineage 1 with the nuclear genes. The middle and late Miocene were also important period for the differentiation of the other member of the Gliridae family with the presence of deeply divergent lineages. The differentiation between *Eliomys melanurus* (Asian garden dormouse) and *Eliomys quercinus* (garden dormouse) and the colonization of Africa by the ancestor of the genus *Graphiurus* (african dormouse) took place during the late Miocene (Montgelard et al. 2003). The intraspecific differentiation within *E. quercinus* took place around 4.2 Ma (Perez et al. 2013) and a recent phylogeographical study on the edible dormouse uncovered a highly divergent lineage in the North of Iran which separated circa 6 Ma (Naderi et *al.,* 2014). The divergence between *Dryomys nitedula* (forest dormouse) and *Dryomys laniger* (wooly dormouse) appears much older (17 Myr ago) (Montgelard et al. 2003).

- Species delimitation

The results obtained with the different methods revealed the complexity of choosing and applying an appropriate criterion to distinguish between species. The DNA-based species delimitation approach developed by Pons *et al.* (2006) on the basis on the *cyt b* dataset, estimated 10 putative species within the genus *Muscardinus*, while the PTP model (Zhang *et al.*, 2013) estimated only three. The situation is even more complex as we found that the number of estimated species differed according to the genetic markers used (one with nuclear markers and five with the combined nuclear and mitochondrial dataset). These two methods (GMYC and PTP) are both based on the phylogenetic species concept (PSC) originally proposed by Cracraft (1983) which defined a species as "the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent". This concept has recently been highly criticized and discussed (Agapow *et al.*, 2004; Hausdorf, 2011; Frankham, 2012; Zachos

497 and Lovari, 2013; Zachos *et al*., 2013). Frankham *et al.* (2012) even concluded that the
498 PSC is unsuitable for use in conservation contexts, especially for classifying allopatric
499 populations. In fact, taxonomic inflation (Isaac *et al.*, 2004) is the major concern with the
500 PSC, sometimes with nearly the double of species newly recognized (Zachos and Lovari,
501 2013; Heller *et al.*, 2013). Increased splitting of species can have serious consequences
502 for conserving biodiversity as the identification of too many taxa (oversplitting) can
503 waste limited conservation resources (Allendorf and Luikart, 2007; Heller *et al*., 2013)
504 and lead to inappropriate management strategies (e.g. translocations, captive breeding
505 decisions) (Zachos and Lovari, 2013). The use of such concepts to define the number of
506 putative species in *Muscardinus* therefore appears to be complicated and debatable.
507 DNA sequence divergence values could also be used as an additional data source for the
508 establishment of an appropriate measure of taxonomic rank (Bradley and Baker, 2001).,
509 Two *M. avellanarius* species would be recognized under the GSC (Baker and Bradley,
510 2006). The cyt *b* divergence between Lineage 1 and Lineage 2 is high (10.3%) and
511 comparable to that found between Asian striped squirrels (genus Tamiops; *T.*
512 *maritimus* and *T. swinhoei*) (Chang *et al*., 2011) or between dormice species (genus
513 *Eliomys; E. quercinus and E. melanurus*) (Montgelard *et al*. 2003). However, recent
514 studies have also revealed that "intraspecific" divergences in species from monotypical
515 genera can be also very deep (e.g. an Iranian lineage within the edible dormouse, Naderi
516 *et al.* 2014 or within the genus *Petaurista,* Li *et al*., 2013). It is therefore difficult to use
517 such information to determine the true taxonomic status of the two hazel dormouse
518 genetic lineages. In addition, results based on a single genetic marker do not necessarily
519 provide conclusive evidence on speciation (Zachos and Lovari, 2013). For instance, the
520 10 putative *Muscardinus* species inferred with GMYC approaches likely represent 10
521 allopatric populations evolving neutrally rather than 10 "real" species. Indeed, a

