

ARTICLE



Different waves of postglacial recolonisation and genomic structure of bank vole populations in NE Poland

Magdalena Niedziałkowska¹✉, Ewa Tarnowska¹, Wiesław Babik², Mateusz Konczal³, Karim Gharbi^{4,5}, Timothee Cezard⁴ and Bogumiła Jędrzejewska¹

© The Author(s), under exclusive licence to The Genetics Society 2023

Previous studies indicated that in some species phylogeographic patterns obtained in the analysis of nuclear and mitochondrial DNA (mtDNA) markers can be different. Such mitonuclear discordance can have important evolutionary and ecological consequences. In the present study, we aimed to check whether there was any discordance between mtDNA and nuclear DNA in the bank vole population in the contact zone of its two mtDNA lineages. We analysed the population genetic structure of bank voles using genome-wide genetic data (SNPs) and diversity of sequenced heart transcriptomes obtained from selected individuals from three populations inhabiting areas outside the contact zone. The SNP genetic structure of the populations confirmed the presence of at least two genetic clusters, and such division was concordant with the patterns obtained in the analysis of other genetic markers and functional genes. However, genome-wide SNP analyses revealed the more detailed structure of the studied population, consistent with more than two bank vole recolonisation waves, as recognised previously in the study area. We did not find any significant differences between individuals representing two separate mtDNA lineages of the species in functional genes coding for protein-forming complexes, which are involved in the process of cell respiration in mitochondria. We concluded that the contemporary genetic structure of the populations and the width of the contact zone were shaped by climatic and environmental factors rather than by genetic barriers. The studied populations were likely isolated in separate Last Glacial Maximum refugia for insufficient amount of time to develop significant genetic differentiation.

Heredity (2023) 130:269–277; <https://doi.org/10.1038/s41437-023-00600-1>

INTRODUCTION

The phylogeographic structure of many extant mammal species in large areas of Europe, especially in northern areas, has been shaped mainly by the postglacial migration waves and admixture of different populations, which survived in several refugia during the Last Glacial Maximum (LGM, 19–26,000 BP; Clark et al. 2009) and recolonised the continent after the glacial retreat (Marková et al. 2020; Plis et al. 2022; Doan et al. 2022). Mitochondrial DNA (mtDNA) is a genetic marker most often used to study the history of populations of different mammal species and to reconstruct their recolonisation routes as this locus is haploid, has a fast mutation rate and it is inherited without recombination in maternal line. However, more recent studies indicated that the phylogeographic pattern obtained using nuclear markers is not always concordant with the results of mtDNA analyses, so more loci are needed to fully recognise the genetic division of the species or population of interest and the factors shaping it (e.g., Towes and Brelford 2012; Wallis et al. 2017).

In central and northern parts of the continent, there are contact zones of different mtDNA lineages of several mammal species such as rodents, ungulates and large carnivores originating from separate LGM refugia (Taberlet et al. 1998; Stojak and Tarnowska 2019; Niedziałkowska et al. 2021a). One of those species is the

bank vole *Myodes (Clethrionomys) glareolus*—a common forest rodent inhabiting almost the whole of Europe and large areas of the Western Asia up to the Lake Baikal (Hutterer et al. 2021). In Europe, there are several mtDNA lineages of the species, which survived during the LGM in various refugia in the southern, central and eastern parts of the continent (Deffontaine et al. 2005; Kotlik et al. 2006; Wójcik et al. 2010), similar to the genetic lineages of other temperate species, e.g. common vole (*Microtus arvalis*), red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) (Stojak et al. 2016; Doan et al. 2022; Plis et al. 2022). Central and northern Europe (Fennoscandia) is inhabited mainly by the bank vole lineages originating from the Eastern and the Carpathian refugia (Wójcik et al. 2010; Marková et al. 2020). These two genetic groups form a secondary contact zone in north-eastern Poland (Wójcik et al. 2010), and their distribution correlates with some plant species and the mean July temperature (Tarnowska et al. 2016).

Further studies, based on microsatellite DNA, showed that, in north-eastern Poland, there are two genetic clusters of the species, and their spatial distribution is concordant with the phylogeographic pattern of the two mtDNA lineages (Tarnowska et al. 2019). However, the analyses also indicated that the width of the hybridisation zone of the bank vole belonging to two different mtDNA lineages ranges from narrow in the south (2 km)

¹Mammal Research Institute Polish Academy of Sciences, 17-230 Białowieża, Poland. ²Institute of Environmental Sciences Jagiellonian University, 30-387 Kraków, Poland.

³Evolutionary Biology Group, Faculty of Biology, Adam Mickiewicz University, 60-614 Poznań, Poland. ⁴Edinburgh Genomics, University of Edinburgh, Edinburgh EH9 3FL, UK.

⁵Present address: Earlham Institute, Norwich NR4 7UZ, UK. Associate editor: Sam Banks. ✉email: mniedz@ibs.bialowieza.pl

Received: 9 May 2022 Revised: 2 February 2023 Accepted: 6 February 2023

Published online: 21 March 2023

to wide in its northern part (180 km). In the southern part of the hybridisation zone, the forest cover is very low and gene flow between the two genetic populations is suppressed by open, deforested areas (Tarnowska et al. 2019). By contrast, the analyses of one of the functional genes—the immune genes of toll-like receptor 2 (TLR2), important for detecting *Borrelia* infection (Hirschfeld et al. 1999)—did not indicate any spatial structure of TLR2 clusters in bank vole populations related to climatic and environmental factors or to the distribution of the mtDNA lineages in north-eastern Poland (Tarnowska et al. 2020). Contrary to the suggestions by Tschirren et al. (2013), we did not find any differences in *Borrelia* infection prevalence between bank voles carrying different TLR2 genotypes, either.

