Genetic structure of a European forest species, the edible dormouse (Glis glis): a consequence of past anthropogenic forest fragmentation?

J. R. MICHAUX1,2,*, H. HÜRNER1, B. KRYSTUFÉK3, M. SARÀ4, A. RIBAS5, T. RUCH6, V. VEKHNIK7 and S. RENAUD8

1Conservation Genetics Laboratory, Institut de Botanique, Chemin de la Vallée, 4, 4000 Liège, Belgium
2CIRAD/INRA UMR117 ASTRE, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France
3Slovenian Museum of Natural History, Presernova 20, SI-1000 Ljubljana, Slovenia
4Dipartimento STEBICEF, Via Archirafi, 18-90123 Palermo, Italy
5Museo de Granollers-Ciències Naturals C/Francesc Macià, 51 08402 Granollers, Spain
6Institute for Cell Biology and Neuroscience, Johann Wolfgang-Goethe Universität Frankfurt am Main, AK NCR, Siesmayerstr. 70, D-60054 Frankfurt am Main, Germany
7445362 Zhiguli State Nature Reserve, Bakhilova Polyana Village, Samara Oblast, Russia
8Laboratoire de Biométrie et Biologie Evolutive, UMR 5558, Université Lyon 1, CNRS, 69622 Villeurbanne, France

Received 17 July 2018; revised 8 October 2018; accepted for publication 8 October 2018

The genetic structure of forest animal species may allow the spatial dynamics of the forests themselves to be tracked. Two scales of change are commonly discussed: changes in forest distribution during the Quaternary, due to glacial/interglacial cycles, and current fragmentation related to habitat destruction. However, anthropogenic changes in forest distribution may have started well before the Quaternary, causing fragmentation at an intermediate time scale that is seldom considered. To explore the relative role of these processes, the genetic structure of a forest species with narrow ecological preferences, the edible dormouse (Glis glis), was investigated in a set of samples covering a large part of its Palaearctic distribution. Strong and complex geographical structure was revealed from the use of microsatellite markers. This structure suggests that fragmentation occurred in several steps, progressively splitting the ancestral population into peripheral isolated ones. The fact that this structure postdates post-glacial recolonization, together with dating based on microsatellite data, supports the hypothesis that the differentiation was recent, starting around 9000 years ago, and took place stepwise, possibly up to Medieval times. This complements a classic phylogeographical interpretation based on the effect of past climate change, and supports the role of anthropogenic deforestation as a trigger of recent intraspecific differentiation.


INTRODUCTION

Anthropogenic changes to the Earth’s biosphere are now so great and so rapid that the concept of a new geological epoch, the Anthropocene, has been proposed. Its start is commonly attributed to the first increase of atmospheric CO₂ and CH₄ concentrations during the 1800s above previous long-term stable values, a view that emphasizes the role of recent anthropogenic changes related to industrialization. Yet, defined by strong human impact on the environment, the Anthropocene might actually have started thousands of years ago (Ruddiman, 2003) as a result of the spread of agriculture and technological innovations in farming that resulted in massive deforestation in several
The impact of humans on the natural fauna might thus be much older than is usually considered when focusing on the effects of industrialization (Martin, 2005) and recent habitat destruction (although these actions are also of great concern because they are ongoing). There is increasing evidence that current habitat fragmentation is strongly affecting the genetic structure of animal populations (Gerlach & Musolf, 2000; Kyle & Strobeck, 2001; Wang & Schreiber, 2001; Hirota et al., 2004; Keller et al., 2004). In contrast, the potential impact of more ancient anthropogenic changes on the patterns of geographical differentiation of European wild species remains largely undocumented. The impact of deforestation (which encompasses habitat loss and fragmentation) can be assumed to be particularly strong for forest animals. Studies on such species could help to decipher the timing and processes relating animal differentiation to anthropogenic changes, although it is often difficult to distinguish between natural climate change and human-induced deforestation.

The edible dormouse (Glis glis) is a good candidate for such a study. Although occasionally gathering food in environments such as orchards (Carpaneto and Cristaldi, 1995), the species relies heavily on food and shelter in mature deciduous woodlands with mast-seeding trees such as beeches (Fagus spp.), oaks (Quercus spp.) and sweet chestnut (Castanea sativa) (Ruf et al., 2006). It is a selective feeder, concentrating on the most nutritious food sources available (Holísova, 1968; Gigirey & Rey, 1999; Hürner & Michaux, 2009). These specific trophic requirements limit its occurrence to high-diversity environments that provide a continuity of adequate food supplies throughout the year. It may even skip reproduction in years of low food availability (Bieber & Ruf, 2009; Krystufek, 2010, Marteau & Sarà 2015). The edible dormouse is also a gregarious animal with large patch requirements and high habitat fidelity (see Krystufek, 2010). Together, these narrow ecological preferences make it highly sensitive to habitat fragmentation (Mortelliti et al., 2008). Its large distribution area (Fig. 1), from the Baltic Sea in the north to the Mediterranean peninsulas (including Sicily) to the south, and from the Iberian region to the west to Russia and the Black Sea and Caspian coasts to the east (Krystufek, 2010), provides the opportunity to trace on a large geographical scale the potential impact of forest habitat structure on the evolutionary dynamics of the species.

Traditional phylogeographical interpretation would predict a differentiation during glacial/interglacial cycles, leading to the establishment of several main lineages in refuge areas, such as the Mediterranean peninsulas, more than 15 000 years ago. It is from these refuge areas that populations would have recolonized Northern Europe after the last glacial maximum (see Taberlet et al., 1998; Hewitt, 1999; Michaux et al., 2003, 2005; Boston et al., 2015).