522 shortcoming of this method is that a single species with a strong spatial population
523 structure could be wrongfully split into several separate GMYC lineages (Pons *et al.*,
524 2006).
525 Sauer and Hausdorf (2012) recommended using multilocus markers for more detailed
526 analyses of taxa but they admit that even this approach has its limits in disentangling
527 species within a single cluster. The application of the PTP model (based on the PSC) on
528 the nuclear markers and the combined dataset (cyt*b* and nuclear markers) resulted in
529 either one or five hazel dormouse species, respectively. However, the phylogenetic
530 reconstructions revealed two geographically separated monophyletic lineages,
531 statistically supported and concordant between nuclear and mitochondrial genes. Under
532 the PSC, these results suggest the existence of two cryptic species of *M. avellanarius*.
533 Several studies have revealed significant geographic variation for the hazel dormouse
534 based on morphological characters (Storch, 1978; Corbet, 1978; Kıvanç, 1983, review in
535 Juškaitis and Büchner 2013), but there is no consensus on the existence of categorical
536 races (subspecies)(Holden 2005). A formal recognition of two species of *M. avellanarius*
537 therefore is not supported by morphological evidence. The Biological Species Concept
538 BSC, uses mating isolation as a criterion to distinguish species. However, there is
539 currently no evidence that different mechanical reproductive isolating mechanisms exist
540 among hazel dormouse populations. This may be because of the lack of studies on any
541 characters associated with reproduction (Simson *et al.*, 1995). It is therefore impossible
542 to establish the presence of one or two species based on this concept.
543 Recently, another concept similar to the BSC, i.e. the Differential Fitness Species Concept
544 (DFSC), was introduced by Hausdorf (2011). It takes into consideration the pre- and
545 post-zygotic reproductive isolation criterion to define species. Under the DFSC, a species
546 is characterized by features that would have negative fitness effects on the other group

and that cannot be regularly exchanged between groups upon contact (Hausdorf, 2011). So far, the DFSC is considered as highly relevant for conservation purposes because it minimizes outbreeding depression and maximizes the fitness (Frankham *et al.*, 2012). We think that this concept might be considered as a consensus for scientists when delineating the species.

Our results highlighted the ambiguity of delimitating species entities. We found that different approaches based on the same concept (see PSC) but also that different concepts based on single-locus or multi-locus markers might lead to different conclusions. To avoid the problem of species definitions, Zachos (2013a) suggested using intraspecific diversity for conservation purposes by delimitating, for instance, evolutionary significant units (ESUs), but this concept is also controversial. This concept was introduced by Ryder (1986) as a potential conservation unit to be applied below the species level instead of subspecies, which is often considered as a subjective concept. Under this definition, a concordant dataset derived from different approaches (life history information, morphometrics, range and distribution records and genetic data) is required (Ryder, 1986). Guia and Saitoh (2006) recommend using the term 'partial ESU' when the results do not fulfill the original definitions of Ryder (1986) and considering the term 'full ESU' when information on both neutral genetic and adaptive variation are available. Under the definition of Moritz (1994), two ESUs would exist for the hazel dormouse in Europe. As these ESU are molecular-based, we should consider that two partial ESUs exist. Further studies are required to confirm the existence of two full ESUs within *M. avellanarius*.

CONCLUSION

570 Our effort to delimitate species or evolutionary entities revealed that the number of
571 possible/putative species for the hazel dormouse is between 1 and 10. Would the
572 genetic evidence on its own not provide conclusive evidence on species limits?
573 Taxonomic uncertainties could certainly be better solved by using an integrative
574 approach. Future research should focus on some aspects that have not been sufficiently
575 studied in *M. avellanarius*, such as social communication, reproduction mechanisms or
576 morphometrical differentiation, etc., in order to gain insight into possible adaptive
577 differentiation among populations in Europe. In addition, a extensive sampling would be
578 highly recommended in the possible zones (see Fig 1) of overlap between the two
579 ancient lineages to reveal a contact zone or an hybrid zone. Beyond the fact that the
580 present study did not clearly reveal the presence of cryptic species of *Muscardinus* in
581 Europe, we argue that the two lineages can no longer be considered as a single entity
582 and that future conservation and management plan such as reintroduction or breeding
583 programs should take into account the presence of two genetic lineages.
584