Until now, there have been no detailed studies analysing the population genetic diversity in the contact zone of different mtDNA lineages of mammal species inhabiting central Europe, including genome-wide data and functional genes of adaptive meaning, which could allow to identify more detailed genetic structure, discordance among different genetic markers and help to determine the main factors shaping observed genetic pattern of the species. The contact zones of genetic lineages are excellent areas to study the evolutionary processes such as drivers of reproductive isolation, species divergence and speciation in natural conditions (e.g., Singhal and Moritz 2012). Results of such studies can be important in delineating of conservation units, understanding how environmental change shapes species distributions and population structure or determining the consequences of postglacial population dynamics on its genetic diversity, structure and fitness (Singhal and Moritz 2012; Toews and Brelsford 2012). One of the important aspects worth studying is the concordance between the nuclear and mitochondrial genomes. Interactions between products of mtDNA and nuclear genomes are important for evolution, ecology and fitness of animals (Sunnucks et al. 2017; Lima et al. 2019). For instance, in most animals, the oxidative phosphorylation complexes responsible for cellular bioenergetics consist of 13 mitochondrial-encoded and 80 nuclear-encoded protein interacting subunits (Sunnucks et al. 2017). Therefore, coadaptation of mitonuclear interactions is crucial for cell respiration.

Incompatibility between mitochondrial and nuclear DNA could occur when they are derived from different species or populations; for example, in hybrids, the parents of which originated from different phylogenetic lines (Lane 2009). These incompatibilities can cause lower fitness, poorer survival and reproduction of hybrids. Until now, the lack of mitonuclear coadaptation in hybrids was experimentally confirmed in copepod (*Tigriopus californicus*) (Ellison and Burton 2006; Lima et al. 2019) and in yeast (*Saccharomyces bayanus* and *S. cerevisiae*) (Lee et al. 2008). There are also many examples of mitonuclear discordance in animal populations, which are a consequence of other factors such as adaptive introgression of mtDNA, demographic disparities and sex-biased asymmetries, hybrid zone movement or human introductions (Toews and Brelsford 2012).

So far, there have been no studies that would analyse the genetic division of the bank vole populations in continental Europe, in the secondary contact zone, on the basis of adaptive loci responsible for coding protein complexes important for cell respiration and would compare it with the genetic structures revealed by analysis of different neutral genetic markers (mtDNA and RAD-seq SNPs). The important role of respiratory-related (such as haemoglobin) genes in adaptation during post-LGM expansion of bank voles has been indicated in Britain in the study of Kotlík et al. (2014; 2018). However, there are not many other studies documenting associations between functional genomic variation and the postglacial recolonisation history in European mammals.

The contact zone between the two mtDNA lineages of bank vole and their hybridisation zone of two microsatellite clusters in the NE Poland covers a wide area (Tarnowska et al. 2016, 2019).

Thus, we hypothesised that the genetic structures of the population obtained using different markers (genome-wide SNPs, mtDNA, microsatellites and functional genes of heart transcriptomes) will be concordant, and genetic divergence among revealed genetic populations will be low. We predicted that SNP analyses would confirm the presence of at least two genetic populations corresponding to the spatial distribution of mtDNA lineages and microsatellite genetic clusters (compare Tarnowska et al. 2016, 2019). We supposed that the differentiation of functional genes of heart transcriptomes between individuals belonging to the two mtDNA lineages will be low, which would suggest that bank voles from the two lineages are not isolated owing to differences in the functional genes of adaptive significance. We hypothesised that the genetic structure of the studied population was shaped mainly by postglacial recolonisation and environmental factors, with a minor contribution from genetic barriers between individuals originating from different LGM refugia.

The aims of our study were to analyse the population genetic structure of the species in the contact zone of its two mtDNA lineages, using high-resolution genetic data (SNPs), and to compare the results with the pattern revealed by mtDNA and microsatellite DNA analyses performed recently in the same study area. We also analysed differentiation in functional genes of the heart transcriptomes obtained from selected individuals from three bank vole populations inhabiting areas outside the contact zone of the two mtDNA lineages. Sequencing the heart transcriptomes allowed us to identify nuclear genes coding for protein-forming complexes, which are involved in the process of cell respiration in mitochondria.

MATERIAL AND METHODS

Study area and trapping

The study was conducted in forest habitats of north-eastern Poland and covered about 45,000 km² (52°21'–54°20' N, 18°59'–23°53' E). The northern part of the region was characterised by postglacial topography formed during the Würm glaciation (115,000–11,500 years BP) (Kondracki 1994), whereas the southern part was mostly a plain cut through by three large rivers (Vistula, Narew and Bug).

For genome-wide SNP analysis, 192 bank vole (*Myodes glareolus*) individuals were used. Samples were collected from 17 localities, referred to as populations (Fig. 1), in 2006 and 2011–2013. These populations were localised on two transects crossing the study area and the contact zone of the Eastern and the Carpathian mtDNA lineages of the bank vole (Tarnowska et al. 2016). Trapping was conducted in summer (June–September) using live traps with a trapping time from 1 to 7 days, depending on the abundance of animals. Bank voles were trapped according to the methods described by Niedziałkowska et al. (2010) and Tarnowska et al. (2016). From each captured animal, a small tail fragment (~4 mm) was taken for genetic analyses. All samples were stored in 96% ethanol.