The phylogeographical structure of the edible dormouse has been assessed by analysis based on mitochondrial DNA (mtDNA) variation (Hürner et al., 2010; Naderi et al., 2014). This has revealed several endemic lineages at the southern fringe of its distribution, corresponding to refuge areas, and a recent post-glacial recolonization of most of Eurasia from a single Mediterranean refugium located in the Iberian region or in northern Italy. This led to a very low level of genetic diversity throughout most of the recolonized area. However, the use of detailed molecular markers such as microsatellites may reveal the existence of more recent events.

Using microsatellites on samples covering the whole distribution area of the edible dormouse (Fig. 1), we characterized its genetic structure with the aim to identify (1) whether finer genetic structure can be seen, post-dating the post-glacial recolonization traced by mitochondrial markers, and (2) the main drivers and the timing of such late differentiation (if any).

MATERIAL AND METHODS
SPEAKS AND DNA ANALYSIS
Tissue samples from 395 edible dormice from 47 sampling localities distributed throughout the geographical range of the species were analysed (Supporting Information, Table S1; Fig. 1). These specimens were obtained from collaborators, museums and our own fieldwork. In Belgium, where this species is protected by law, hair samples were collected non-invasively under the authorization of the Ministry of the Environment. All samples used were recent tissue specimens stored in ethanol except for the Hungarian samples which were dried hairs. DNA was isolated using the Qiagen DNeasy Tissue kit following the manufacturer’s instructions. Samples were subject to genotyping at ten microsatellite loci using the multiplex sets defined by Hürner et al. (2009). PCR conditions were as reported by Hürner et al. (2009). Amplified DNA was analysed for length variations on an automated capillary ABI PRISM 310 Genetic Analyser (Applied Biosystems) using GENEMAPPER v.4.0 software (Applied Biosystems). To avoid rate errors, the dried hairs were analysed following the methods used for degraded DNA and the ‘multitube’ approach proposed by Taberlet et al. (1996), which requires that three positive PCRs are obtained for each locus and sample to be considered correct.

MICRO-CHECKER 2.2.3 (Van Oosterhout et al., 2004)
was used to detect stutter errors and to estimate the proportion of null alleles at each locus. Genotypes were then corrected accordingly. This method revealed no error rates for the Hungarian samples. Tests for linkage disequilibrium (10^4 permutations) were also performed using ARLEQUIN v.3.1, adjusting for multiple tests using sequential Bonferroni correction.

Clustering Analysis

We applied two complementary Bayesian clustering algorithms, namely STRUCTURE v.2.3.3 (Pritchard et al., 2000) and GENELAND v.4.0.3 (Guillot et al., 2012), to infer population structure (i.e. the number of clusters, $K$) and to assign individuals to clusters probabilistically based on each multilocus genotype and, for GENELAND, on each spatial origin. These two approaches assume that populations are panmictic units with distinct allele frequencies. The estimation procedure consisted of runs using trial values of the number of populations ($K$) and then comparing the estimated log probability of the data under each $K$. For STRUCTURE, we conducted a series of independent runs with different proposals for $K$, testing all values from 1 to 15. Each run used $5 \times 10^6$ iterations after a burn-in of $5 \times 10^5$, utilizing a model allowing for admixture and correlated allele frequencies. To ensure convergence of the Markov chain Monte Carlo (MCMC) procedure, we used ten replicates for each value of $K$ and then checked the results visually for consistency. Generally, the number of distinct populations ($K$) is estimated from the model of $K$ that maximizes the probability of the data, $\ln P(D)$, but for complex population structures a modal value of $\ln P(D)$ may not be apparent (Rosenberg et al., 2002). In such instances, the value of $K$ at the beginning of a ‘plateau’ of estimates of $\ln P(D)$ may be selected as this is the smallest value of $K$ that captures the major population structure of GLIS GLIS.
structure in the data set (Pritchard & Wen, 2003). We also used the ΔK method (Evanno et al., 2005) implemented with STRUCTURE HARVESTER (Earl & vonHoldt, 2012) to find the most likely value of K present in the dataset.

The second algorithm, implemented in GENELAND 4.0.3, differs from that of Pritchard et al. (2000) mainly by explicitly taking into account the spatial dependence of individuals expected for a species whose range is much larger than the average inter-generation movement of individuals. This model assumes no admixture and aims to infer and locate genetic discontinuities between populations in space from individual geo-referenced multilocus genotypes, while taking into account uncertainty in the location of sampled individuals (Guillot et al., 2012). All the parameters (including K) are co-estimated simultaneously by the MCMC algorithm. To optimize estimation of the parameters, a two-step procedure was applied: a first run to infer K and a second run with K fixed at the modal value to estimate the other parameters (mainly assignment of individuals to the inferred populations). The first step was replicated ten times to check for convergence, allowing K to vary from 1 to 15 clusters and using run parameters as follows: 10⁶ MCMC iterations, maximum rate of Poisson process fixed at 700, maximum number of nuclei in the Poisson–Voronoi tessellation fixed at 500, and an uncertainty associated with the spatial coordinates of 1 km. We used the Dirichlet model of allelic frequencies as this has been demonstrated to perform better than the alternative model (Guillot et al., 2005). We inferred the number of clusters (K) from the modal value of K for these ten runs, and then ran the MCMC procedure again 100 times with K fixed for this value, 1 × 10⁶ MCMC iterations and the other modelling options unchanged. We calculated the mean logarithm of the posterior probability of the data (PPD) for each of the 100 runs and selected the ten runs with the highest PPD. These ten runs were then post-processed (with a burn-in of 3 × 10⁴ iterations) to obtain posterior probabilities of population membership for each individual.

A visual output of the STRUCTURE and GENELAND results was generated using DISTRUCT (Rosenberg, 2004).