585 ## DATA ARCHIVING

586 GenBank accession numbers can be found in supplementary Table 5.
587

588 ## CONFLICT OF INTEREST

589 The authors declare there are no conflicts of interest.
590

591 ## ACKNOWLEDGMENTS

592 We thank everyone who provided tissue samples of *M. avellanarius*: Peter Vogel, Valdis

Pilats, Luis Popa, Josef Bryja, Achim Schumacher, Helle Vilhelmsen, Nilson Goran (Goteborg Natural History Museum), Gabor Csorba (Hungarian Natural History Museum), Anita Gamauf (Wien Natural History Museum), Hans J Baagøe (National History Museum of Denmark). We thank Adrien Rieux, Jiajie Zhang for their kind help and advice for the phylogenetic analyses. This project was supported by the network "Bibliothèque du Vivant" funded by CNRS, the MNHN, INRA and CEA. A. Mouton is supported by a Belgian research fellowship from FRIA (Fonds pour la Formation et la Recherche dans l'Industrie et dans l'Agriculture) and a financial grant from the Belgian FNRS (crédits pour brefs séjours à l'étranger to A. Mouton) and from the University of Liège (Patrimoine) and J. R. Michaux (mandat Maitre de recherches) is supported by a Belgian research fellowship from FNRS (Fonds National pour la Recherche Scientifique) and financial grants from the Belgian FNRS (crédits aux chercheurs to J. R. Michaux). Part of this work was supported by the INTERREG-project "BioGrenzKorr" carried out by Naturstyrelsen, Stiftung Naturschutz Schleswig-Holstein and Schleswig-Holsteinische Landesforsten.

REFERENCES

Aaris-Sørensen K (1998). Danmarks Forhistoriske Dyreverden. Gyldendal, Denmark.

Agapow P-M, Bininda-Emonds OR, Crandall K a, Gittleman JL, Mace GM, Marshall JC, et al. (2004). The impact of species concept on biodiversity studies. Q. Rev. Biol. 79: 161–79.

Allendorf FW, Luikart G (2007). Conservation and the Genetics of Populations. Wiley-Blackwell, New York, NY.

Baker RJ, Bradley RD (2006). Speciation in Mammals and the Genetic Species Concept. J. Mammal. 87: 643–662.

616 Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Mol
617 Ecol 13:729–744.
618 Belkhir K, Borsa P, Chikhi L, Raufaste N, et al. (1996-2004). GENETIX 4.05, Logiciel Sous
619 Windows TM Pour la Génétique des Populations. Laboratoire Génome, Populations,
620 Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier.
621 Bradley RD, Baker RJ (2001). A test of the genetic species concept: Cytochrome-b
622 sequences and mammals. J. Mammal. 82: 960–973.
623 Casanovas-Vilar I, Moyà-Solà S, Agustí J, Köhler M (2005) The geography of a faunal
624 turnover : tracking the Vallesian Crisis. In: Migration of Organisms. pp 247–300.
625 Casanovas-Vilar I, García-Paredes I, Alba DM, Van Den Hoek Ostende LW, Moyà-Solà S
626 (2010). The European Far West: Miocene mammal isolation, diversity and turnover in
627 the Iberian Peninsula. J. Biogeogr. 37: 1079–1093.
628 Chang SW, Oshida T , Endo H, Nguyen ST, Dang CN, Nguyen DX, Jiang X, Li ZJ, Lin LK
629 (2011). Ancient hybridization and underestimated species diversity in Asian striped
630 squirels (genus *Tamiops*): Inference from paternal, maternal and biparental markers. J
631 Zool. 285 :128-138.
632 Colangelo P, Bannikova a a, Krystufek B, Lebedev VS, Annesi F, Capanna E, et al. (2010).
633 Molecular systematics and evolutionary biogeography of the genus *Talpa*
634 (Soricomorpha: Talpidae). Mol. Phylogenet. Evol. 55: 372–80.
635 Corbet GB (1978). The mammals of the Palearctic Region: a taxonomic review. – British
636 Museum (Natural History). London & Ithaca, NY: Cornell University Press.
637 Costeur L, Legendre S, Aguilar J-P, Lécuyer C (2007). Marine and continental
638 synchronous climatic records: Towards a revision of the European Mid-Miocene
639 mammalian biochronological framework. Geobios 40: 775–784.