To analyse functional differentiation between two Eastern and Carpathian lineages, we studied the heart transcriptomes of 20 bank voles from three populations: 10 individuals in the Białowieża Forest—BIA, 2 in Miłomłyn—MI and 8 in Włocławek—WL (Fig. 1). According to the genetic analyses (based on mtDNA, microsatellite and SNP datasets), the BIA population belonged to the same mtDNA lineage and genetic population as neighbouring populations (e.g., AL and AJ) in the region of the north-eastern Poland (Wójcik et al. 2010; Tarnowska et al. 2016, 2019; Marková et al. 2020). The trapped animals were euthanised in the place of capture by cervical dislocation. Then, their internal organs were extracted. Heart samples were preserved in the RNAlater (Sigma) immediately after dissection and transported to the laboratory.

The capture and handling of animals were conducted according to the procedure approved by the Local Ethics Committee of Animal Research in Białystok, Poland (permissions 15/2006, 16/2011, 43/2011 and 12/2013).

Laboratory analyses

Genomic DNA was extracted using Qiagen DNeasy Blood & Tissue Kit (Valencia, CA, USA). DNA concentration and quality were checked using

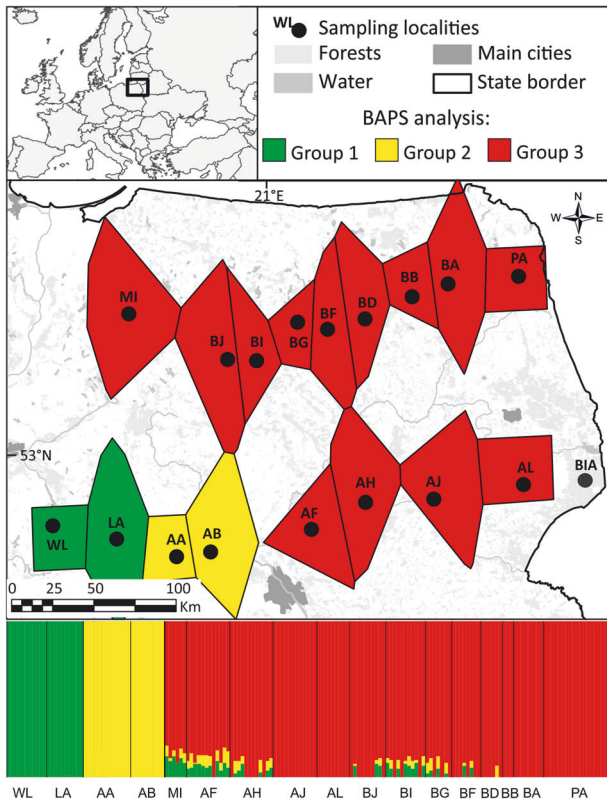


Fig. 1 Sampling localities of bank voles *Myodes glareolus* in NE Poland with results of SNP clustering indicated in BAPS analyses (spatial models). Each colour corresponds to one cluster and each bar denotes one individual. BIA—locality, where we analysed only heart transcriptome and mtDNA sequences of bank voles.

NanoDrop and electrophoresis in 1% agarose gel with FastRuler Low Range DNA Ladder (Fermentas). RAD-tag library preparation and sequencing were performed at Edinburgh Genomics (<https://genomics.ed.ac.uk/>), as described in Gonen et al. (2014) after Etter et al. (2011). In brief, each sample was digested with Sbf and ligated to an individual-specific, barcoded P1 adapter. P1-ligated samples from multiple individuals were pooled into a single library, followed by shearing, size selection, P2 adapter ligation and PCR amplification. Each amplified library was quantified by qPCR and pooled before Illumina sequencing on a HiSeq 2000 platform, using 50-base paired-end reads (v3 chemistry). Samples from the BIA population were not used for the RAD-tag sequencing as the DNA extracted from these samples was not of sufficient quality for this kind of analyses.

Total RNA from heart samples was extracted using RNAzol (MRC) and used for poly-A selection, library preparation (Illumina TruSeq RNA kit) and sequencing. Sequencing was performed on a single lane of the Illumina HiSeq2000 platform with 2×100 bp mode.

SNP filtering and statistical analyses

Raw RAD-tag reads were demultiplexed using `process_radtags` as part of the Stacks v1.12 package (Catchen et al. 2011, 2013). Output files were further processed using `ustacks` and `cstacks` in Stacks with default parameters. The number of called SNPs was 2,495,157. An output `vcf` file was filtered in program `VCFtools v 0.1.13` (Danecek et al. 2011), with the following parameters: `mac = 3` (minor allele count), `minQ = 30` (minimum sites quality), `max-missing = 0.9` (proportion of missing data where 0 allows sites that are completely missing and 1 indicates no missing data), `min-meanDP = 10` (minimum mean depth values for genotypes), `maf = 0.05` (minor allele frequency). After initial filtering, we obtained a total of 60,945 SNPs.