**DIFFERENTIATION BETWEEN POPULATIONS**

To determine the level of genetic differentiation at microsatellite loci, Fst (Weir & Cockerham, 1984) values among inferred clusters were calculated. A significance test was performed for Fst values after 10⁴ permutations in ARLEQUIN v.3.1 (Excoffier et al., 2005). To depict the overall relationship among all inferred clusters, we applied derived pairwise values of Fst to construct a neighbour-joining (NJ) tree with the program NEIGHBOUR in PHYLIP (Felsenstein, 2004). Moreover, to confirm the NJ tree obtained using Fst values, we used the program POPULATIONS 1.2.31 (Langella, 1999) using different genetic distances, the Da index (Nei, 1983) and the standard genetic distance Ds index (Nei, 1987). Takezaki & Nei (2008) suggest using these indices when constructing phylogenetic trees from microsatellite data.

**GENETIC DIVERSITY AMONG CLUSTERS AND BOTTLENECK ANALYSIS**

Allelic richness (Rs) corrected for differences in sample size (fewer than ten samples), observed (Ho) and expected (He) heterozygosity and Fis values were calculated within each inferred cluster using FSTAT v.2.9.3 (Goudet, 2001). The significance of Fis values was assessed by comparing the observed values against distributions obtained by randomization (10⁴ times) and adjusting for multiple tests using Bonferroni correction.

**ISOLATION BY DISTANCE ANALYSES**

Isolation by distance (IBD) among the identified clusters was tested by comparing pairwise geographical distance (log-transformed) with pairwise genetic distances, estimated as Fst/(1 − Fst) and Rat/ (1 − Rat) using ARLEQUIN and SPAGeDI 1.3 (Hardy & Veekemans, 2002), respectively. IBD was calculated to assess whether genetic differentiation was distributed according to geographical distance.

**TESTING ALTERNATIVE SCENARIOS BY ABC MODELLING**

Approximate Bayesian computation (ABC) analyses are useful when investigating complex population genetics patterns based on microsatellite markers (e.g. Bryja et al., 2010; Sunnåker et al., 2013; Ali et al., 2014; Olafsson et al., 2014).

A set of different logical scenarios was therefore created (Supporting Information, Fig. S1) and tested using the ABC procedure (Beaumont et al., 2002) in DIYABC v.2.0 (Cornuet et al., 2008, 2014). This analysis allowed us to determine potential relationships among different clusters, which were then interpreted as populations (Fig. S1), and to estimate the timing of their differentiation.

To test alternative sequences of divergence of the different clusters, 17 scenarios were constructed. The first three were based on the topology obtained based on Fst values, and Da and Ds genetic distances. The fourth scenario was constructed according to the progressive relationships among STRUCTURE clusters from the hypothesis of K = 2 to K = 12.
Thirteen further scenarios were built by rearrangement of branches with the ‘Fst’ scenario (as depicted in Fig. S1) used as a ‘template’, which was modified according to the geographical proximity of the different populations as well as the biology of the species.

For each of these 17 scenarios, $5 \times 10^5$ datasets were simulated to build a reference table. To check if the combination of scenarios and previous distributions of their parameters was able to generate datasets similar to the one observed, a principal components analysis (PCA) was performed on the first 10 000 simulated datasets of the reference table in the space of summary statistics. The reference table was produced and recorded during the first step of ABC analyses. Each row of this table corresponds to a simulated dataset and contains the parameter values used to simulate the dataset and summary statistics computed based on the simulated dataset. Parameter values are drawn from previous distributions (Cornuet et al., 2008). The summary statistics used were based on Fst values, mean numbers of alleles and mean genetic diversity. To determine the most likely scenario among the 17 tested scenarios, normalized Euclidean distances between each simulated dataset of the reference table and the observed dataset were then computed and the 1% closest simulated datasets were used to estimate the relative posterior probability (with 95% confidence intervals) of each scenario based on logistic regression.

The most likely scenario had the highest posterior probability and non-overlapping 95% confidence intervals. Finally, the posterior density distributions of the effective population size of each population as well as the split times between them were estimated from the most likely scenario. Power analyses were used to evaluate the level of confidence.

We estimated type-I and type-II errors following Cornuet et al. (2010), by simulating 500 pseudo-observed datasets displaying properties similar to the observed dataset under the best scenario, and we computed the proportion of cases in which this scenario did not display the highest posterior probability. We estimated the false positive error rate (type-II error) by simulating 500 pseudo-observed datasets under each alternative scenario and computed the proportion of cases in which the scenario with the highest posterior probability was incorrectly selected as the most likely.

The prior distributions of parameters (effective population size, timing of events, chosen mutation model parameters for the microsatellites) used in the ABC analyses are summarized in Table S2. These parameters were chosen according to previous studies on Glis (Hürner et al., 2010) as well as general mutation rates for microsatellite markers observed in rodents (Dietrich et al., 1992). Broad intervals were chosen for each parameter, so that uncertainties in these parameters would not affect the output of the analysis. To estimate the timing of differentiation events, we considered one generation per year, as generally considered for this species (Gazárková & Adamik, 2016).

