Heredity (2014), 1–11
© 2014 Macmillan Publishers Limited All rights reserved 0018-067X/14
www.nature.com/hdy

ORIGINAL ARTICLE

Revisiting the phylogeography and demography of European badgers (*Meles meles*) based on broad sampling, multiple markers and simulations

AC Frantz[1,2], AD McDevitt[3], LC Pope[4], J Kochan[5], J Davison[6], CF Clements[1], M Elmeros[7], G Molina-Vacas[8], A Ruiz-Gonzalez[9], A Balestrieri[10], K Van Den Berge[11], P Breyne[11], E Do Linh San[12], EO Ågren[13], F Suchentrunk[14], L Schley[15], R Kowalczyk[16], BI Kostka[17], D Ćirović[18], N Šprem[19], M Colyn[20], M Ghirardi[21], V Racheva[22], C Braun[23], R Oliveira[24], J Lanszki[25], A Stubbe[26], M Stubbe[26], N Stier[27] and T Burke[1]

Although the phylogeography of European mammals has been extensively investigated since the 1990s, many studies were limited in terms of sampling distribution, the number of molecular markers used and the analytical techniques employed, frequently leading to incomplete postglacial recolonisation scenarios. The broad-scale genetic structure of the European badger (*Meles meles*) is of interest as it may result from historic restriction to glacial refugia and/or recent anthropogenic impact. However, previous studies were based mostly on samples from western Europe, making it difficult to draw robust conclusions about the location of refugia, patterns of postglacial expansion and recent demography. In the present study, continent-wide sampling and analyses with multiple markers provided evidence for two glacial refugia (Iberia and southeast Europe) that contributed to the genetic variation observed in badgers in Europe today. Approximate Bayesian computation provided support for a colonisation of Scandinavia from both Iberian and southeastern refugia. In the whole of Europe, we observed a decline in genetic diversity with increasing latitude, suggesting that the reduced diversity in the peripheral populations resulted from a postglacial expansion processes. Although MSVAR v.1.3 also provided evidence for recent genetic bottlenecks in some of these peripheral populations, the simulations performed to estimate the method's power to correctly infer the past demography of our empirical populations suggested that the timing and severity of bottlenecks could not be established with certainty. We urge caution against trying to relate demographic declines inferred using MSVAR with particular historic or climatological events.
Heredity advance online publication, 30 April 2014; doi:10.1038/hdy.2014.45

INTRODUCTION

The phylogeography of European mammals has been extensively studied since the 1990s and a broad consensus has been reached regarding the importance of the Mediterranean peninsulas (Iberia, Italy and the Balkans) as refugia during the Last Glacial Maximum (LGM) (Taberlet *et al.*, 1998; Hewitt, 1999). However, many of these findings were based on limited data sets in terms of both the number of samples and molecular markers used, ultimately leading to incomplete postglacial recolonisation scenarios. Increased sampling in particular has led to the identification of important refugia present further north in Europe, such as the Carpathian Basin (Kotlík *et al.*, 2006; McDevitt *et al.*, 2012). Therefore, it is clear that continent-wide sampling—including adequate sampling of putative refugia—is vital for accurate inference of the phylogeographic history of European species. Furthermore, over recent years phylogeographic studies have included data from multiple markers in order to gain new insights into complex colonisation histories (McDevitt *et al.*, 2011) and processes over different time scales (Avise, 2004).

In addition to broad sampling and the use of multiple markers, a number of methodological innovations can also contribute to a better

[1]NERC Biomolecular Analysis Facility, Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK; [2]Musée National d'Histoire Naturelle, Luxembourg, Luxembourg; [3]School of Biology and Environmental Science, University College Dublin, Dublin, Ireland; [4]School of Biological Science, University of Queensland, St Lucia, Queensland, Australia; [5]Department of Genetics and Animal Breeding, Wrocław University of Environmental and Life Sciences, Wrocław, Poland; [6]Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia; [7]Department of Bioscience, Aarhus University, Rønde, Denmark; [8]Animal Biology Department, University of Barcelona, Barcelona, Spain; [9]Department of Zoology, Biogeography and Population Dynamics Research Group, University of the Basque Country, UPV/EHU, Vitoria-Gasteiz, Spain; [10]Department of Biosciences, University of Milan, Milan, Italy; [11]Research Institute for Nature and Forest, Geraardsbergen, Belgium; [12]Department of Zoology and Entomology, University of Fort Hare, Alice, South Africa; [13]National Veterinary Institute, Department of Pathology and Wildlife Diseases, Uppsala, Sweden; [14]Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Vienna, Austria; [15]Administration de la nature et des forêts, Luxembourg, Luxembourg; [16]Mammal Research Institute, Bialowieza, Poland; [17]Queen's University Belfast, Northern Ireland, UK; [18]Faculty of Biology, University of Belgrade, Belgrade, Serbia; [19]Department of Fisheries, Beekeeping, Game Management and Special Zoology, University of Zagreb, Zagreb, Croatia; [20]CNRS, UMR 6553, ECOBIO, Université de Rennes 1, Rennes, France; [21]Università degli Studi di Torino, Torino, Italy; [22]Balkani Wildlife Society, Sofia, Bulgaria; [23]9 chemin du Kilbs, Bischoffsheim, France; [24]Departamento de Zoologia e Antropologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal; [25]Department of Nature Conservation, University of Kaposvár, Kaposvár, Hungary; [26]Domplatz 4, Halle/Saale, Germany and [27]Institute of Forest Botany and Forest Zoology, Dresden University of Technology, Tharandt, Germany
Correspondence: Dr AC Frantz, Musée National d'Histoire Naturelle, 25 rue Münster, Luxembourg, L-2160, Luxembourg.
E-mail: alainfrantz@yahoo.co.uk
Received 6 September 2013; revised 4 February 2014; accepted 14 February 2014

understanding of the factors influencing the broad-scale genetic structure of a species. Recent advances in approximate Bayesian computation (ABC), for example, have paved the way for robust comparison of phylogeographic/colonisation scenarios (Beaumont, 2010). In addition, likelihood-based methods paired with Monte Carlo sampling are becoming increasingly popular tools to estimate past demographic parameters (Girod *et al.,* 2011 and references therein), as standard summary statistic-based bottleneck tests have low power at the typical sample sizes of phylogeographic studies and their results can depend on the choice of mutation model parameters (Peery *et al.,* 2012). The full Bayesian model developed by Beaumont (1999) and Storz and Beaumont (2002), for example, infers posterior probability distributions of population parameters using information from the full microsatellite allelic distribution in a coalescent-based framework under a stepwise mutation model. The method by Storz and Beaumont (2002), implemented in the program MSVAR v.1.3, estimates the posterior distributions of three demographic parameters—the log of the current, $\log(N_0)$, and ancestral, $\log(N_1)$, effective population sizes and the log of time since the demographic change, $\log(t_a)$—assuming an exponential or linear decrease in a panmictic, isolated population.

The phylogeography of the European badger (*Meles meles*) is an example where extensive sampling, multiple markers and up-to-date analytical techniques might help to resolve outstanding issues. It is now well established that the European badger is one of the four Eurasian badger species (Del Cerro *et al.*, 2010 and references therein). The broad-scale genetic structure of the European species is of interest, as it may be the result of historic restriction to and expansion from glacial refugia and/or recent impact by extensive anthropogenic interference (Pope *et al.,* 2006). Indeed, badger densities declined in some European countries owing to the large-scale gassing of setts during the rabies control campaigns of the 1960s and 1970s (Griffith and Thomas, 1997). Consequently, a number of studies have attempted to describe the genetic variation of the species in a European context (Marmi *et al.,* 2006; Pope *et al.,* 2006; O'Meara *et al.*, 2012). However, they were essentially based on the same data set of samples from western and north–central Europe, with little to no sampling from eastern Europe and Italy, making it difficult to draw robust conclusions about the location of refugia, patterns of postglacial expansion and recent demography.