In the next step, individuals with more than 10% missing data ($N = 26$) were identified and removed. Plink v1.90b6.17 software (Chang et al. 2015; www.cog-genomics.org/plink/1.9/) was used for removing SNPs with $LD r^2 > 0.05$ (10,989 SNPs were left). Next, all variants with missing call rates

exceeding a value of 0.1 were filtered out (with “`geno`” command), and all SNPs with HWE exact test p -value below 0.00001 were removed. Following filtering, we retained a total of 10,819 SNPs across 166 individuals for further analysis.

To detect relationships among populations of the bank vole, hierarchical principal component analysis (PCA) was performed in Plink (Purcell et al. 2007). In the first step, PCA was performed for all populations and, in the second one, without the 4 most distinct populations (to better explore admixture in the contact zone). PCA plots were prepared in R 3.2.3 (R Development Core Team 2015).

To define genetic ancestry and population structure, two approaches implemented in ADMIXTURE 1.3 (Alexander et al. 2009) and BAPS 2.4.7 software (Corander et al. 2008) were used. ADMIXTURE uses maximum-likelihood estimation, whereas BAPS is based on Bayesian approach, which also incorporates spatial analysis. The ADMIXTURE analysis was run for 1 to 10 genetic populations (K) with 10 replicates for each K . The correct value of K was detected on the basis of cross-validation (CV) error (the CV procedure was run 10 times with different random seeds). Results were plotted using R. The BAPS analysis was performed with spatial and spatial admixture models. The most optimal number of clusters was calculated on the basis of the log marginal likelihood ($\log(\text{ml})$) values estimated for different numbers of clusters. The highest value of $\log(\text{ml})$ indicates the most optimal number of K . The spatial models were run for 1 through 10 to 24 clusters. All other parameters were set to minimum values indicated in the program manual. Results were transferred onto the map in ArcGIS 10.3 (ESRI 2012).

Summary statistics of genetic diversity for the final set of SNPs (proportion of polymorphic loci— P and heterozygosity described here as mean nucleotide diversity— π) for 17 local populations were calculated using ARLEQUIN 3.5.1.2 (Excoffier and Lischer 2010). F_{ST} values between all pairs of populations and genetic clusters, indicated by the BAPS analysis, were calculated using ARLEQUIN software.

Cytochrome b analysis

A maximum likelihood (ML) tree was constructed for the haplotypes detected in our previous study (Tarnowska et al. 2016; GenBank accession numbers: KT438074-6, KT438078, KT438080, T438082-4, KT438089-90, KT438098-9, KT438101-2, KT438104-6, KT438112, KT438115-8, KT438125, KT438127-8, KT438135, KT438138-9 and KT438159) and Marková et al. (2020) (GenBank accession numbers: MN102750–103006) by MEGA version 7.0.26 (Kumar et al. 2016), using the HKY+I+G substitution model, as the best-fit model indicated by MEGA (according to BIC values) and 1,000 bootstrap replications. Where sequences did not overlap completely, they were shortened to 346 bp before analysis.

Reconstructing the reference transcriptome

To assemble the heart transcriptome de novo, we first trimmed low-quality sequences using DynamicTrim, removed adaptors with Cutadapt and removed reads shorter than 20 bp with LengthSort (Cox et al. 2010; Martin 2011). We then employed Trinity software with default parameters (Grabherr et al. 2011). The transcripts assembled by Trinity that were likely derived from the same genomic locations were merged into transcriptome-based gene models, as described previously in Stuglik et al. (2014) and Konczal et al. (2015). The transcriptome was searched for long open-reading frames, which were later annotated using Trinotate software.

SNP calling and statistical analyses

To identify polymorphic sites, we first trimmed low-quality reads and removed adaptors using `trimmomatic` software (version 0.39; Bolger et al. 2014). Trimmed reads were mapped to the assembled transcriptome using `BWA mem` and default parameters (version 0.7.10-r789; Li and Durbin 2010). Duplicated reads were marked using `samtools markdup`, and the output files were used for genotype calling with `bcftools mpileup` (`-C50, -max-depth 8000, -min-MQ 20, -min-BQ 20`) and `bcftools call`. Sites were later filtered using `BCF filter` to exclude positions with summed depth across all samples smaller than 200 (`INFO/DP < 200`), sites quality smaller than 30 (`Q < 30`), and mean mapping quality smaller than 30 (`MQ < 30`). Subsequently, we applied `vcftools` (Danecek et al. 2011) with the following parameters: `max-missing = 0.9` (proportion of missing data), `min-meanDP = 10` (minimum mean depth values for genotypes), `-min-alleles 2, -max-alleles 2` (to allow only for biallelic SNPs) and `-remove-indels` (to remove indels).

Table 1. Parameters of SNP genetic diversity within 17 populations of bank voles *Myodes glareolus* in NE Poland.

Population	mtDNA lineage	<i>N</i>	<i>P</i>	π (SD)
WL	C/W	11	0.660	0.237 (0.117)
LA	C	10	0.655	0.237 (0.118)
MI	C	6	0.549	0.219 (0.113)
AA	C/E/W	13	0.659	0.232 (0.114)
AB	E/C	9	0.633	0.229 (0.115)
AF	E	12	0.601	0.223 (0.110)
AH	E	12	0.639	0.233 (0.115)
AJ	E/C	12	0.653	0.237 (0.117)
AL	E	9	0.478	0.197 (0.098)
BJ	E/C	10	0.486	0.202 (0.101)
BI	E/C	11	0.607	0.224 (0.111)
BG	E/C	7	0.539	0.217 (0.111)
BF	E/C	8	0.556	0.218 (0.110)
BD	E/C	6	0.646	0.239 (0.124)
BB	E	3	0.424	0.218 (0.126)
BA	E/C	8	0.579	0.221 (0.112)
PA	E	19	0.684	0.241 (0.117)

C, Carpathian; E, Eastern; W, Western mtDNA lineage; *N*, the number of individuals; *P*, the proportion of polymorphic loci, π , mean nucleotide diversity (heterozygosity), *SD* standard deviation.