640 Costeur L, Montuire S, Legendre S, Maridet O (2007). The Messinian event: What
641 happened to the peri-Mediterranean mammalian communities and local climate?
642 Geobios 40: 423–431.
643 Cracraft J (1983). Species concepts and speciation analysis. In: Johnston RF, ed. *Current*
644 *ornithology*. 159–187. New York: Plenum Press
645 Drummond AJ, Suchard M a, Xie D, Rambaut A (2012). Bayesian phylogenetics with
646 BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29: 1969–73.
647 Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for
648 visualizing STRUCTURE output and implementing the Evanno method. Cons Gen Res 4:
649 359–361.
650 Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals
651 using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611–20.
652 Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: A new series of programs to
653 perform population genetics analyses under Linux and Windows. Mol. Ecol.Res. 10: 564-
654 567.
655 Estoup A, Jarne P, Cornuet JM (2002) Homoplasy and mutation model at microsatellite
656 loci and their consequences for population- genetics analysis. Mol Ecol, 11, 1591–1604
657 Fortelius M, Eronen J, Liu L, Pushkina D, Tesakov A, Vislobokova I, et al. (2006). Late
658 Miocene and Pliocene large land mammals and climatic changes in Eurasia. Palaeogeogr.
659 Palaeoclimatol. Palaeoecol. 238: 219–227.
660 Frankham R (2010). Challenges and opportunities of genetic approaches to biological
661 conservation. Biol. Conserv. 143: 1919–1927.
662 Frankham R, Ballou JD, Dudash MR, Eldridge MDB, Fenster CB, Lacy RC, et al. (2012).
663 Implications of different species concepts for conserving biodiversity. Biol. Conserv. 153:
664 25–31.

665 Fujisawa T, Barraclough TG (2013). Delimiting species using single-locus data and the
666 Generalized Mixed Yule Coalescent approach: a revised method and evaluation on
667 simulated data sets. Syst. Biol. 62: 707–24.
668 Galtier N, Enard D, Radondy Y, et al. (2006) Mutation hot spots in mammalian
669 mitochondrial DNA. Genome Res 16:215–22.
670 Garcia-Alix A, Minwer-Barakat R, Martin-Suarez E, Freudenthal M (2008). *Muscardinus*
671 *meridionalis sp*. nov., a new species of Gliridae (Rodentia, Mammalia) and its
672 implications for the phylogeny of Muscardinus. J. Vertebr. Paleontol. 28**:** 568–573.
673 Guia APO, Saitoh T (2006). The gap between the concept and definitions in the
674 Evolutionarily Significant Unit: the need to integrate neutral genetic variation and
675 adaptive variation. Ecol. Res. 22: 604–612.
676 Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and
677 analysis program for Windows 95/98/ NT. Nuc. Acid S*.* 41**:** 95–98.
678 Hardy OJ, Charbonnel N, Fréville H, Heuertz M (2003). Microsatellite allele sizes: a
679 simple test to assess their significance on genetic differentiation. Genetics 163: 1467–
680 1482.
681 Hardy O, Vekemans X (2002). SPAGeDi: a versatile computer program to analyse spatial
682 genetic structure at the individual or population levels. Mol. Ecol. Notes 2: 618-620.
683 Hare MP, Palumbi SR (1999) The accuracy of heterozygous base calling from diploid
684 sequence and resolution of haplotypes using allele-specific sequencing. Mol Ecol 8:
685 1750–1752.
686 Hausdorf B (2011). Progress toward a general species concept. Evolution 65: 923–31.
687 Heller R, Frandsen P, Lorenzen ED, Siegismund HR (2013). Are there really twice
688 asmany bovid species as we thought? Syst Biol 62:490–493