### RESULTS

#### Clustering analysis

Using STRUCTURE, the increase in likelihood when $K$ (number of clusters) increases showed several important steps at $K = 2, K = 5, K = 8$ and $K = 11$ ($\Delta K$ method, Evanno et al., 2005) (Fig. 2A). This suggests that the populations are hierarchically partitioned into several clusters of genetically similar populations. However, when considering the probability of successive partitions of the data when $K$ increases (Fig. 2B), the estimated likelihood increases progressively and smoothly without reaching a clear plateau. This suggests a complex structure, although on this basis alone it is difficult to select a value of $K$ capturing the major structure in the data, a problem exemplified by the fact that the assignments of the studied specimens when $K$ was varied from 8 to 12 were often unclear (data not shown). The results of GENELAND, which considers the geographical locations of the samples, revealed a clearer picture. Indeed, posterior distributions of the estimated number of inferred clusters across ten replicates displayed a clear mode at $K = 12$ (Fig. 2C). The probability of membership of all specimens to these 12 clusters was obtained with DISTRACT based on the highest-probability run (of ten) at that value of $K$ (Fig. 3A). All individuals within each locality had a proportion of assignment to the inferred cluster of 100%, indicating that these results are robust. The geographical distribution of the mean probability per geographical area (Fig. 3B) showed that each cluster includes a set of geographically close populations. The 12 groups of populations corresponding to an individualized genetic cluster are: (1) North-Western Balkans (NWBa: Slovenia,
Figure 2. Estimation of the number of populations based on microsatellite data, analysed by STRUCTURE and GENELAND. A, differential variation of the probability of partitions into $K$ and $K + 1$ clusters (Evanno et al., 2005) obtained with STRUCTURE. Variation is calculated as $\Delta(\ln P(D)) = \ln P(D)_{K+1} - \ln P(D)_K$. B, probability of successive partitions of the data into an increasing number of clusters obtained with STRUCTURE. Probability $[\ln P(D)]$ is estimated as the mean of ten replicated runs. For both A and B, the variance estimate is represented, which was calculated after five independent runs, using STRUCTURE HARVESTER. C, posterior distributions of the estimated number of inferred clusters across ten replicates obtained with GENELAND, displaying a clear mode at $K = 12$.

Figure 3. A, partitioning of the population structure of the edible dormouse for the modal solution $K = 12$, based on GENELAND analysis of the microsatellite data. Each individual is represented by a thin vertical line divided into $K$ coloured segments representing the probability of membership of this individual to the $K$ clusters. The different colours represent the different geographical groups emerging from the analysis. Individuals are arranged based on their country of origin and are sorted with increasing latitude. B, geographical representation of the outcomes for 12 genetic clusters ($K = 12$).
north-eastern Italy, Croatia), (2) Belgium (BEL), (3) Italian peninsula (IT: north-western, central and southern Italy, Sardinia), (4) Balkans (BAL: Bosnia and Montenegro), (5) Southern Italy (CAL: Calabria), (6) Sicily and Aeolian Islands (SIC), (7) Eastern Europe (EEur: Russia and Latvia), (8) Germany (GER), (9) Northern Spain and French Pyrenees (PYR), (10) Western Central Europe (WCEur: south-east France/Switzerland), (11) Central Europe (CEur: Hungary, Czech Republic) and (12) South-Eastern areas (SEar: Turkey and Macedonia).

One exception to this geographical congruence emerged, between the genetically identified clusters and the sets of populations: the United Kingdom group (UK) appeared to be associated with the Italian peninsula group IT. This apparent geographical inconsistency is related to the recent artificial introduction of the edible dormouse to Great Britain. Hence, UK samples were not included in subsequent analyses.

**Differentiation between populations**

Divergence between populations was estimated first by Fst values (Table S3). All pairwise Fst values were significantly different from zero, indicating that the clusters represented differentiated populations. The highest Fst values were recorded in pairwise comparisons involving northern peripheral populations: Russia/Latvia (EEur: mean Fst = 0.58), Belgium (BEL) (mean Fst = 0.56) and Germany (GER) (mean Fst = 0.45).

NJ analysis (Fig. 4A) based on Fst values provided an assessment of the hierarchical organization of the 12 clusters.

Different genetic groups were resolved. The first included populations from the Mediterranean area: Italian Peninsula (IT, CAL), Sicily (SIC), Balkans (BAL) and South Eastern areas (SEar). The Central European populations (CEur) appeared to be relatively closely related to this group. A set of populations clustered close together at the centre of the tree: North-Western Balkans (NWBa), Western Central Europe (WCEur) and Pyrenees (PYR). The remaining populations diverged from this central group of populations with longer branches: Belgium (BEL) branching close to Pyrenees (PYR) and Western Central Europe (WCEur) populations; Germany (GER) genetic distances index. C, neighbour-joining tree using the standard Ds genetic distance index. Colours are the same as in Figure 3. Values at the nodes in B and C correspond to bootstrap support values calculated with the program POPULATIONS.
branching together with the North-Western Balkans (NWBa); and Eastern Europe (EEur) comprising a group on its own.

The NJ trees obtained based on Da and Ds distances provided two further topologies, which showed many common features with the Fst-based pattern, but also some noticeable discrepancies. The tree based on Da distances (Fig. 4B) provided clearer support for the association of the westernmost populations (PYR, WCEur, BEL). On this basis, the Mediterranean cluster was split into two clades, one including only animals from the Italian Peninsula (IT and CAL) and the other including apparently disjoint populations (SIC, BAL and EEur). The German (GER) population was associated with the Central European one (CEur). NWBa again appeared close to the centre of the tree. Finally, the tree based on Ds distances (Fig. 4C) was quite similar to that obtained based on Da distances. The Western clade (WEur, PYR and BEL) again emerged and NWBa was close to the centre of the tree as in other topologies; GER was associated with CEur as seen based on Ds distances. Discrepancies were mostly found regarding Southern Europe, with some similarities to the Fst tree and others to the Dst tree.

IT and BAL appeared isolated but branched close to the North Western Balkan (NWBa) population. CAL, SIC, EEur and SEar clustered within a clade corresponding to the Southern and Eastern fringe of the geographical range. However, bootstrap values characterizing these last nodes were low and the difference among the trees should be taken with caution.

Isolation by distance

IBD was not detected among the 12 populations using both Fst and Rst values ($R^2 = 0.11$, $P = 0.36$; and $R^2 = 0.10$, $P = 0.29$, respectively). This suggests that despite the geographical coherence of the clusters of populations, the genetic structure observed in our dataset cannot be explained merely by geographical distance, but rather corresponds to patches of interconnected populations isolated from each other.