Although subfossil evidence suggests that the badger was present in Iberia, Italy, the Carpathians and the Balkans during the LGM (Pazonyi, 2004; Sommer and Benecke, 2004), Marmi *et al.* (2006) inferred the presence of only one European mitochondrial DNA (mtDNA) clade. However, despite the absence of samples from southeastern Europe, the haplotype network led these authors to suggest recolonisation of the continent from multiple refugia. Working with microsatellites, Pope *et al.* (2006) interpreted the presence of a broad-scale isolation-by-distance signal between western European populations to result from an expansion out of a single glacial refugium. These authors also found a higher genetic diversity in badgers from western central Europe, compared with populations from the southwest and the north but could not ascribe this pattern with confidence to either historic process or recent demographic declines. Although in principle, MSVAR allows recent and ancient population declines to be distinguished (see, for example, Goossens *et al.,* 2006; Heller *et al.,* 2008), the simulation study by Girod *et al.* (2011) suggested that both the accuracy and precision of the estimated demographic parameters were greatest for old and severe events. As suggested by Peery *et al.* (2012), it would be advisable to estimate a method's power to detect a bottleneck—or assess the likelihood of a false bottleneck signal—given realistic

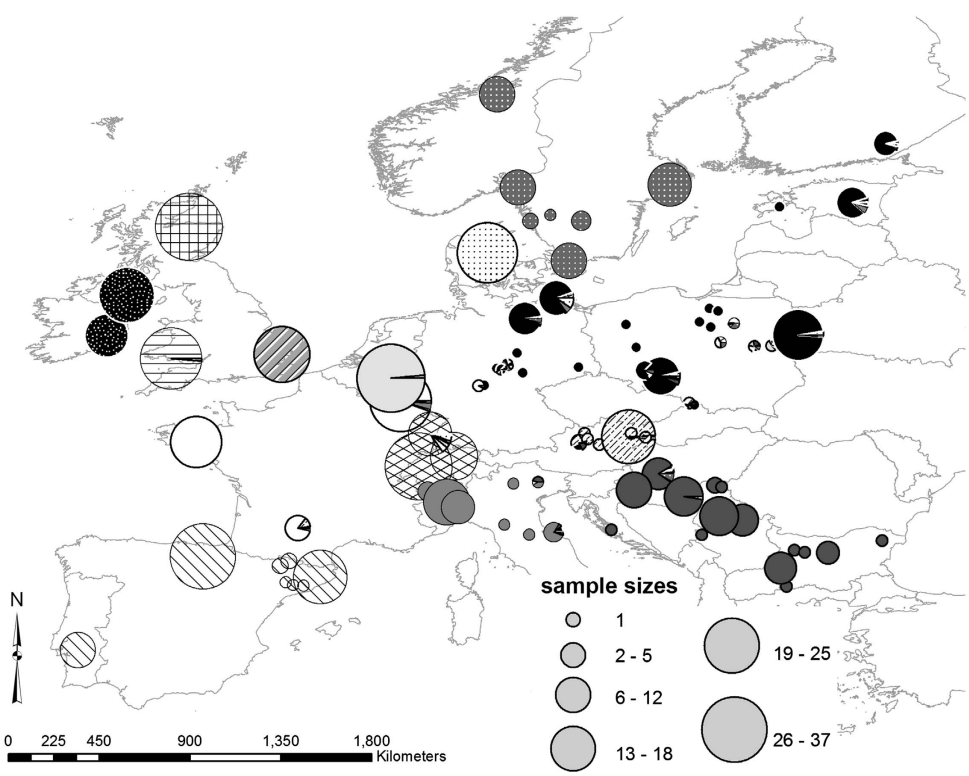

Figure 1 Sample distribution and location of the genetic subpopulations inferred using the BAPS algorithm. The size of the pie charts indicates the number of samples collected from a locality, whereas the pattern of the pie chart indicates the identity of the genetic clusters.

demographic scenarios and the available number of samples and genetic markers.

The overall objective of the current work was to gain a thorough understanding of the factors influencing the broad-scale genetic structure of the European badger. Based on continent-wide sampling, we specifically aimed to infer postglacial colonisation routes of the species in Europe, analysing mtDNA sequence data and microsatellite loci with traditional phylogeographic analyses coupled with ABC. Specifically, we wanted to test whether badgers recolonised Europe out of a single or from multiple refugia. Furthermore, we aimed to use microsatellite data to investigate the demography of European populations. Specifically, we wanted to test whether the reduced genetic diversity in peripheral badger populations was due to either historic or recent processes. Finally, we performed limited simulations to test the power of the MSVAR method to allow robust inference of past demographic events in the specific context of this study.

MATERIALS AND METHODS

Laboratory work

We collected muscle, ear or hair samples from a total of 675 badgers from 21 different European countries (Figure 1). Following geographic criteria, samples from five countries were further subdivided into different sampling locations, giving rise to a total of 30 predefined populations (Supplementary Table S1). Altogether, 177 samples (originating from Finland, Great Britain, Luxembourg, Norway, the Republic of Ireland and Zurich) had already been microsatellite genotyped for the study by Pope *et al.* (2006). Tissues samples from road-killed or legally hunted individuals were stored in absolute ethanol until extraction. DNA was extracted from tissue samples using an ammonium acetate-based salting-out procedure (Miller *et al.,* 1988) and from hair samples (3 localities, 43 samples) using a Chelex protocol (Walsh *et al.,* 1991). We used primers *Mel*CR1 and *Mel*CR6 (Marmi *et al.,* 2006) to amplify a 594-base pair (bp) fragment of the 5′-end of the control region of 327 individuals (Supplementary Table S1, Supporting Information, Supplementary Appendix S1). Samples were also genotyped using 18 microsatellite markers in three multiplex reactions using the QIAGEN Multiplex kit (QIAGEN, Hilden, Germany; see also Pope *et al.*, 2006). Further information on the multiplex composition and the PCR conditions are given in the Supporting Information, Supplementary Appendix S2. The genetic profiles of all samples consisted of at least 12 loci.

Data analysis: mtDNA

The number of polymorphic sites, as well as nucleotide and haplotype diversity within Europe overall, were calculated using DnaSP v.5 (Librado and Rosaz, 2009). Expected heterozygosity and haplotype richness were calculated for populations containing at least seven individuals in CONTRIB v.1.02 (Petit *et al.,* 1998). Furthermore, our own data set was supplemented with 82 European individuals from two previous phylogeographic studies of Eurasian badgers (Marmi *et al.,* 2006; Tashima *et al.,* 2011) and a phylogenetic network was constructed using the software NETWORK v.4.6 (www.fluxus-engineering.com) with a median-joining algorithm based on maximum parsimony (Bandelt *et al.,* 1999). Previous studies using multiple markers have demonstrated that European badgers are a monophyletic group (Marmi *et al.,* 2006; Tashima *et al.,* 2011), and under the circumstances of closely related sequences, simulation studies have demonstrated that this method provides reliable estimates of the true genealogy (Cassens *et al.,* 2005). Because of missing data, all the sequences were truncated to 467 bp.

Past demographic expansion in European badgers as a whole was tested using three methods implemented in ARLEQUIN v. 3.5.1.2 (Excoffier and Lischer, 2010): Fu's (1997) F_S statistic, Tajima's (1989) D statistic and mismatch distributions of pairwise nucleotide differences. Mismatch distributions were calculated and compared with expected values for an expanding population (Rogers and Harpending, 1992) by testing for goodness-of-fit statistics based on the sum of square deviations for a model of sudden expansion. Tests for goodness of fit for all three methods were generated using parametric bootstrapping with 10 000 replicates.

Data analysis: microsatellite loci

We estimated observed (Ho) and unbiased expected (He_u) heterozygosities (Nei, 1978) for the 25 predefined populations with 10 or more sampled individuals (Supplementary Table S1) using GENETIX 4.05.2 (Belkhir, 2004, unpublished software). We tested for the significance of heterozygote deficiency or excess in these 25 populations using the Markov chain method in GENEPOP 3.4 (Raymond and Rousset, 1995), with 10 000 dememorisation steps, 100 batches and 10 000 subsequent iterations. The populations were tested for linkage disequilibria among loci using an exact test based on a Markov chain method as implemented in GENEPOP 3.4. The false discovery rate technique was used to eliminate false assignment of significance by chance (Verhoeven *et al.,* 2005).

We used the STRUCTURE v2.3.1 (Pritchard *et al.,* 2000) and GENELAND v.4.0.3 (Guillot *et al.,* 2005) Bayesian clustering algorithms to analyse the population genetic structure of European badgers. To estimate the number of subpopulations (K), 10 independent runs of $K = 1$–15 were carried out with 10^6 Markov chain Monte Carlo (MCMC) iterations after a burn-in period of 10^5 iterations. Further details on the exact parameters used for running the programs can be found in the Supporting Information, Supplementary Appendix S3. We also used BAPS v5.4 (Corander *et al.,* 2004) to cluster the data at group level, as it partitions the sampling units into populations with nonidentical allele frequencies (Corander *et al.,* 2004). The program was run 10 times for each $K = 15$–25, using the 30 sampling locations (Supplementary Table S1) as predefined groups. We used GENETIX v.4.05.2 to perform a factorial correspondence analysis to visualise the genetic distance between the 30 predefined badger populations.