689 Holden ME (2005). Family Gliridae. In: Wilson DE, Reeder DM, eds. Mammal species of
690 the world, 3rd edn. Washington and London: Smithsonian Institute Press, 819–841
691 Isaac NJB, Mallet J, Mace GM (2004). Taxonomic inflation: its influence on macroecology
692 and conservation. Trends Ecol. Evol. 19: 464–9.
693 Juškaitis R, Büchner S (2013). The Hazel Dormouse Vol.2. NBB English Edition.
694 Kivanç E (1983). Die Haselmaus, *Muscardinus avellanarius* L., in der Türkei. Bonn zool
695 Beitr. 34 :419-428.
696 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. and Mayrose, I. (2015).
697 Clumpak: a program for identifying clustering modes and packaging population
698 structure inferences across *K*. Mol Ecol Res, 15: 1179–1191
699 Kotsakis T (2003) Fossil glirids of Italy : the state of the art Glíridos fósiles de Italia :
700 situación actual. Coloquios Paleontol 1:335–343.
701 Lambeck K (1995). Late Devensian and Holocene shorelines of the British Isles and
702 North Sea from models of glacio-hydro- isostatic rebound. J. Geol. Soc.London152:437–
703 448
704 Legendre S, Montuire S, Maridet O, Escarguel G (2005). Rodents and climate: A new
705 model for estimating past temperatures. Earth Planet. Sci. Lett. 235: 408–420.
706 Li S, He K, Yu F-H, Yang Q-S (2013). Molecular Phylogeny and Biogeography of *Petaurista*
707 Inferred from the Cytochrome b Gene, with Implications for the Taxonomic Status of *P.*
708 *caniceps, P. marica and P. sybilla*. PLoS ONE 8(7): e7046
709 Librado P, Rozas J (2009). DnaSP v5: A software for comprehensive analysis of DNA
710 polymorphism data. Bioinformatics 25**:** 1451-1452
711 Luck GW, Daily GC, Ehrlich PR (2003). Population diversity and ecosystem services.
712 Trends Ecol. Evol. 18:331–336.

713 Ludt CJ, Schroeder W, Rottmann O, Kuehn R (2004). Mitochondrial DNA phylogeography
714 of red deer (*Cervus elaphus*). Mol. Phylogenet. Evol. 31: 1064–83.
715 Masters PM, Flemming NC (1983). Quaternary coast- lines and marine archaeology:
716 towards the prehistory of land bridges and continental shelves. London: Academic Press.
717 MatLab.
718 Mills C, Dawson DA,Horsburgh GJ, Godley BJ,Hodgson DJ (2013). Isolation and
719 characterisation of hazel dormouse (*Muscardinus avellanarius*) microsatellite loci.
720 Conservation Genet Resour 5: 687-692.
721 Montgelard C, Bentz S, Tirard C, Verneau O, Catzeflis FM (2002). Molecular systematics
722 of sciurognathi (rodentia): the mitochondrial cytochrome b and 12S rRNA genes support
723 the Anomaluroidea (Pedetidae and Anomaluridae). Mol. Phylogenet. Evol. 22: 220–233.
724 Montgelard C, Matthee CA, Robinson TJ (2003). Molecular systematics of dormice
725 (Rodentia: Gliridae) and the radiation of *Graphiurus* in Africa. Proc. Biol. Sci. 270:1947–
726 55.
727 Mortelliti A, Amori G, Capizzi D, Rondinini C, Boitani L (2010). Experimental design and
728 taxonomic scope of fragmentation studies on European mammals: current status and
729 future priorities. Mamm. Rev. 40: 125–154.
730 Mortelliti A, Santulli Sanzo G, Boitani L (2008). Species' surrogacy for conservation
731 planning: caveats from comparing the response of three arboreal rodents to habitat loss
732 and fragmentation. Biodivers. Conserv. 18: 1131–1145.
733 Moritz C (1994) Defining 'Evolutionarily Significants Units' for conservation. Trends
734 Ecol. Evol. 9: 373-375
735 Mouton A, Grill A, Sara M, Kryštufek B, Randi E, Amori G, et al. (2012a). Using
736 phylogeography to promote dormouse conservation: the case of *Muscardinus*
737 *avellanarius* (Rodentia, Gliridae). Peckiana 8: 255–264.