Genetic diversity within populations

Expected and observed heterozygosity and Fis values were estimated for all populations except Sardinia, which was too poorly sampled ($N = 2$, Table S1). The majority of the sampled populations conformed to Hardy–Weinberg (HW) equilibrium at most of the loci after sequential Bonferroni correction. The locus/population combinations that were not in HW equilibrium (heterozygote deficiency) were not specific to a particular population or locus. We therefore decided to maintain all of the loci in further analyses.

There was no evidence of linkage disequilibrium after Bonferroni correction. These results confirm that our dataset is compatible with the clustering approach, the assumptions of which require that clusters are sampled from discrete genetic units at HW and linkage equilibrium.

Mean allelic richness ($Rs$) ranged from 1.8 for BEL to 6 for SIC (Table 1). Average observed heterozygosity within populations varied from 0.14 to 0.63 and expected heterozygosity within populations from 0.18 to 0.75 (Table 1).

Multivariate analysis of the microsatellite data

Peripheral populations (BEL, GER and EEur) appeared to be the most divergent on the two principal axes of the PCoA, in total representing more than 38% of the variance (Fig. 5). Furthermore, along the first axis, a gradient emerged contrasting the western (PYR, WCEur and BEL) and eastern groups (EEur as the most extreme; SEar, CAL, SIC and, to a lesser extent, BAL). Geographically intermediate locations were also intermediate along the first PCoA axis (IT close to the UK, CEur and NWBa populations). Overall, the structure that emerged was similar to those retrieved from Fst, Da and Ds distances, but there was differentiation of peripheral populations.

Population history inferred by ABC analysis

The Bayesian approach allowed us to make inferences regarding the reconstruction of population history based on several steps. First, a PCA performed on the first 10 000 simulated datasets of the reference table in the space of summary statistics revealed

Table 1. Estimates of multilocus genetic variation for all groups comprising five or more individuals

<table>
<thead>
<tr>
<th>Group</th>
<th>$N$</th>
<th>Rs</th>
<th>Ho</th>
<th>He</th>
<th>Fis</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEL</td>
<td>70</td>
<td>1.8</td>
<td>0.18</td>
<td>0.18</td>
<td>0.199</td>
</tr>
<tr>
<td>PYR</td>
<td>70</td>
<td>3.6</td>
<td>0.37</td>
<td>0.51</td>
<td>0.283</td>
</tr>
<tr>
<td>WCEur</td>
<td>30</td>
<td>3.4</td>
<td>0.45</td>
<td>0.51</td>
<td>0.111</td>
</tr>
<tr>
<td>GER</td>
<td>26</td>
<td>2.6</td>
<td>0.23</td>
<td>0.33</td>
<td>0.367</td>
</tr>
<tr>
<td>EEur</td>
<td>65</td>
<td>2.2</td>
<td>0.14</td>
<td>0.19</td>
<td>0.255</td>
</tr>
<tr>
<td>CEur</td>
<td>10</td>
<td>3.7</td>
<td>0.43</td>
<td>0.56</td>
<td>0.234</td>
</tr>
<tr>
<td>CAL</td>
<td>7</td>
<td>5.1</td>
<td>0.51</td>
<td>0.64</td>
<td>0.304</td>
</tr>
<tr>
<td>BAL</td>
<td>8</td>
<td>3.6</td>
<td>0.40</td>
<td>0.42</td>
<td>0.059</td>
</tr>
<tr>
<td>UK</td>
<td>11</td>
<td>3.4</td>
<td>0.56</td>
<td>0.57</td>
<td>0.122</td>
</tr>
<tr>
<td>NWBa</td>
<td>59</td>
<td>3.6</td>
<td>0.41</td>
<td>0.45</td>
<td>0.165</td>
</tr>
<tr>
<td>SEar</td>
<td>5</td>
<td>2.0</td>
<td>0.26</td>
<td>0.27</td>
<td>0.012</td>
</tr>
<tr>
<td>SIC</td>
<td>19</td>
<td>6.0</td>
<td>0.48</td>
<td>0.75</td>
<td>0.358</td>
</tr>
</tbody>
</table>

$N$, number of specimens per group; Rs, allelic richness corrected for sample size; Ho, observed heterozygosity; He, expected heterozygosity; Fis, inbreeding coefficient.
that the observed dataset was encompassed by simulated datasets, indicating that our model was able to produce datasets similar to the one observed (Fig. S2). The final step of our Bayesian approach was a set of logistic regressions, which were performed on the closest 1% of the simulated datasets. This allowed us to unambiguously distinguish the most likely of the tested scenarios. The observed topology based on Ds distances (i.e. scenario 3 in Fig. 4C) was statistically the most robust, supported by posterior probabilities three times higher than the second ranked scenario, based on Da distances (Ds-based topology: $P = 0.579 \pm 0.322–0.669$ vs. Da-based topology: $P = 0.212 \pm 0.197–0.265$) whereas all other topologies had a $P$-value < 0.05. (Supporting Information, Fig. S3). Power analyses based on test datasets simulated under the 17 scenarios indicated that, given the size and polymorphism of the dataset, the method has high power to distinguish between the competing scenarios tested. More precisely, we estimated the performance of the analysis: of the thousands of pseudo-observed datasets generated using scenario 3, 46.4% were correctly identified as having been generated under this model. A moderate proportion (10.2%) were incorrectly attributed under scenario 2. The type-II error rate was low for the datasets generated under most scenarios (Table S4).