We used program ADZE v1.0 (Szpiech *et al.,* 2008) to calculate the allelic richness (A_R) of the 25 predefined populations with a sample size of $N \geqslant 10$, excluding genetic markers that had been typed at $<50\%$ of all individuals in at least one population (loci *Mel110* and *Mel112*). The estimates of allelic richness were based on a sample of 16 individuals in all 25 populations. We used regression analysis in R 2.13.0 (R Development Core Team, 2011) to test for the effect of both longitude and latitude on diversity measures based on microsatellites (He_u and A_R) and mitochondrial sequence data (haplotype richness and haplotype diversity). The analysis was performed for all 25 predefined populations, but also, in order to avoid issues of spatial autocorrelation, using only one population—the one with the largest sample size—per BAPS-defined genetic clusters (see Results). Geographic coordinates for each predefined population were obtained by averaging coordinates of individuals.

Phylogeographic reconstruction

The ABC was implemented in DIYABC v1.0.4.46 (Cornuet *et al.,* 2008) to further investigate the dynamics of the recolonisation process of badgers in Europe. This software can produce estimates of the relative likelihood of alternative phylogeographic scenarios in a coalescent framework. Here we used a relatively simplified situation, with three 'lineages' of interest: Iberia, Scandinavia and the Balkans. We aimed to clarify the phylogeographic origin of the Scandinavian populations (see Results).

The ABC analysis was performed with the microsatellite data only, as preliminary analyses with a combined mtDNA and microsatellite data set did not yield satisfactory support (admixture between groups was one of the scenarios to be tested). The microsatellite data set consisted of 74 individuals from Spain and Portugal (Iberia), 51 individuals from Norway and Sweden (Scandinavia) and 73 individuals from Serbia, Croatia, Hungary and Bulgaria (southeast Europe). After preliminary analyses, effective population sizes were allowed to vary between 10 and 10 000 for Scandinavia and Iberia, as well as between 10 and 20 000 for the Balkans. In scenario 1, the Scandinavian lineage split from the Iberian lineage at $t1$ (13 000 years to 1 year before present (BP)) to coincide with the emergence of various landbridges connecting mainland Europe to Scandinavia (Björck, 1995), the first appearance of badger fossils in Scandinavia (Sommer and Benecke, 2004) and to allow this event to have occurred any time since then (Supplementary Figure S1). In scenario 2, the Scandinavian lineage split from the Balkan lineage at $t1$. In scenario 3, each of the Iberian, Scandinavian and Balkan 'lineages' coalesced at $t2$ (LGM; 19 000–26 000 BP) and had independent histories since. Finally, scenario 4 has

the Scandinavian lineage forming as a result of admixture between the Iberian and Balkan lineages at $t1$ (Supplementary Figure S1).

Simulated data sets were created by requesting a total of 21 summary statistics, including the number of alleles and heterozygosity (per population and per pairs of populations), Garza and Williamson's M-ratio (per population) and both F_{ST} and $\delta\mu^2$ pairwise divergence statistics. One million simulated data sets per scenario were used to produce posterior distributions. Each scenario was considered equally probable and reliability of scenarios was visualised through principal component analysis, whereas posterior probabilities of scenarios were compared by means of logistic regression, using the closest 1% of simulated data sets to the observed data (Cornuet *et al.,* 2008).

Demographic reconstruction

We used the method by Storz and Beaumont (2002), implemented in the program MSVAR 1.3, to infer past population dynamics of the badger populations. MSVAR 1.3 assumes that a stable population of size N_1 started to either increase or decrease a time of t_a ago to a current population size N_0. The loci are assumed to be evolving according to a strict single-step mutation model with a mutation rate μ. Prior distributions for the parameters are assumed to be log-normal. The means and s.d. of these prior log-normal distributions are themselves drawn from prior (or hyperprior) distributions. Hyperpriors for the means were specified by normal distributions with a mean of α and a s.d. of σ. Hyperpriors for the standard deviations were assumed to be zero-truncated normal distributions with a mean of β and a s.d. of τ (see Supplementary Table S2).

We applied MSVAR to the 19 predefined populations with a sample size of $N \geqslant 20$. Two loci (*Mel115* and *Mel14*) had repeat length variations that were not a consistent multiple of two or four and were therefore excluded from the MSVAR analyses. We modelled an exponential change in population size as well as a generation time of 1 so that the estimate of $\log(t_a)$ should be indicative of the number of generations since the change in population size. Initially, 12 independent runs were performed using different random seeds, starting values, priors and run lengths (Supplementary Table S2). We discarded the first 10% of each MCMC chain to avoid biases due to starting conditions.

Convergence among chains was tested with the Gelman and Rubin (1992) statistic using the CODA library (Plummer *et al.,* 2006) of the R package. A point estimate of <1.1 is normally taken as an indicator of good convergence (Gelman *et al.,* 2004), with a value of <1.2 sometimes being used as a guideline for approximate convergence (see, for example, Brooks and Gelman, 1998). As MSVAR has been shown to reach convergence with difficulty, especially with recent and severe bottlenecks (Girod *et al.,* 2011), we also considered chains that only converged approximately as giving rise to informative point estimates. If the 12 independent chains did not converge, we ran 2 further chains for a total of 10^{10} steps (Supplementary Table S2) and, in the case of successful convergence, limited the inference to the 3 longer chains. Because of the computational complexity of MSVAR, running a larger number of longer chains was impractical (see also Girod *et al.,* 2011). Independent runs were pooled into one data set to produce larger samples of the posterior distribution. The marginal posterior distributions of the model parameters were estimated using the R library LOCFIT (Loader, 1999). Point estimates of $\log(N_0)$, $\log(N_1)$ and $\log(t_a)$ were obtained from the mode of their marginal posterior distribution. The 90% highest probability density intervals were obtained with the CODA package.

We performed a limited number of simulations to assess the power of the method by Storz and Beaumont (2002) to correctly infer the past demography of the different badger populations, given the sample sizes of the empirical data sets and the number of genetic markers used. We used the program DIYABC to simulate nine different scenarios of population decline. We fixed the number of simulated microsatellites to 16 (evolving according to a strict stepwise mutation model) and the ancestral effective population size to $N_1 = 5000$, but varied the current effective population size N_0 (50, 150, 1000) and the generation time since the decline t_a (25, 100, 500). For each combination of these parameters, we generated five genetic data sets consisting of 20 individuals and five data sets of 50 individuals. We analysed these simulated data using the same methodology as used for the empirical data. If approximate convergence was not achieved for the $\log(N_0)$ estimate, we simulated and analysed a new data set to ensure a balanced sampling design for the statistical analysis. We performed three-way analysis of variance in R to test for the influence of simulated sample size, current effective population sizes and times since decline on the accuracy of the MSVAR point estimates for $\log(N_0)$ and $\log(N_1)$. The difference between the simulated and estimated effective population size was log-transformed to improve normality.

RESULTS

Altogether, 49 haplotypes were observed in the control region data set, with 28 of these newly described in this study (GenBank accession nos.: KJ161328–KJ161355). A total of 27 polymorphic nucleotide sites were found, of which 16 were parsimony informative. Overall nucleotide and haplotype diversities were equal to 0.00591 (s.d. $\pm$ 0.00018) and 0.894 (s.d. $\pm$ 0.01), respectively. The median-joining network did not reveal clear genetic structuring within European badgers (Figure 2), with closely related haplotypes differing only by a single bp (Figure 2). A star-like pattern was found with core haplotypes, meles1–3, distributed throughout eastern and central Europe, the Balkans and Britain and a newly identified haplotype mm20 found in the Balkans, Poland and Estonia. However, a group of haplotypes—mm1–4, mm6–14 and meles15–20—were largely restricted to Iberia (with a few occurrences in Western Europe). Furthermore, haplotypes meles12–14 were almost exclusively Scandinavian (Sweden and Norway). Most Irish individuals also had these 'Scandinavian' haplotypes (and mm4). These 'Scandinavian' haplotypes formed a third group in the median-joining network that was more closely related to the Iberian than the eastern haplotype groups. 'Scandinavian' haplotypes also occurred in Finland, Estonia and central Russia. The entire European data set conformed to a model of demographic expansion (sum of square deviations: 0.0067; $P = 0.42$) and was also consistent with a model of spatial expansion (sum of square deviations: 0.0079; $P = 0.21$). Fu's (1997) F_S value was significant (-26.08331; $P < 0.001$) but Tajima's (1989) D was not (-0.88218; $P = 0.19$). The mismatch distribution revealed that there was evidence of a bimodal distribution of pairwise differences (Supplementary Figure S2).