MtDNA lineage determined according to Tarnowska et al. (2016). See Fig. 1 for geographic location of populations.

To detect relationship among samples, hierarchical PCA was performed in Plink as mentioned earlier. The measures of genetic diversity (*P*—proportion of polymorphic loci, π —nucleotide diversity) and divergence F_{ST} were calculated using vcftools. SNP annotation was performed with SnpEff software (Cingolani et al. 2012), and functions of outlier SNPs were identified using the Harmonizome database (Rouillard et al. 2016; available at <https://maayanlab.cloud/Harmonizome>).

RESULTS

Genomic diversity of bank vole population in the NE Poland

Most of the samples in the study area belonged to two mtDNA lineages: the Carpathian and the Eastern, which formed a contact zone in the NE Poland (Table 1 and Fig. S1). Only two individuals belonged to the Western mtDNA lineage (one in population WL and one in AA, Tarnowska et al. 2016). Haplotype and nucleotide diversities were highest in the local populations occurring in the contact zone of mtDNA lineages (Tarnowska et al. 2016).

On the basis of the analyses of 10,819 SNPs, we estimated that the proportion of the polymorphic loci (*P*) in the studied 17 bank vole populations ranged from 0.424 to 0.684 (Table 1 and Fig. S2) and was correlated with the number of samples ($r = 0.65$, $p < 0.01$). The nucleotide diversity varied from 0.197 to 0.241 (Table 1, Fig. S2) and was not significantly correlated with the number of samples in the populations ($r = 0.42$, NS). The highest genetic diversity was recorded in the PA population and in populations in the southern part of the study area (Fig. S2). The highest number of samples was analysed in PA, so that this could inflate diversity relative to other northern populations. The genetic divergence (F_{ST}) among the pairs of the populations varied from very low (0.016, populations BB and PA) to moderate (0.112, WL and AJ) and in most cases was statistically significant (Table S1). F_{ST} was not significant among the population BB and some other populations, but this was likely owing to low number of samples in BB.

Table 2. Parameters of genetic diversity within three SNP clusters of bank voles in NE Poland.

Cluster	Populations	<i>N</i>	<i>P</i>	π (SD)
1	2 (WL, LA)	21	0.839	0.247 (0.119)
2	2 (AA, AB)	22	0.833	0.241 (0.116)
3	13 (all other)	123	0.849	0.232 (0.110)

N, the number of individuals; *P*, the proportion of polymorphic loci; mean π the nucleotide diversity (heterozygosity), *SD* standard deviation.

Genomic structure of bank vole population

BAPS analyses identified three genomic clusters of bank vole in NE Poland (Table S2). The first consisted of the populations WL and LA, the second was made of individuals from AA and AB and the third, which was the largest, combined the remaining populations (Fig. 1). All except one individual of the first cluster belonged to the Carpathian mtDNA lineages. In addition, most specimens (about 73%) of the second cluster belonged to this lineage (Table S3). Most individuals (about 86%) in the third, largest cluster belonged to the Eastern mtDNA lineage (Table S3). The proportion of polymorphic loci and the nucleotide diversity, calculated for the three clusters, had similar values, although the number of the analysed samples in each of the first two clusters was about six times lower than that in the third cluster (Table 2). The F_{ST} values among pairs of the three clusters were low, although statistically significant (Supplementary Materials p. 12).

Cross-validation error values, calculated using ADMIXTURE, had the lowest and similar values for $K = 2-4$ (Fig. S3). For $K = 2$, the studied populations were divided into two genetic groups: one consisting of populations WL, LA, AA and AB, and the second combining the rest of the populations (Figs. 2, S4). This division was concordant with the distribution of the two main mtDNA lineages of the species. The population genetic structure for $K = 3$ indicated grouping similar to those indicated by the BAPS analysis (compare Fig. 1 and Figs. 2, S4). The genetic division for $K = 4$ revealed one more cluster, which occurred mainly in the northern part of the study area, where most of the individuals possessed mtDNA haplotypes of the Eastern lineage. Its frequency increased from east to west (Figs. 2, S4).

The PCA indicated three genetic clusters, similar to the groups indicated in BAPS and ADMIXTURE analyses ($K = 3$; compare Figs. 1, 2, 3 and S5). Further division, according to the PCA (PC2 – PC4), distinguished yet two other genetically different populations AJ and BJ (results not shown). These two populations were also indicated when the PCA was performed, excluding the most genetically different populations WL, LA, AA and AB (Fig. S5). In populations AJ and BJ, the private mtDNA haplotype H25 (in AJ), belonging to the Carpathian mtDNA lineage, and H55 (in BJ), of the Eastern lineage, occurred with high frequencies (Fig. S1).

Functional differentiation between the Eastern and the Carpathian lineages

The assembled transcriptome contains 252,000 sequences (with min length 200 bp), summing up to 135 M bp with $N_{50} = 650$ bp (half of the total length of the transcriptome is assembled into sequences longer than 650 bp). In total, we identified 25,000 sequences longer than 1 kb and 24,852 long open-reading frames within 21,961 sequences.