These results provide estimates of the periods of differentiation (Table S5) that would have led to the present genetic structure of the edible dormouse (Fig. 6A). Although these estimates should be considered with caution given that the population of origin may not have been panmictic, as postulated in our analyses, they do suggest four successive periods of differentiation, all clearly after the last Ice Age, which corroborates the results obtained with mtDNA (Hürner et al., 2010). These periods are:

1. an initial step of differentiation, at the beginning of the Holocene (around 9000 years ago), separating four ancestral stocks approximately corresponding to South and East Europe, Balkans, Northern Central Europe, and Central to Western Europe (Fig. 6B);

2. a succeeding split of these ancestral stocks between 6000 and 7000 years ago, especially affecting North Western Balkans (NWBa) and Eastern areas;

3. a third phase of differentiation during the Neolithic (between 5000 and 4000 years ago), which drove the separation among (i) populations from Calabria (CAL) and Sicily (SIC), (ii) populations from Germany (GER) and Central Europe (CEur), and (iii) populations from continental Italy (IT) and the Western European clade; and

4. finally the last period of differentiation during Medieval times (between the 5th and 11th centuries AD), driving the separation of the Western European group [Belgium (BEL), Pyrenees (PYR), West Central Europe (WCEur)].
DISCUSSION

The present study represents the most thorough research on the population genetics of the edible dormouse. Previous analyses based on mtDNA (Hürner et al., 2010; Naderi et al., 2014) have shown a Europe-wide genetic homogeneity for the species, except at the southern fringe of its distribution (Italian peninsula, Sicily, the Balkans, the Middle East, Iran), locations characterized by a slightly higher genetic diversity and which perhaps acted as refuges for this species.

Results suggest that Quaternary glaciations led to the differentiation of multiple lineages within this species, as in other rodents such as the hazel dormouse (Muscardinus avellanarius; Mouton et al., 2012, 2016), woodmice (Michaux et al., 2003, 2005; Herman et al., 2017) and bank voles (Deffontaine et al., 2005; Filipi et al., 2015), but that only lineage contributed to the post-glacial recolonization of the rest of Europe.

Microsatellite markers, having a higher evolutionary rate than mtDNA, have the potential to provide evidence of more recent and precise diversification events. Indeed, beyond the genetic homogeneity within the edible dormouse in most of Europe shown by mtDNA analysis, the present analysis revealed a recent complex evolutionary history. Even allowing for the methodological limits for the estimates of divergence time, the microsatellite results support the hypothesis that the basal divergence between the Western, Central and Eastern European groups occurred after the last glacial maximum (22,000–15,000 years BP) (Fig. 6A; Table 2). Accordingly, we propose the following scenario of genetic differentiation of the edible dormouse populations in Europe. First, an initial split of the main lineages took place during the mid-Holocene (between 9000 and 8000 years BP); second, differentiation in the South-Eastern region between 7000 and 6000 years ago; third a further split in different regions of Europe, including Central and South-Western Europe, occurring between 5000 and 4000 years ago; and finally, a phase of local divergence within Western European regions as late as Medieval times (1500–900 years BP).

The edible dormouse is a thermophilous rodent that is highly dependent on mature deciduous woodlands. As such, it was presumably severely affected by the climatic fluctuations of the Quaternary that drove it from its preferred habitat into a few refugia, namely the classic European refuge areas (Sommer & Nadachowski, 2006) – Iberia, Italian peninsula, the Balkans and the Hyrcanian region – and led to the differentiation of mtDNA lineages (Hürner et al., 2010; Naderi et al., 2014). These southern lineages, however, did not contribute to the recolonization of most of Europe or even Eurasia. Rather, recolonization occurred from a single refuge in the Iberian region or northern Italy, where most mtDNA diversity occurs.

Figure 6. A, representation of the best scenario corresponding to the neighbour-joining (NJ) tree obtained with Ds values and validated using ABC analyses. Split times corresponding to the mean values of the posterior probabilities (see Table 2) are given at each node. The grey colour corresponds to the first period of diversification (8500–9000 years BP), pink to the second one (6000–7000 years BP), blue to the third one (4000–5000 years BP) and red to the last period of differentiation (900–1500 years BP). B, representation on a European map of the best scenario of splitting events between populations (corresponding to the NJ tree based on Ds values). The four main genetic groups, which differentiated during the first period of diversification (8500–9000 years BP), are represented as shaded rectangles in A and shaded polygons with the same colours in B. The names of the studied populations are also given (BEL, PYR, WCEur, GER, EEur, NWBa, CEur, IT, CAL, SIC, EEur, BAL). Colours are as for Figures 3 and 4.
today and where those haplotypes characterizing all northern European populations (mtDNA lineage 3) are present (Hürner et al., 2010). In agreement with this, fossil evidence of edible dormice has been recorded in the Iberian Peninsula for the last glacial maximum (Sommer & Nadachowski, 2006). The edible dormouse shows low dispersal and is a specialist feeder that is strongly adapted to temperate forests and is temporally and spatially dependent on seed-mast crops (Bieber, 1998; Scinski & Borowski, 2008; Bieber & Ruf, 2009). This specialization may have made it prone to random extinction in an unstable environment such as that of northern Europe at the end of the last glacial maximum, explaining its absence during the coldest periods from almost all regions (Sommer and Nadachowski, 2006).

The timing of differentiation (between 9000 and 4000 years BP) between the southern groups (IT, CAL, BAL, NWBa) based on microsatellite data, as well as their very weak genetic differentiation based on microsatellite markers (see PCoA), however, contradicts with the persistence in these areas of mtDNA genetic lineages inherited during glacial/interglacial cycles (Hürner et al., 2010).

While the haploid nature and uniparental transmission of mtDNA can contribute to stronger genetic drift at local or regional geographical scales, these results could also be explained in part by female philopatry (Weber et al., 2018) and male-mediated dispersal.