Although the microsatellite He_u of the 25 predefined populations with $N \geqslant 10$ was almost always larger than Ho (Supplementary Table S1), no locus systematically deviated from Hardy–Weinberg equilibrium after correcting for multiple tests (Supplementary Table S3). Three different pairs of loci were in linkage disequilibrium in three predefined populations after correcting for multiple tests. All loci were therefore included in the analyses. The results of the three clustering methods did not converge on one optimal solution. In the case of STRUCTURE, $K = 10$ was the highest K at which the log-likelihood values converged reasonably well (Supplementary Figure S3), whereas GENELAND inferred the presence of only nine genetic populations. The badgers in Iberia, Ireland, Britain, Denmark and Scandinavia always formed separate clusters in both analyses (Supplementary Figure S4). GENELAND also inferred the presence of a separate Scottish population. However, the composition of the clusters in the rest of mainland Europe differed between STRUCTURE runs and between both methods (Supplementary Figure S4). The BAPS algorithm, run by clustering 30 predefined groups, inferred the presence of 14 clusters in the optimal partition and appeared to resolve the incompatibilities between the different STRUCTURE and GENELAND runs (compare Figure 1 and Supplementary Figure S4). It also identified Iberia, Ireland, Denmark and Scandinavia as being genetically distinct. Furthermore, it split the mainland data set into a further seven geographically coherent clusters and classified the three sampling locations in Britain as being genetically different.

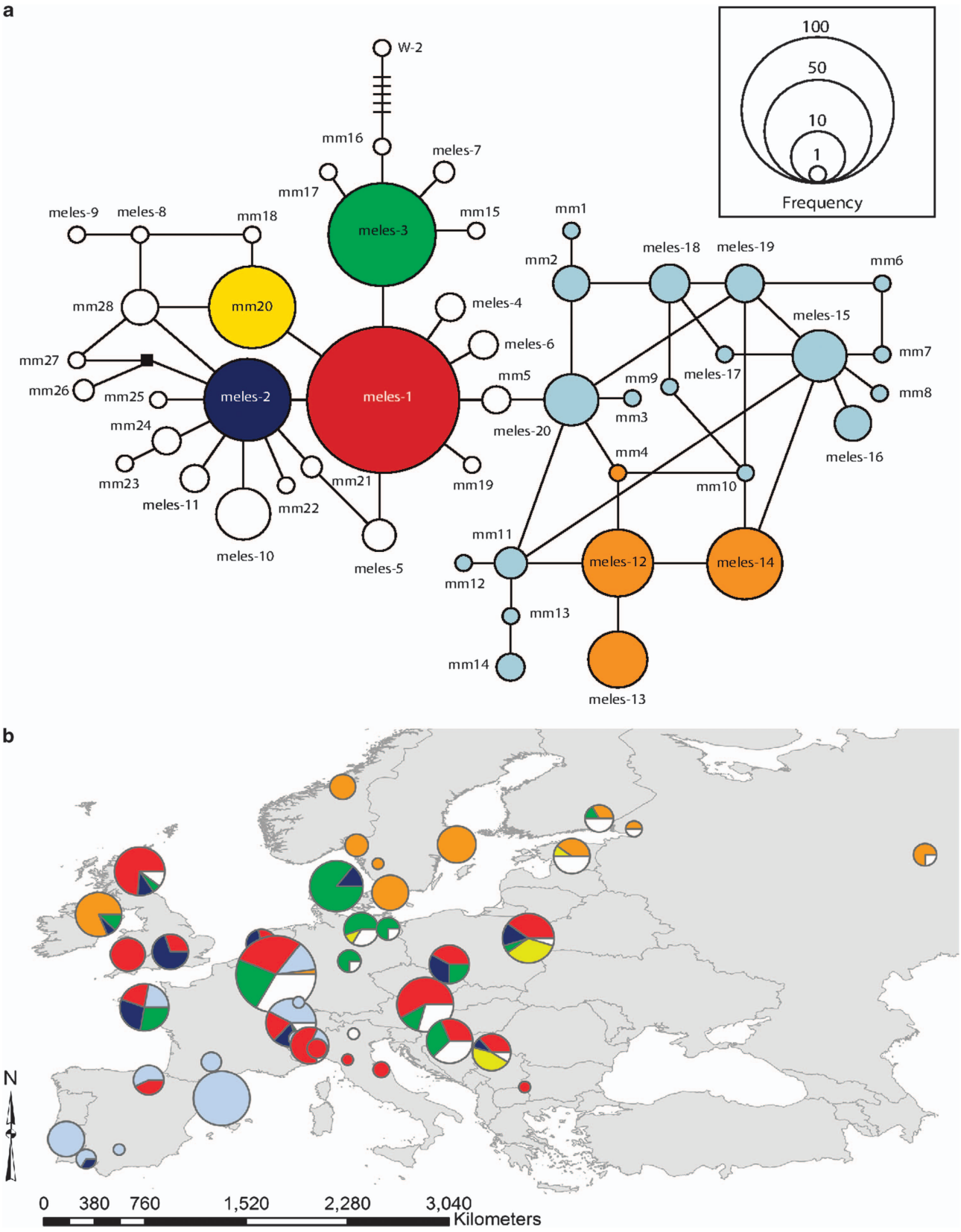

Figure 2 Median-joining network (**a**) and geographic distribution of 49 badger mitochondrial control region haplotypes (**b**). The same colours represent the same groups of haplotypes in both figures. Missing haplotypes are indicated by a small black square in the network. Horizontal bars represent mutational steps when greater than one. The sizes of the symbols in the network and the map are representative of haplotype frequency and sample sizes, respectively.

A factorial correspondence analysis (Figure 3) illustrated a strong differentiation of the badgers in Ireland, Britain and Scandinavia. Scandinavian badgers were—in contrast to the results of the mtDNA analysis—closely related to populations from eastern Europe and British/Irish badgers and to western–central Europe. The DIYABC analysis supported scenario 4 as the most probable, demonstrating that the Scandinavian population resulted from admixture between the Iberian and Balkan 'lineages' (Figure 4). This event was estimated to have occurred at 1760 BP (95% confidence interval: 525–6390 BP). The principal component analysis demonstrating the reliability of the chosen scenarios parameters and the posterior distributions of model parameters for the most likely scenario are provided in Supplementary Figure S5 and Supplementary Table S4, respectively.

The microsatellite-based genetic diversity measures (A_R and He_u) showed a significant decline in variability from east to west and from south to north (Supplementary Table S5), irrespective of whether all 25 predefined populations were used or only one per BAPS-defined genetic cluster. These relationships mainly resulted from badgers in Iberia, Ireland, Britain, Denmark and Scandinavia having a reduced genetic diversity compared with the remaining populations in mainland Europe (Supplementary Figure S6). In the case of mtDNA, there was evidence of a decline in haplotype richness (and in haplotype diversity using all populations) from south to north (Supplementary Table S4), with badgers in Britain, Ireland and Scandinavia in particular being less variable than the southern populations (Supplementary Figure S7).

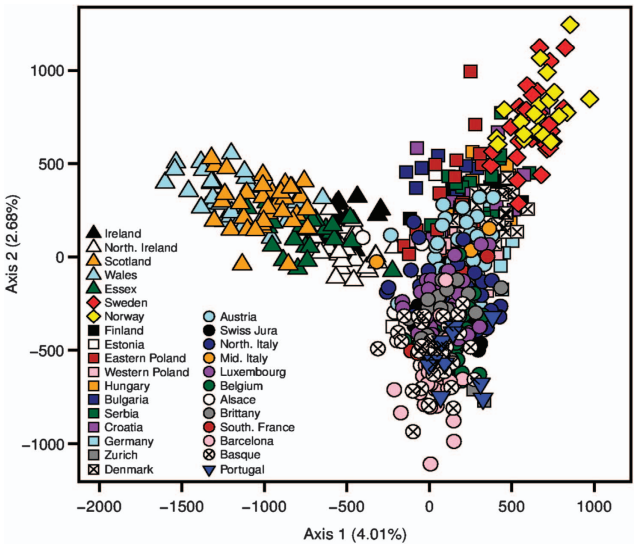

Figure 3 Factorial correspondence analysis of badgers from different predefined European populations. The analysis was based on 18 microsatellite loci. The percentage of the total variation explained by each of the two axes is given.