Twenty samples, originating from three populations, were sequenced, yielding on average of 8.5 M paired-end reads (2×100 bp) per individual. Among 7,981,698 sites matching initial coverage and quality thresholds, we identified 82,465 biallelic transcriptomic SNPs (we meant here all SNPs found in transcriptomes), including 11,091 and 21,870 SNPs, annotated as non-synonymous and synonymous, respectively. Transcriptomic SNPs were found in 5833 sequences, of which 3541 and 2669

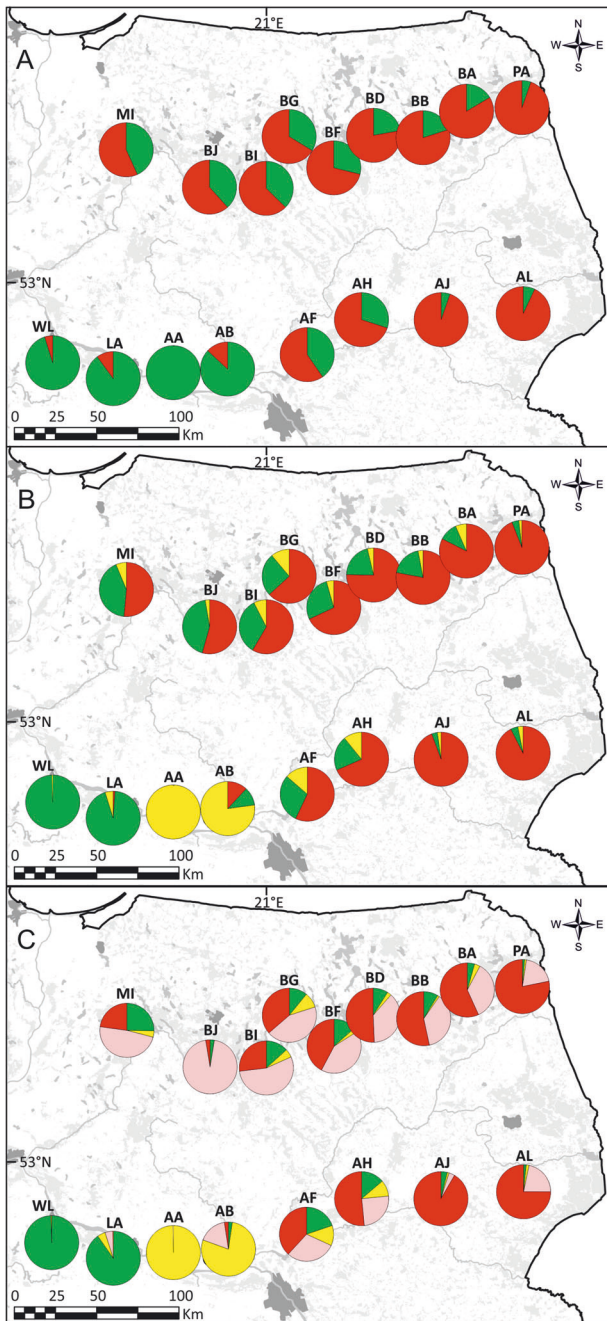


Fig. 2 Distribution of different genomic clusters identified in ADMIXTURE analyses. **A** $K=2$, **B** $K=3$, **C** $K=4$. Each colour means one cluster. Pie-chart shares correspond to the frequency of occurrence of each cluster in local populations.

included at least one synonymous and one non-synonymous SNP, respectively.

The transcriptomic SNP analysis clearly separated the three populations, with the samples from MI being placed on the PCA plot closer to the BIA population (Fig. 4). As expected, genetic variation was lower in non-synonymous sites, comparing with synonymous ones (Table 3). Genetic variation was elevated in BIA population when compared with the WL population (Table 3).

We then calculated F_{ST} between WL and BIA populations. The mean across F_{ST} of all transcriptomic SNPs equalled 0.056 ($SD=0.125$, Fig. 5). The mean F_{ST} , plus three times standard deviation, was used as a threshold to identify the most

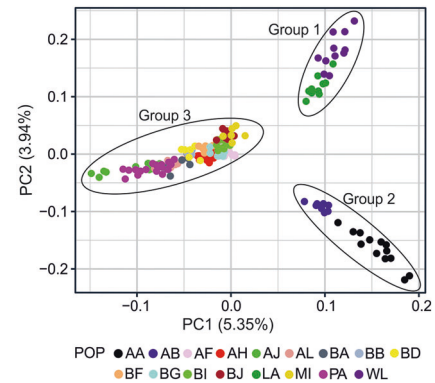


Fig. 3 Results of PCA based on the genomic data. Pop—populations of bank voles (acronyms of populations as in Fig. 1 and Table 1). Each point denotes one individual. Groups 1–3 indicate the populations identified in BAPS analyses (see Fig. 1 for comparison).