Indeed, this can explain the strong structure of female-linked markers, while differentiation of biparentally transmitted loci is homogenized by male-mediated gene flow (Fahey et al., 2014). Therefore, genetic exchange might have occurred among the southern populations soon after the last glacial maximum, blurring the signal of glacial/interglacial structure at the nuclear level, but still maintaining the high mtDNA genetic structure observed in these populations, especially given that mtDNA markers evolve much more slowly than microsatellites. Hence, despite the high Fst values, the observed interpopulation differentiation is quite recent.

CURRENT GENETIC STRUCTURE OF THE EDIBLE DORMOUSE: A CONSEQUENCE OF EARLY HUMAN DEFORESTATION?

A striking feature emerging from our results is that all phases of fragmentation between populations of the edible dormouse clearly postdated the last ice age, ranging up to the mid-Holocene and Medieval times. This agrees with the Europe-wide homogeneity evidenced by more slowly evolving markers (mtDNA), which suggest range expansion after the last glacial maximum (Hürner et al., 2010). The relative effects of climate change and human impact on vegetation during the Holocene are difficult to disentangle (Jalut et al., 2009; Marquer et al., 2017). Climate might have been the primary driver of vegetation changes at the beginning of the Holocene (Jalut et al., 2009; Marquer et al., 2017), but later, human impact progressively increased (Roberts et al., 2018).

Hence, climate change might at first have favoured an expanded range of the edible dormouse, since the area of forest in Europe was at a maximum during the mid-Holocene thermal optimum (Roberts et al., 2018). Thereafter, the effects of climate change and human activity became intermingled, with increasing human impacts during the Late Holocene (Marquer et al., 2017; Roberts et al., 2018). In particular, the phases of population splits occurring after ~5000 years BP correspond roughly to periods when land use became a major driver of vegetation change (Marquer et al., 2017). Because deciduous forests, the preferred habitat of the edible dormouse, were often associated with land that was suitable for conversion to agriculture (Roberts et al., 2018), they were especially impacted by this increase in human land use.

Human impacts on the distribution of forest occurred in different phases. The Neolithic in Europe, corresponding to the transition from hunter-gatherers to the first farmers, spread from the Near East into south-eastern Europe as early as ~9500 years BP and progressed northward and westward along the Mediterranean coast and across the continent (Danubian stream), a process completed on the continent by 7500–7000 years BP (Zohary & Hof, 1993; Tresset & Vigne 2011). Splits among edible dormouse populations started around 9000 years BP in south-east Europe, progressing thereafter westward and northward, and these were more complex in areas such as the Balkans, which were affected early during the Neolithic. Overall, this suggests that the isolation of different populations of edible dormice and their progressive genetic differentiation might have been due to a combination of climate change, especially for the early differentiation events, and increasing human impact on the landscape, starting soon after the Neolithic switch to farming practices.

The latest phase of differentiation, corresponding to a split of Western European populations, seems to have occurred during Medieval times, a period known for intense deforestation related to monastic activity (Gimpel, 1976; Dubois, 1989; Bechmann, 1990). Together, the divergence time estimates suggest that early anthropogenic impact might have severely affected the distribution of the edible dormouse. This is corroborated by fossil records for Iberia, which show continuous occupation after its recolonization following the last glacial maximum, but its disappearance as a result of anthropogenic change during the Bronze Age, ~3000 years ago (Cuenca-Bescos et al., 2009). A similarly recent structuring of populations of edible dormice has
been suggested to have occurred along the Black Sea coast in relation to partly anthropogenic fragmentation of forest habitats (Helvaci et al., 2012).

If confirmed, this dynamic population differentiation linked to forest fragmentation is probably ongoing, as shown by the congruence between the distribution of population clusters of the edible dormouse and areas with remnant patches of primary forests in Europe (Sabatini et al., 2018). The rarity of such preserved habitats may contribute strongly to the rarity of the edible dormouse in the northern part of its distribution, where it is of conservation concern. Three effects may have contributed to the low genetic diversity in these remaining populations: low initial diversity related to post-glacial recolonization from a single source population in peripheral environments; a location at the leading edge of the expansion, which may also explain the lower genetic diversity compared to other populations; and the creation of forest fragments that probably shrank over time in highly populated areas such as Belgium and Germany. Restricted gene flow between populations possibly promoted genetic drift and led to high genetic differentiation among populations (previously unsuspected). However, this fragmentation increased inbreeding, with an associated loss of genetic diversity, reducing the capacity of small, isolated populations to adapt to environmental change (Saccheri et al., 1998; Frankham et al., 2002).

12.30
ISLANDS: SINK OR HOTSPOTS OF DIVERSITY?
Insular populations would appear to be particularly sensitive to extinction and, hence, to be less likely than continental counterparts to survive in the long term, as their genetic diversity is generally lower (Frankham, 1997). Nevertheless, two insular populations included here (Sicily and Great Britain) displayed surprisingly high genetic diversity (Table 1), suggesting unique evolutionary histories for these populations of dormice. The role of human activities is clear for the English population, which comprises specimens that were deliberately introduced at the beginning of the 20th century (Morris, 1997). In contrast to previous hypotheses, which suggested the origin of this population in Central Europe (Hungary, Germany or Switzerland) (Harris & Yalden, 2008), our results suggest strongly that the origin of these animals is continental Italy. As this last region is considered the main glacial refugium for the species and as it is characterized by having a high level of genetic diversity, it is not surprising that the British population also appears to be genetically diverse.

In Sicily, a high level of genetic diversity is preserved in various species of mammals, and more generally among vertebrates and invertebrates, despite its isolation from the continent (e.g. Michaux et al., 1998, 2003; Nieberding et al., 2006; Stöck et al., 2008). This unique phylogeographical history is related to the Mediterranean Sea, which (1) buffered the impact of the last ice age and allowed the island to play a role as a refugium during the last ice age, and (2) impeded the invasion of many competitors and predators to the island (Michaux et al., 2002). Moreover, for a small rodent like the edible dormouse, the large area of Sicily would be enough to sustain important population sizes and maintain a high level of genetic diversity.