When trying to estimate values for $\log(N_0)$, $\log(N_1)$ and $\log(t_a)$ using the method by Storz and Beaumont (2002), the 12 initial MCMC chains for the three demographic parameters converged in 7 populations at the <1.1 level and in a further 7 populations the chains reached approximate convergence at least (Supplementary Table S6). In the case of 5 populations, the 12 initial chains relating to the $\log(N_0)$ and $\log(t_a)$ estimates did not converge. When these data were analysed using the three longer chains (a total of 10^{10} steps), the analyses for Serbia did not converge and were not considered further.

The MSVAR-based point estimates of $\log(N_1)$ were all relatively similar and varied between 3.7 and 4.1 (which does correspond to a range of $\sim$5000–12 500 individuals though; Figure 5). The point estimates of $\log(N_0)$ were particularly low ($\leqslant 1.9$ or $\leqslant 79$ individuals) for Belgium, Northern Ireland, Norway, Scotland and Sweden. In contrast, Croatia had the highest estimate of $\log(N_0)=3.2$ ($\sim$1580 individuals), followed by Barcelona (2.9; 800 individuals) and Northern Italy (2.8; 630 individuals). All other populations had point estimates of $2.0 \leqslant \log(N_0) \leqslant 2.7$ ($\cong$200–500 individuals). There was a significant positive correlation between point estimates of $\log(N_0)$ and $\log(t_a)$ ($r_s=0.72$; $P<0.001$). The five populations with estimates of $\log(N_0) \leqslant 1.9$ were also estimated—with one exception—to have suffered the most recent declines ($2.1 \leqslant \log(t_a) \leqslant 2.4$; Figure 5). However, the point estimates suggested that no decline occurred before 100 generations ago. On the opposite end of the spectrum, the populations in Luxembourg and Barcelona were estimated not to have declined before 2500 generations ($\log(t_a) \geqslant 3.4$) ago. It should be emphasised that all estimates of $\log(N_0)$ and $\log(t_a)$ had wide 90% credible intervals.

When analysing the simulated data with the MSVAR method, the majority of the chains estimating $\log(N_0)$ only converged approximately; in 10 cases, the three longer chains needed to be run to achieve convergence and two data sets needed to be replaced because of nonconvergence (Supplementary Table S7). Although sample size

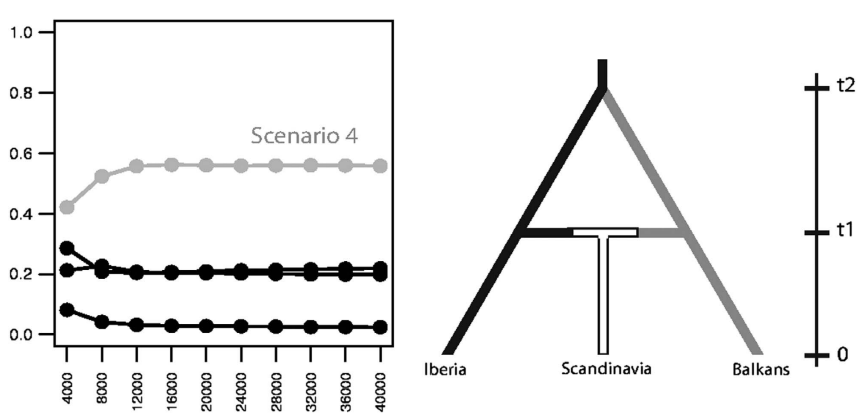

Figure 4 Results from the ABC analysis: graph of linear regressions showing posterior probabilities of the scenarios on the Y axis and the number of simulations used to calculate it (1% of total simulations) on the X axis. The plot for the best-supported scenario 4 is in grey graphic and represented on the right.

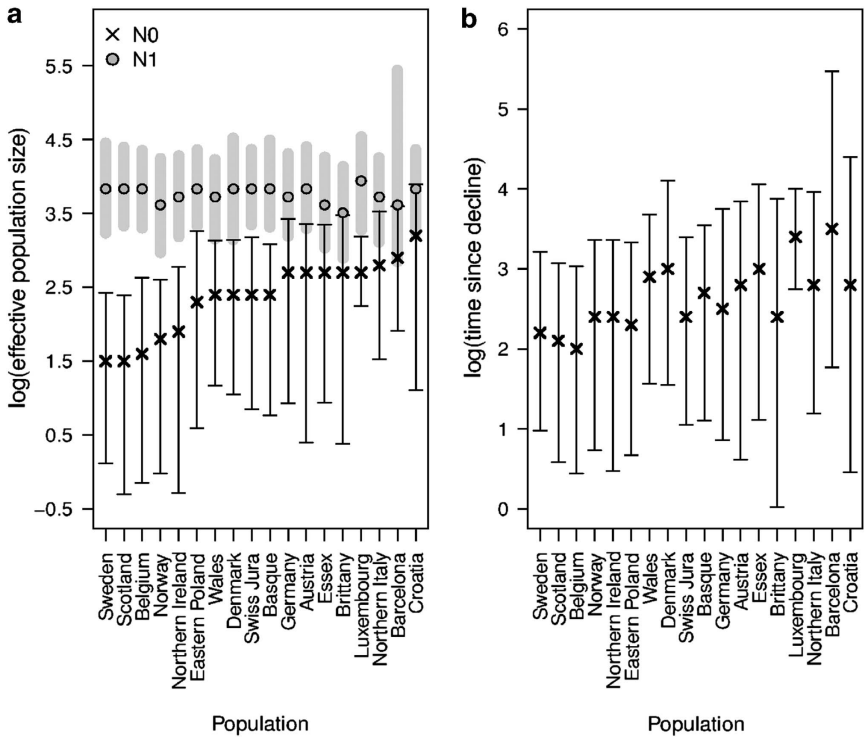

Figure 5 Quantification of populations size changes using the MSVAR method by Storz and Beaumont (2002), consisting of (**a**) estimates of the log of present (N_0) and past (N_1) effective population sizes and (**b**) log of time since the decline in population size. The point estimates represent the mode of the highest posterior density interval, and the error bars and grey bars represent the corresponding 90% credible intervals.

did not affect accuracy, estimates of $\log(N_0)$ were more accurate for older and severe events (Figure 6, Supplementary Figure S8 and Supplementary Table S8). Estimates of $\log(N_0) < 2.0$ were obtained with 20 simulated data sets, all but one of which corresponded to a 100-fold simulated decline (5000 to 50 individuals; Figure 6). Point estimates of $2.0 \leqslant \log(N_0) \leqslant 2.7$ were obtained for 34 simulated data sets, 33 of which contained at least two of every possible combination of N_0 (50, 150) and t_a (25, 100, 500). The 10 data sets with a simulated N_0 of 150 and a decline 25 generations ago gave rise to point estimates of $2.6 \leqslant \log(N_0) \leqslant 3.5$. The chains estimating N_1 achieved good convergence (Supplementary Table S6). The strength of the bottleneck did affect the accuracy of the $\log(N_1)$ point estimate (Supplementary Table S9). MSVAR tended to underestimate $\log(N_1)$ for severe bottlenecks occurring more than 100 generations ago (the interaction was not significant though; Supplementary Figure S9 and Supplementary Table S9).

When using MSVAR to estimate the $\log(t_a)$ of the simulated data, around a third of all runs did not achieve approximate convergence and/or had a bimodal posterior distribution and these were excluded from further analysis (Supplementary Figure S10 and Supplementary Table S7). This appeared to be particularly the case for less severe and more recent declines (Supplementary Figure S10). In the remaining results, the point estimates were always higher than the simulated values. Point estimates of time since decline varied between $1.9 \leqslant \log(t_a) \leqslant 3.2$ and $2.0 \leqslant \log(t_a) \leqslant 3.2$ when analysing data sets with a simulated decline 25 and 100 generations ago, respectively. Point estimates of $\log(t_a) \geqslant 2.9$ were obtained for data sets with a simulated t_a of 500. In the case of the simulated data, there was also a significant positive correlation between point estimates of current effective population size and time since decline ($r_s = 0.38$; $P < 0.00$).