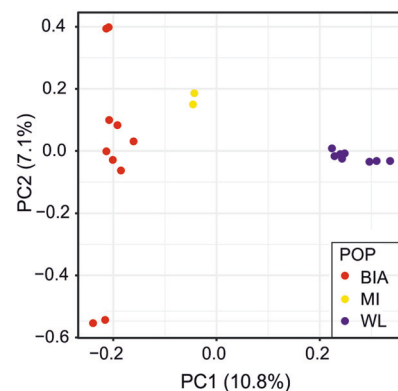


Fig. 4 Results of PCA based on the transcriptomic data. Acronyms of populations as in Fig. 1 and explanations as in Fig. 3.

differentiated transcriptomic SNPs ($F_{ST}=0.432$). We found 1661 transcriptomic SNPs above this threshold, 236 of which were annotated as non-synonymous. The differentiated non-synonymous transcriptomic SNPs were in 122 genes (Table S4). The most differentiated genes included complement factor H, spectrin beta-chain, haemoglobin subunit beta, beta-1-syntrophin, RNA-binding protein 40 and Nesprin-1.

DISCUSSION

Concordant genetic structure indicated by different genetic markers

The results of different analyses of the genomic structure of the bank vole populations showed that there were 2–4 SNP clusters of the species in the contact zone of their two mtDNA lineages in NE Poland. The divisions into two and three clusters were, to a large extent, concordant with the genetic structures revealed by mtDNA and microsatellite DNA analyses (compare Tarnowska et al. 2016, 2019). Most individuals originating from populations WL, LA, AA and AB belonged to the Carpathian lineage. They formed one or two separate SNP clusters, according to the ADMIXTURE, BAPS and PCA analyses and microsatellite clusters based on the STRUCTURE and Geneland analyses (compare Figs S1, S4–S6; Tarnowska et al. 2016, 2019). Most of the bank voles of the remaining one to three SNP clusters (depending on the type of analysis) belonged to the Eastern mtDNA lineage. Only in the case of one population – (MI) – did the majority of individuals possess the mtDNA of the Carpathian lineage (Fig. S1) and belonged to the same microsatellite STRUCTURE group as WL and LA populations;

Table 3. Parameters of genetic diversity within 3 populations of bank voles in NE Poland calculated for synonymous (syn), nonsynonymous (nsyn) and all sites identified in the sequenced heart transcriptomes.

Pop	N	P_{syn}	π_{syn} (SD)	P_{nsyn}	π_{nsyn} (SD)	P_{all}	π_{all} (SD)
BIA	10	0.793	0.212 (0.175)	0.753	0.181 (0.167)	0.777	0.201 (0.173)
WL	8	0.584	0.189 (0.195)	0.528	0.164 (0.189)	0.562	0.180 (0.193)
MI	2	0.366	0.200 (0.267)	0.319	0.174 (0.258)	0.350	0.192 (0.265)

N, the number of individuals; P, the proportion of polymorphic loci; π , mean nucleotide diversity (heterozygosity), SD standard deviation.

whereas, according to the Geneland (compare Fig. S6; Tarnowska et al. 2019), SNPs (BAPS and PCA) analyses, it formed one cluster with the majority of individuals in the Eastern mtDNA lineage. Furthermore, ADMIXTURE indicated that individuals of the MI population were hybrids of two to three SNP clusters. A similarly admixed population of bank voles was revealed in Denmark, where the individuals possessing mtDNA of the Eastern lineage had the closest genomic relationship with the SNP cluster originating from the Carpathian refugium (Marková et al. 2020). Furthermore, in the bank vole study in Britain, some discrepancy was reported between the distribution of different mtDNA lineages and variants of the haemoglobin genes (Kotlík et al. 2018). Such incompatibility between the phylogenetic patterns obtained using different genetic markers in those cases has been explained as an effect of admixture and partial replacement of individuals from one recolonising population by another wave of migrants (Kotlík et al. 2018; Marková et al. 2020).

The PCA of heart transcriptomic data divided the studied individuals into three groups, consistent with the genetic structure described earlier. These results confirm our hypothesis that the genetic structures obtained, based on the different genetic markers in the bank vole population in NE Poland, were concordant. A similar concordance between mtDNA and nuclear DNA analyses was also reported in other studies conducted not only on bank voles (Kotlík et al. 2014; Marková et al. 2020; see also Table S3 for comparison) but also on larger mammal species such as wolves *Canis lupus* (Czarnomska et al. 2013) and moose *Alces alces* (Niedziałkowska et al. 2016).

Different waves of recolonisation

The present-day genetic structure of bank voles in central and northern Europe has been shaped mainly by the postglacial colonisation of the continent by the species from several LGM refugia (Kotlík et al. 2006; Wójcik et al. 2010). Those refugial areas were located not only in the most southern parts of the continent but also in more northern regions in western, central and eastern Europe (Sommer and Nadachowski 2006; Queirós et al. 2019; Niedziałkowska et al. 2021b). In some temperate species, including the bank vole, the most important LGM refugial areas for contemporary populations inhabiting central, northern and eastern Europe were located in the Carpathians and the eastern part of the continent (Wójcik et al. 2010; Tarnowska et al. 2016; Marková et al. 2020), and not in the southern (Iberian, Apennine and Balkans) peninsulas, as has been found for many other mammalian species such as red deer (Niedziałkowska et al. 2011; Doan et al. 2022), brown bear *Ursus arctos* and some small mammals (e.g. Taberlet et al. 1998).

The analyses performed using fragments of mtDNA sequences and selected microsatellite loci indicated that there were two postglacial recolonisation waves of bank vole to north-eastern Poland: one from the Carpathian LGM refugium and one from the Eastern refugium (Tarnowska et al. 2016, 2019). Similar results were obtained in the analysis of mtDNA of the species, performed by Marková et al. (2020). The expansion time calculated, based on mtDNA, was much older for the Carpathian than for the Eastern lineages, which has been still in expansion (Tarnowska et al. 2016)

and in some areas (e.g. in present-day Denmark), the first of them was to some extent replaced by the latter (Marková et al. 2020). After the LGM, northern Europe was colonised by small mammals several times, before or during and after Younger Dryas (12,900 to 11,700 years BP) (Searle et al. 2009; Marková et al. 2020).