12.60
CONCLUSION
The significant role of early human activity on wild fauna has long been appreciated, especially with regard to extinctions of emblematic species such as those belonging to the Pleistocene megafauna (Alroy, 2001; Roberts et al., 2001), but also extinctions of less emblematic taxa of small rodents (Vigne & Valladas, 1996), and even microevolutionary adaptive processes in response to anthropogenic environmental perturbations (Bochaton et al., 2017). The present study further suggests a discrete but decisive early anthropogenic impact on modern faunas, namely of early human-driven deforestation on the genetic structure of a forest species. Increasing evidence indicates that humans strongly impacted their environment much earlier than currently accepted, being responsible in the early Holocene for an increased greenhouse effect (see Ruddiman, 2003, 2007, 2011) and for distinct landscape modifications (Kaplan et al., 2010; Klein Goldewijk et al., 2010). The present study suggests another aspect of this ancient anthropogenic impact on wildlife, namely profound changes in the population structure of some mammals, such as the edible dormouse. How this species will be affected by current environmental change remains of debate. Our results emphasize that conservation measures for this species should target connections between areas of deciduous forest to limit the impact of forest fragmentation. However, the edible dormouse may benefit from modern social, political and agricultural trends that are favouring the reforestation of European landscapes (Moreira & Russo, 2007; Sitzia et al., 2010; Garcia-Feced et al., 2011).

12.100
ACKNOWLEDGMENTS
We are grateful to Alec de Selliers de Morenville for financial support that enabled the creation of the bank of microsatellites. We are grateful to the many people who assisted with this study by providing edible dormouse samples from throughout Europe: P. Adamik, A. Arrizabalaga, S. Bertolino, P. Gauberville, R. Libois, P. Morris and P. Vogel as well as P. Chevret et al. (Saccheri et al., 2002). Moreover, for a small rodent like the edible dormouse, the large area of Sicily would be enough to sustain important population sizes and maintain a high level of genetic diversity.

© 2018 The Linnean Society of London, Biological Journal of the Linnean Society, 2018, XX, 1–16
for numerous discussions. H.H. was supported by a Belgian research fellowship from the FRIA (Fonds pour la Formation et la Recherche dans l’Industrie et dans l’Agriculture) and J.R.M. from the FNRs (Mandat ‘Directeur de Recherches’, Fonds National pour la Recherche Scientifique) and financial grants of the Belgian FNRs (crédits aux chercheurs à J.R.M.), and M.S. from the project supported by the Italian Ministry of Education, University and Research: PRIN 2010/2011 - 20108TZKHC “Climate change and animal populations: predictive models for mammals and birds”. V.A. Vekhnik was funded by a regional grant of the Russian Foundation for Basic Research for Samara oblast according to the research project N 17-44-630288. B.K. acknowledges financial support from the Slovenian Research Agency (research core funding no. P1-0255).

13.20

REFERENCES


© 2018 The Linnean Society of London, Biological Journal of the Linnean Society, 2018, XX, 1–16


© 2018 The Linnean Society of London, Biological Journal of the Linnean Society, 2018, XX, 1–16


**SUPPORTING INFORMATION**

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site.

Fig. S1. Seventeen evolutionary scenarios for the differentiation of European groups of edible dormice, tested using ABC analyses. The first three scenarios were based on the topology obtained by NJ analyses on (1) Fst values (Fig. 3A); (2) the Da genetic distance index (Fig. 3B); and (3) the Ds genetic distance index (Fig. 3C). The fourth scenario was constructed according to the results obtained with the STRUCTURE software. The tree topology was created according to the progressive relationships among clusters from the hypothesis of $K = 2$ to $K = 12$. To explore possible space around these scenarios, 13 further scenarios, which make sense based on the biology of *Glis glis* and its phylogeographical history, were built by rearrangement of the branches in scenario 1. The names of populations correspond to those described in Figures 3–6.

Fig. S2. Results of PCA obtained with the diYABC software, showing the fit of the observed data (in yellow) to the simulated fit (light blue for prior and black circled blue for posterior Bayesian distributions).

Fig. S3. Graph representing the logistic regressions performed on 1% of the closest simulated datasets, distinguishing unambiguously the most likely of all tested scenarios at each step. The third scenario is the most likely, i.e. that based on the observed topology resulting from Ds distance values (Fig. 3C).

Table S1. Reference number, country abbreviation, inferred populations and number of edible dormouse specimens collected in the present study.

Table S2. Prior distributions of parameters used in our ABC analysis.

Table S3. Genetic distances between edible dormouse populations. Pairwise Fst values (*Weir & Cockerham, 1984*) calculated in ARLEQUIN are provided below the diagonal. Statistical significance of the difference is provided above the diagonal ($**p < 0.00075$). N corresponds to the number of specimens per group.

Table S4. Performance analysis of the model choice procedure. $D =$ proportion of cases in which the simulation-based model choice procedure selected a scenario as the most probable with non-overlapping confidence intervals of the posterior probabilities of each scenario. *Type-I or α-error rate. *Type-II or β-error rate ($1 - \Sigma \beta_i$ is used to determine the power of the model choice procedure).

Table S5. Posterior probabilities obtained for the most likely scenario (1) (see figure in Appendix 1). The effective sizes [from Na (ancestral) to N12] of each studied population as well the time split among them (from tC1 to tC9) are given. Confidence intervals (CI) (95%) are also given for each parameter.

**DATA ARCHIVING**

The microsatellite dataset is available on the Dryad data repository under code XXXX.