738 Mouton A, Grill A, Sara M, Kryštufek B, Randi E, Amori G, Juškaitis R, Aloise G, Mortelliti
739 A, Panchetti F, Michaux J (2012). Evidence of a complex phylogeographic structure in the
740 common dormouse, *Muscardinus avellanarius* (Rodentia: Gliridae). Biol. J. Linn. Soc.
741 105: 648–664
742 Nabholz B, Glemin S, Galtier N (2008) Strong Variations of Mitochondrial Mutation Rate
743 across Mammals—the Longevity Hypothesis. Mol Biol Evol 25 (1): 120-130.
744 Nadachoswki A, Daoud A (1995). Patterns of myoxid evolution in the Pliocene and
745 Pleiostocene of Europe. *Hystrix* 6: 141–149.
746 Naderi G, Kaboli M, Koren T, Karami M, Zupan S, Rezaei HR, Krystufek B (2014).
747 Mitochondrial evidence uncovers a refugium for the fat dormouse (*Glis glis* Linnaeus,
748 1766) in Hyrcanian forests of northern Iran. Mamm Biol - Zeitschrift für Säugetierkd
749 79:202–207.
750 Naim D, Kemp SJ, Telfer S, Watts PC (2009) Isolation and characterization of 10
751 microsatellite loci in the common dormouse *Muscardinus avellanarius*. Mol Ecol Res
752 9:1010–1012.
753 Nunome M, Yasuda SP, Sato JJ, Vogel P, Suzuki H (2007). Phylogenetic relationships and
754 divergence times among dormice (Rodentia, Gliridae) based on three nuclear genes.
755 Zool. Scr. 36: 537–546.
756 Parnmen S, Rangsiruji A, Mongkolsuk P, Boonpragob K, Nutakki A, Lumbsch HT (2012).
757 Using phylogenetic and coalescent methods to understand the species diversity in the
758 *Cladia aggregata* complex (Ascomycota, Lecanorales). PLoS One 7: e52245.
759 Perez GCL, Libois R, Nieberding CM (2013). Phylogeography of the garden dormouse
760 *Eliomys quercinus* in the western Palearctic region. J Mammal 94:202–217.

761 Pons J, Barraclough T, Gomez-Zurita J, Cardoso A, Duran D, Hazell S, et al. (2006).
762 Sequence- Based Species Delimitation for the DNA Taxonomy of Undescribed Insects.
763 Syst. Biol. 55: 595–609.
764 Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution.
765 *Bioinformatics* **14:** 817– 818.
766 Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using
767 multilocus genotype data. Genetics, 155, 945– 959
768 Pritchard JK, Wen W (2004). Documentation for STRUCTURE Software Version 2.
769 Available from http://pritch.bsd. uchicago.edu
770 De Queiroz K (2007). Species concepts and species delimitation. Syst. Biol. 56: 879–86.
771 Rambaut A, Drummond AJ (2009). Tracer v1.5. Available: http://
772 beast.bio.ed.ac.uk/Tracer.
773 Rousset F (2008) GENEPOP ' 007: a complete re-implementation of the GENEPOP
774 software for Windows and Linux. Mol Ecol Res, 8, 103-106.
775 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S et al. (2012).
776 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large
777 model space. Syst Biol 61:539–54
778 Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population
779 structure. Mol Ecol Notes. 4: 137-138.
780 Ryder OA (1986). Species conservation and systematics: the dilemma of subspecies. Trends Ecol.
781 Evol. 1: 9-10.
782 Santucci F, Emerson BC, Hewitt GM (1998). Mitochondrial DNA phylogeography of
783 European hedgehogs. Mol. Ecol. 7: 1163–72.