DISCUSSION

The results of our median-joining network were consistent with those presented by Marmi *et al.* (2006), revealing no clear mtDNA genetic structuring within European badgers. However, the geographic distribution of some groups of haplotypes in European badgers strongly suggested the presence of at least two separate source populations. Specifically, we had haplotypes with an Iberian distribution and another group of haplotypes spread throughout most of the continent, with secondary contact of the haplotypes from both 'Iberian' and 'Eastern' groups in western Europe. The evidence for a bimodal mismatch distribution (Patarnello *et al.*, 2007) further supported a scenario of expansion out of at least two glacial refugia (Supplementary Figure S2). Although not converging on an optimal solution regarding the population genetic structure of European badgers, the different microsatellite-based analytical methods showed that the badgers in Iberia together with populations in Ireland, Britain, Denmark and Scandinavia were well differentiated from the remaining mainland populations.

The presence of many unique haplotypes in Iberia, coupled with extensive fossil data from the region (Sommer and Benecke, 2004), clearly points to the peninsula having been a glacial refugium for badgers. Fossil evidence also shows that badgers were present in the Carpathians and the Balkans during the LGM (Pazonyi, 2004; Sommer and Benecke, 2004). Southeastern and eastern Europe are also the only regions where all four main 'Eastern' haplotypes (meles1–3 and mm20) are found, suggesting that expansion started outwards from here. Neither the mtDNA nor the microsatellite data provided evidence for genetic divergence between the Carpathian and Balkan refugia, in contrast to results from smaller mammals with more limited dispersal ability (Kotlík *et al.*, 2006; McDevitt *et al.*, 2012). The greater dispersal ability possessed by larger mammals may

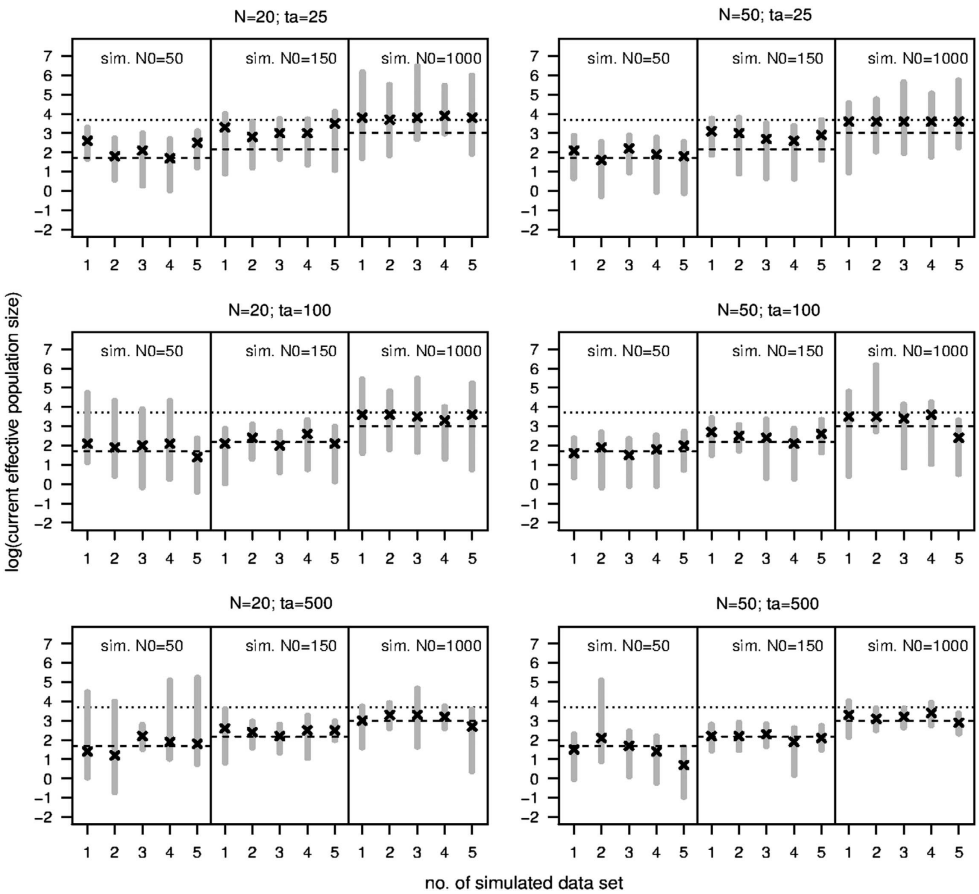

Figure 6 Inference of the current effective population size, log(N_0), of simulated data sets using MSVAR. The simulated data sets had different samples sizes (N), time since population decline (t_a) and current effective population size (sim N_0). Past effective population sizes were fixed to 5000 individuals. In each graph, the grey bars represent the 90% credible intervals of the point estimate represented by a cross. The dotted and the dashed lines indicate the simulated past and current effective population sizes, respectively.

have allowed connectivity between the Carpathian and Balkan refugia, as a similar pattern of distinct refugia but with apparently uninterrupted gene flow during the LGM was shown for another large mammal, red deer (Zachos and Hartl, 2011).

The origin of Scandinavian badgers is more uncertain and results from the two marker types suggested differing colonisation scenarios. The mtDNA haplotypes observed in Scandinavia were closely related to the Iberian haplotypes (Figure 2a). Considering that the first subfossil records from southern Sweden have been dated to ~9000 BP (Sommer and Benecke, 2004), it appears most likely that the Iberian lineage reached Scandinavia first over a landbridge (Björck, 1995) after the Younger Dryas (because of the dominance of the closely related mtDNA haplotypes). Scandinavian haplotypes were also found as far eastwards as Russia, and this suggests an extensive expansion after this initial colonisation from Iberia. Based on the extensive distribution of these haplotypes and the recent nature of this recolonisation (after the Younger Dryas), an argument could be made for a third, unsampled refugium with badgers colonising Scandinavia from the east as opposed to Iberia. However, the close relationship between Iberian and Scandinavian haplotypes suggests this to be unlikely. The microsatellite-based ABC analysis provided most support for a colonisation of Scandinavia from both Iberian and southeastern refugia (Figure 4). This later admixture from the southeast (via Finland) occurred more recently (525–6390 BP). This pattern of admixture in Scandinavia has been previously observed in other mammalian species (Brunhoff *et al.*, 2003; Ruiz-Gonzalez *et al.*, 2013). The mtDNA haplotypes from the southeast of Europe do not occur in Scandinavia however (Figure 2b), but this may be because of a selective advantage of the haplotypes of the first wave of colonisers in the region (Ruiz-Pesini *et al.*, 2004; McDevitt *et al.*, 2012).

The colonisation of Ireland provides another clear example of conflicting results obtained from mtDNA and microsatellite data. The question of how the island acquired its fauna and flora has long presented a problem (McDevitt *et al.*, 2011). Using data from a shorter fragment of the control region and six microsatellites, O'Meara *et al.* (2012) recently concluded that badgers have colonised Ireland naturally, but failed to identify how and when this occurred. We show here that Irish badgers share the majority of their haplotypes with Scandinavia but are closely related to Britain at the microsatellite level (Figures 2a and 3). Natural colonisation from an Iberian source (as we have proposed for Scandinavia) is unlikely because of the absence of evidence for suitable landbridges post LGM (McDevitt *et al.*, 2011) or appropriate fossil evidence before the presence of humans in Ireland (McCormick, 1999). The only securely dated badger fossil in Ireland dates to 1554 BP (P Woodman and M O'Dowd, personal communication) and other fossils are of a Bronze Age context (4000–1500 BP; McCormick, 1999). Therefore, human-mediated introduction(s), which has been reported for other Irish mammals (see, for example, McDevitt *et al.*, 2011), is the most likely scenario and our genetic results support this. Early introductions

from Britain (as supported by a number of mtDNA haplotypes shared between the islands and microsatellite data) would have been followed by later introductions during the Viking invasions from Scandinavia from ~800 AD onwards (supported by the Scandinavian haplotypes found on the island). The Vikings were also responsible for the introduction of the house mouse (*Mus musculus*) onto Ireland (Searle *et al.* 2009) and badger bones were associated with Viking sites in Dublin from the tenth and eleventh centuries AD, possibly being a valuable food item at the time (McCormick, 1999).

Our results show that the populations on the European periphery—the western- and northern-most populations—have a reduced genetic diversity (at least at the microsatellite level) compared with the populations in central, eastern and southeastern Europe. As a result of successive founder events, species that expanded from a refugium may exhibit reduced genetic diversity in populations furthest from their origin (Hewitt, 1999), especially if the population in the refugium had a small effective population size. It is therefore likely that the reduced diversity in the peripheral populations results from historic processes. On the other hand, badgers might have undergone recent declines and genetic bottlenecks as a result of human-induced habitat modification (see, for example, van der Zee *et al.,* 1992) and/or persecution in the context of disease management (Griffith and Thomas, 1997).

We aimed to use the MSVAR approach to assess the likelihood of recent demographic processes having contributed to a reduced diversity in some populations. Taken at face value, the estimates of $\log(N_0)$ are very encouraging, suggesting that five populations have undergone a severe decline and that three populations located in glacial refugia (Barcelona, Italy and the Balkans) have the largest current effective population sizes, with the remaining populations having intermediate estimates. MSVAR also suggested, however, that no decline occurred before 100 generations ago. Because the characteristics of a bottleneck, such as its timing and severity, can influence the accuracy and precision of the MSVAR analysis (Girod *et al.,* 2011), we performed our own simulations—based on realistic demographic scenarios and the number of samples and genetic markers available for the empirical analyses—to assess whether our empirical results were a realistic indication of past demographic events.

Similar to Girod *et al.* (2011), many chains—both with the empirical and the simulated data—only reached approximate convergence (Gelman and Rubin statistic <1.2), with a small number not converging at all. A less stringent convergence criterion thus appears necessary to obtain point estimates of the demographic parameters, despite the larger credible intervals this necessarily will entail. Results by Girod *et al.* (2011) suggested that scaled parameters ($\theta_0 \equiv 4N_0\mu$, $\theta_1 \equiv 4N_1\mu$ add $t_f \equiv t_a/(2N_0)$; $\mu =$ mutation rate) provide more precise estimates of the natural parameters for contractions. However, most researchers are likely to find point estimates of the demographic parameters intuitively more appealing.

Although the size of the empirical data sets varied between 20 and 37 individuals, our simulations suggested that the accuracy of the MSVAR point estimates was not influenced by sample size. However, the simulation results also suggested that it was very difficult to draw robust conclusions about the severity and timing of past bottlenecks in our empirical data. The estimates of time since decline appeared to be particularly unreliable, as the MSVAR point estimates of $\log(t_a)$ were always higher (sometimes substantially so) than the simulated values. However, five empirical populations had an estimated current effective population size of $\log(N_0) < 2.0$, with an estimated decline $\log(t_a) \leqslant 2.4$ generations ago. Only five of the simulated data sets had similarly low estimates of both $\log(N_0)$ and $\log(t_a)$, and all of these had a simulated N_0 of 50 individuals after experiencing a 100-fold decline 25 generations ago (Figure 6 and Supplementary Figure S10). In other words, the simulations provided support for a recent severe bottleneck of badger populations in Belgium, Northern Ireland, Norway, Scotland and Sweden.

However, drawing conclusions about the past demography of the remaining badger population seems problematic. Generally, our simulation results agreed with those by Girod *et al.* (2011), in that the estimates of the current effective population size tended to be more accurate for older and severe events. In other words, point estimates of $\log(N_0)$ for bottlenecks of the same severity tended to be larger for more recent bottlenecks. For example, values in the range of $2.0 \leqslant \log(N_0) \leqslant 2.7$ were obtained for all combinations of N_0 (50, 150) and t_a (25, 100, 500), and these data sets—especially those with a simulated decline 25 and 100 generations ago—also gave rise to estimates of $\log(t_a)$ that were comparable to the ones obtained with the empirical data. Although it appeared encouraging that the populations in the 3 classical glacial refugia had the highest estimated values for $\log(N_0)$, the 10 data sets with simulated $N_0 = 150$ and $t_a = 25$ gave rise to similarly high estimates of $\log(N_0)$ and $\log(t_a)$. In other words, the simulations suggested that for the majority of the empirical populations, the timing and severity of a possible bottleneck could not be established with certainty. Furthermore, given the convergence problems, the empirical estimates of the demographic parameters (with the possible exception of Luxembourg) had very large 90% credible intervals and were thus very imprecise as well.

During the twentieth century, badgers have declined dramatically in Belgium, partially as a result of the rabies eradication campaigns (Griffiths and Thomas, 1997): In 1992, only 28 badger setts were known from Flanders (where our samples were from). Although Scandinavian badgers are abundant and currently expanding their range, the population was restricted to the southern tip of the peninsula during the nineteenth century (Bevanger and Lindström, 1995). It is thus possible that the Scandinavian results were caused by a population expansion from small effective population size (MSVAR performs poorly when analysing data from expanding populations; Girod *et al.,* 2011). Although badgers in Denmark have declined over the past 50 years and exhibit reduced genetic diversity, Pertoldi *et al.* (2005) have convincingly shown that they have not suffered from a genetic bottleneck during the past half-century. We therefore cannot exclude the possibility that the genetic diversity in Scandinavia and Denmark was historically low, even before possible recent human-mediated declines.

There are—to our knowledge—no records of recent badger declines in Northern Ireland or Scotland (Griffiths and Thomas, 1997; Reid *et al.,* 2011). Although not directly comparable to this study, the estimates of allelic richness presented by both Pope *et al.* (2006) and O'Meara *et al.* (2012) suggested that most, if not all, British and Irish badger populations have similarly low genetic diversity than those analysed in this study. It is therefore likely that the reduced diversity observed both with microsatellites and mtDNA on the British Isles resulted from founder effects during natural colonisation, or—as probably in the case of Ireland—human-mediated introduction. It is possible that MSVAR proves the occurrence of recent unrecorded bottlenecks in Scotland as Pope *et al.* (2006) did not observe a decline in genetic diversity in Britain with latitude.

Given the general correlation of genetic diversity with latitude, and the support for southern refugia, it is likely that reduced genetic diversity in northern populations was contributed to by historic

processes. Unfortunately, despite multiple markers and extensive sampling, any additional effects of recent processes remain unresolved. As also shown by others, the precision and the accuracy of the MSVAR method is influenced by the timing and severity of a bottleneck. Based on simulated data sets representing specific demographic scenarios and the available number of samples and genetic markers, we have established that there is insufficient power to differentiate between historic founder effects and recent declines with certainty in the context of this study. We therefore urge caution when trying to relate demographic declines inferred using MSVAR with particular historic or climatological events (see, for example, Goossens *et al.*, 2006; Heller *et al.*, 2008), especially if there is uncertainty about the study species' generation time (which is required to estimate of the time since the decline). Other researchers should undertake similar testing if they wish to employ MSVAR in this way.

In summary, continent-wide sampling of badgers and analyses with multiple markers provided evidence for two glacial refugia (Iberia and southeast Europe) contributing to the genetic variation observed in badgers in Europe today. Because of lack of sampling in Russia, there remains a certain doubt, however, regarding the phylogeographic origin of badgers in the far east of Europe. The pattern of decline of genetic diversity with increasing latitude suggested that the reduced diversity in the peripheral populations probably resulted from a postglacial expansion processes. Because of methodological limitations, it was not possible to ascertain whether some of these peripheral populations also have undergone recent genetic bottlenecks. Therefore, despite our best efforts, additional questions remain to be answered about the large-scale genetic structure of *M. meles* in Europe.

DATA ARCHIVING

Sample locations and microsatellite data: DRYAD entry doi:10.5061/dryad.5nm5g.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The laboratory study was performed at the NERC Biomolecular Analysis Facility, supported by the UK Natural Environment Research Council. We thank Joe Chipperfield for discussions on MCMC methodologies and Ilaria Coscia and Stefano Mariani for advice on ABC. The following people contributed to sample collection: L Elorza, P Lizarraga, WRC Martioda (Basque Country), AB Madsen (Denmark), M Semchenko (Estonia), M Isomursu (Finland), V Boyaval (France), H Bocklisch, F Müller, K Weidemann (Germany), M Statham, P Turner (Ireland), E Randi, L Remonti (Italy), M Schaul (Luxembourg), H Brøseth, Petter Fakler (Norway), R Myslajek, K Zalewski (Poland) and K Hindelang (Switzerland). Sampling in Serbia was supported by the Serbian Ministry of Education, Science and Technological Development.