784 Sauer J, Hausdorf B (2012). A comparison of DNA-based methods for delimiting species
785 in a Cretan land snail radiation reveals shortcomings of exclusively. Cladistics 28: 300 –
786 316.
787 Searle JB, Kotlík P, Rambau RV, Marková S, Herman JS, McDevitt AD (2009). The Celtic
788 fringe of Britain: insights from small mammal phylogeography. Proc. R. Soc. B Biol. Sci.
789 276, 201– 207
790 Simson S, Ferrucci L,Kurtonur C, Ozkan B, Filippucci MG (1994). Phalli and bacula of
791 european dormice: description and comparison. Hystrix 6:231-244.
792 Snell C, Tetteh J, Evans IH (2005). Phylogeography of the pool frog (*Rana lessonae*
793 Camerano) in Europe: evidence for native status in Great Britain and for an unusual
794 postglacial colonization route. Biol. J. Linn. Soc. 85: 41–51.
795 Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses
796 with thousands of taxa and mixed models. Bioinformatics 22, 2688-2690.
797 Stamatakis A, Hoover P, Rougemont J (2008). A rapid bootstrap algorithm for the RAxML
798 web-servers. Syst. Biol. 57: 758-771.
799 Storch G (1978). Gliridae – Schlafer. In: Niethammer J, Krapp F, eds. *Handbuch der*
800 *Saugetiere Europas*. Wiesbaden: Akad. Verlagsges 1. 201–280.
801 Tamura K, Peterson D, Peterson N, Stecher G, Nei N, Kumar S. (2011). MEGA5: molecular
802 evolutionary genetics analysis using maximum likelihood, evolutionary distance, and
803 maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.
804 Tobien H (1967) Subdivision of Pontian mammal faunas. Giornale di Geologia 35: 1–5.
805 Vilhelmsen H (2003). Status of dormice (*Muscardinus avellanarius)* in Denmark. Acta
806 Zool. Acad. Sci. Hungaricae 49: 139–145.

807 Zachos FE, Apollonio M, Bärmann E V, Festa-bianchet M, Göhlich U, Christian J, et al.
808 (2013). Species inflation and taxonomic artefacts — A critical comment on recent trends
809 in mammalian classification. Mamm. Biol. - Zeitschrift für Säugetierkd. 78: 1-6.
810 Zachos JC, Pagani M, Sloan L, Thomas E, Billups K (2001). Trends, rhythms, and
811 aberrations in global climate 65 Ma to present. Science 292, 686–693.
812 Zachos FE, Lovari S (2013). Taxonomic inflation and the poverty of the Phylogenetic
813 Species Concept – a reply to Gippoliti and Groves. Hystrix (Online First).
814 Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations:
815 practice, problems and prospects. Mol Ecol 12:563–84.
816 Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013). A General Species Delimitation Method
817 with Applications to Phylogenetic Placements. Bioinformatics: 1-8.
818 Zink RM, Barrowclough GF (2008). Mitochondrial DNA under siege in avian
819 phylogeography. Mol. Ecol. 17: 2107–21.
820
821 TITLES AND LEGENDS TO FIGURES
822 **Fig. 1** Geographic location of the *M. avellanarius* samples used in the study. The shaded zone
823 corresponds to the distribution area of the species. The symbols refer to Lineage 1 (●) and
824 Lineage 2 (★) in Figures 2, 4, 5. The black line represent the fictive contact or hybrid zone
825 between the lineages.
826
827 **Fig. 2** Bayesian tree summarizing the phylogenetic relationship among the studied
828 populations based on a) the mitochondrial cyt*b* dataset, b) the nuclear dataset (BFIBR,
829 APOB), c) the combined dataset (cyt*b*, BFIBR, APOB). Numbers indicated at the root of the
830 branches correspond to Bayesian Inference (BP) on the left and bootstrap support (BS) for
831 ML analyses on the right. Haplotype and alleles distributions are summarized in Table 1. The

832 black shades represent the numbers of putative species based on the Phylogenetic Species
833 Concept, PSC (Zhang *et al*., 2013) identified by the Poisson Tree Process Model (PTP) and
834 the number of putative species based on the Genetic Species Concept (Baker and Bradley,
835 2006).
836
837 **Fig. 3** Ultrametric tree obtained with BEAST on the mitochondrial haplotype dataset.
838 Numbers indicate the posterior mean estimates divergence time (Millions years, Mya) for the
839 mitochondrial sequence dataset with the values of the 95% of the highest posterior density
840 (HPD). Clusters corresponding to putative species (GMYC) (Pons *et al*., 2006), based on the
841 Phylogenetic Species Concept, PSC,are indicated in red. Haplotypes distributions are
842 summarized in Table 1.
843
844 **Fig. 4** a) A two-dimensional plot of the FCA performed using GENETIX and b) Estimated
845 population structure from Structure analyses for K=2. Each individual is represented by a thin
846 vertical line divided into K coloured segments that represent the individual's estimated
847 membership fractions in K clusters. Colours yellow and blue indicate the membership for the
848 Lineage 1 and 2 respectively.