Avise JC (2004). *Molecular Markers Natural History, and Evolution*, 2nd edn. Sinauer & Associates: Sunderland, MA.
Bandelt H-J, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* **16**: 37–48.
Beaumont MA (1999). Detecting population expansion and decline using microsatellites. *Genetics* **153**: 2013–2029.
Beaumont MA (2010). Approximate Bayesian computation in evolution and ecology. *Annu Rev Ecol Evol Syst* **41**: 379–406.
Belkhir K (2004). *Genetix 4.05.2*. University of Montpellier II, Laboratoire Genome et Populations, Montpellier, France.
Bevanger K, Lindström ER (1995). Distributional history of the European badger *Meles meles* in Scandinavia during the 20th century. *Ann Zool Fenn* **32**: 5–9.
Björck S (1995). A review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quatern Int* **27**: 19–40.
Brooks SP, Gelman A (1998). General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* **7**: 434–455.
Brunhoff C, Galbreath KE, Fedorov VB, Cook JA, Jaarola M (2003). Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Mol Ecol* **12**: 957–968.
Cassens I, Mardulyn P, Milinkovitch MC (2005). Evaluating intraspecific 'network' construction methods using simulated sequence data: do existing algorithms outperform the Global Maximum Parsimony approach? *Syst Biol* **54**: 363–372.
Corander J, Waldmann P, Marttinen P, Sillanpää MJ (2004). BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics* **20**: 2363–2369.
Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ *et al.* (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**: 2713–2719.
Del Cerro Marmi J, Ferrando A, Chashchin P, Taberlet P, Bosch M (2010). Nuclear and mitochondrial phylogenies provide evidence for four species of Eurasian badgers (Carnivora). *Zool Scr* **39**: 415–425.
Excoffier L, Lischer HEL (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetic analyses under Linux and Windows. *Mol Ecol Resour* **10**: 564–567.
Fu Y-X (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.
Gelman A, Rubin DB (1992). Inference from iterative simulation using multiple sequences. *Stat Sci* **7**: 457–511.
Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC: New York.
Girod C, Vitalis R, Leblois R, Fréville H (2011). Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the MSVAR method. *Genetics* **188**: 165–179.
Goossens B, Chikhi L, Ancrenaz M, Lackman-Ancrenaz I, Andau P, Bruford MW (2006). Genetic signature of anthropogenic population collapse in orang-utans. *PLoS Biol* **4**: 285–291.
Griffiths HI, Thomas DH (1997). The conservation and management of the European badger (*Meles meles*). *Nat Environ* **90**: 1–77.
Guillot G, Estoup A, Mortier F, Cosson J (2005). A spatial statistical model for landscape genetics. *Genetics* **170**: 1261–1280.
Heller RE, Lorenzen D, Okello JBA, Masembe C, Siegismund HR (2008). Mid-Holocene decline in African buffalos inferred from Bayesian coalescent-based analyses of microsatellites and mitochondrial DNA. *Mol Ecol* **17**: 4845–4858.
Hewitt GM (1999). Post-glacial re-colonization of European biota. *Biol J Linn Soc* **68**: 87–112.
Kotlík P, Deffontaine V, Mascheretti S, Zima J, Michaux JR, Searle JB (2006). A northern glacial refugium for bank voles (*Clethrionomys glareolus*). *Proc Natl Acad Sci USA* **103**: 14860–14864.
Loader C (1999). *Local Regression and Likelihood*. Springer-Verlag: New York.
Librado P, Rosaz J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
Marmi J, Lopez-Giraldez F, Macdonald DW, Calafell F, Zholnerovskaya E, Domingo-Roura X (2006). Mitochondrial DNA reveals a strong phylogeographic structure in the badger across Eurasia. *Mol Ecol* **15**: 1007–1020.
McCormick F (1999). Early evidence for wild animals in Ireland. In: Benecke N (ed) *The Holocene History of the European Vertebrate Fauna: Modern Aspects of Research*. Verlag Marie Leidorf GmbH: Rahden: Germany, pp 355–371.
McDevitt AD, Vega R, Rambau RV, Yannic G, Herman JS, Hayden TJ *et al.* (2011). Colonization of Ireland: revisiting 'the pygmy shrew syndrome' using mitochondrial, Y chromosomal and microsatellite markers. *Heredity* **107**: 548–557.
McDevitt AD, Zub K, Kawałko A, Oliver MK, Herman JS, Wójcik JM (2012). Climate and refugial origin influence mitochondrial lineage distribution of weasels (*Mustela nivalis*) in a phylogeographic suture zone. *Biol J Linn Soc* **106**: 57–69.
Miller SA, Dykes DD, Polesky HF (1988). A simple salting-out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* **16**: 1215.
Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
O'Meara DB, Edwards CJ, Sleeman DP, Cross TF, Statham MJ, McDowell JR *et al.* (2012). Genetic structure of Eurasian badgers *Meles meles* (Carnivora: Mustelidae) and the colonization history of Ireland. *Biol J Linn Soc* **106**: 893–909.
Patarnello T, FAMJ Volckaert, Castilho R (2007). Pillars of Hercules: is the Atlantic-Mediterranean transition a phylogeographic break? *Mol Ecol* **16**: 4426–4444.
Pazonyi P (2004). Mammalian ecosystem dynamics in the Carpathian Basin during the last 27,000 years. *Palaeogeogr Palaeocl Palaeoecol* **212**: 295–314.
Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Bëer E, Robinson S *et al.* (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Mol Ecol* **21**: 3403–3418.
Pertoldi C, Loeschcke V, Madsen AB, Randi E, Mucci N (2005). Present and past microsatellite variation and assessment of genetic structure in Eurasian badger (*Meles meles*) in Denmark. *J Zool* **265**: 387–394.
Petit RJ, El Mousadik A, Pons O (1998). Identifying populations for conservation on the basis of genetic markers. *Conserv Biol* **12**: 844–855.
Plummer M, Best N, Cowles K, Vines K (2006). Coda: output analysis and diagnostics for MCMC. *R News* **6**: 7–11.
Pope LC, Domingo-Roura X, Erven K, Burke T (2006). Isolation by distance and gene flow in the Eurasian badger (*Meles meles*) at both a local and broad scale. *Mol Ecol* **15**: 371–386.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
R Development Core Team (2011). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna: Austria, ISBN 3-900051-07-0, URL. http://www.R-project.org/.
Raymond M, Rousset F (1995). GENEPOP (version 1.2): a population genetics software for exact tests and ecumenicism. *J Hered* **86**: 248–249.
Reid N, Etherington TR, Wilson GJ, Montgomery WI, McDonald RA (2011). Monitoring and population estimates of the European badger *Meles meles* in Northern Ireland. *Wildlife Biol* **18**: 46–57.
Rogers AR, Harpending H (1992). Population-growth makes waves in the distribution of pairwise genetic distances. *Mol Biol Evol* **9**: 552–569.
Ruiz-Gonzalez A, Madeira MJ, Randi E, Abramov AV, Davoli F, Gomez-Moliner BJ (2013). Phylogeography of the forest-dwelling European pine marten (Martes martes): new insights into cryptic northern glacial refugia. *Biol J Linn Soc* **109**: 1–18.
Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004). Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* **303**: 223–226.
Searle JB, Jones CS, Gündüz İ, Scascitelli M, Jones EP, Herman JS *et al*. (2009). Of mice and (Viking?) men: phylogeography of British and Irish house mice. *Proc Roy Soc Lond Ser B* **276**: 201–207.
Sommer RS, Benecke N (2004). Late- and post-glacial history of the Mustelidae in Europe. *Mammal Rev* **34**: 249–284.
Storz JF, Beaumont MA (2002). Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56**: 154–166.
Szpiech ZA, Jackobson NA, Rosenberg NA (2008). ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**: 2498–2504.
Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998). Comparative phylogeography and post-glacial colonization routes in Europe. *Mol Ecol* **7**: 453–464.
Tajima F (1989). The effect of change in population size on DNA polymorphism. *Genetics* **123**: 453–464.
Tashima S, Kaneko Y, Anezaki T, Baba M, Yachimori S, Abramov AV *et al.* (2011). Phylogeographic sympatry and isolation of the Eurasian badgers (*Meles,* Mustelidae, Carnivora): implications for an alternative analysis using maternally as well as paternally inherited genes. *Zool Sci* **28**: 293–303.
van der Zee FF, Wiertz J, ter Braak CJF, Apeldoorn RC (1992). Landscape change as a possible cause of the badger Meles meles L. decline in the Netherlands. *Biol Conserv* **61**: 17–22.
Verhoeven KJF, Simonsen KL, McIntyre LM (2005). Implementing false discovery rate control: increasing your power. *Oikos* **108**: 643–647.
Walsh PA, Metzger DA, Higuchi R (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**: 506–513.
Zachos FE, Hartl GB (2011). Phylogeography, population genetics and conservation of the European red deer *Cervus elaphus. Mammal Rev* **41**: 138–150.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)