The results of studies, based on SNP data and much more dense sampling in northern and central Europe, were generally concordant with mtDNA data but also revealed more complex recolonisation patterns (this study, Marková et al. 2020). On the basis of the results obtained in our study (SNP genetic structure indicated by Admixture, BAPS and PCA, geographic distribution of different mtDNA haplotypes—compare Figs. S1, S7), we suppose that there were probably several recolonisation waves of bank voles to the north-eastern Poland: two from the Carpathian refugium and one, two or three from the Eastern refugium.

A majority of individuals belonging to the populations WL and LA (Group 1 indicated by BAPS) possessed mtDNA haplotypes H10 and H66 (Tarnowska et al. 2016), which were not found in more northern regions of Europe but were indicated in Slovakia, in the proximity of the Carpathian LGM refugium (Table S3). These haplotypes were located on different branches of the phylogeographic tree (Fig. S7) and network (Tarnowska et al. 2016) than haplotypes H1 and H7 occurring with high frequencies in populations AA and AB (Group 2 indicated by BAPS). The latter haplotypes were also found in more northern parts of the study area (e.g. in the population MI, Fig. S1) and in Scandinavia (Table S3). A significant genetic difference between Groups 1 and 2, indicated by both mtDNA and SNP data, is most probably an effect of different migration events from the Carpathian refugium. Such a scenario is concordant with multiple independent migration waves from this refugium, confirmed in this region of Europe by Marková et al. (2020).

Most of the performed analyses indicated that the majority of the studied populations belonged to one genetic cluster (e.g. Group 3 indicated by BAPS), originating from the Eastern LGM refugium, which is in agreement with the results of other phylogenetic studies of bank vole performed in this part of the continent (Wójcik et al. 2010; Tarnowska et al. 2016; Marková et al. 2020).

However, more detailed genetic analyses performed in this study indicated further structuring of this population. Admixture analyses ($K=4$) and PCA of Group 3 indicated additional one or two distinct genetic clusters. One of them occurred mainly in the northern and north-western part of the study area, where individuals possessing endemic (not present in other populations) mtDNA haplotypes were also found (Fig. S1 and Table S3). More genetic data of bank voles, especially from eastern Europe, are needed to confirm whether such substructure is an effect of different migration waves from the Eastern LGM refugium. However, such an explanation is in agreement with the results of the study by Marková et al. (2020), which indicated that there were several (including two in the eastern part of the continental Europe) independent migration waves from the Eastern refugium.

The results of our study are in agreement with other recently published genomic studies, showing that the contemporary phylogeographic pattern of bank vole in Europe is rather complex

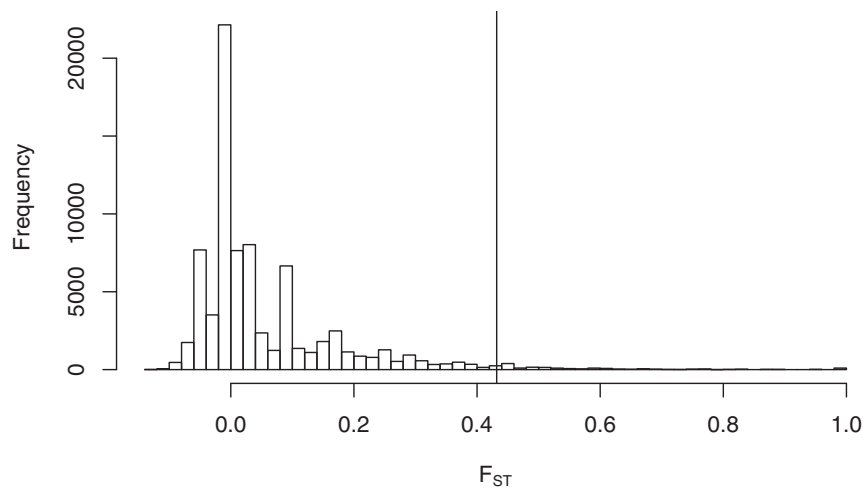


Fig. 5 Distribution of F_{ST} values between BIA and WL populations, calculated for transcriptome-derived SNPs. The vertical line shows the threshold used to identify the most differentiated SNPs between populations (e.g., mean plus $3 \times SD$).

and has been an effect of multiple migration waves from the LGM refugia, an admixture of different recolonising populations and, in some areas, the replacement of one mtDNA lineage by another (Kotlík et al. 2018; Marková et al. 2020; Horníková et al. 2021).

Owing to the admixture of different recolonisation waves of migrants from several distinct bank vole populations, the mtDNA genetic diversity of populations of various mammal species inhabiting central Europe is higher than that in other areas of the continent (e.g., Niedziałkowska et al. 2014), including the past LGM refugia (Veličković et al. 2016; Marková et al. 2020; Niedziałkowska et al. 2021a). In addition, the genetic diversity parameters (P , π) calculated for bank vole populations and SNP clusters in NE Poland were higher than such values obtained for populations inhabiting southern LGM refugia (compare Marková et al. 2020). Furthermore, the nucleotide diversity of bank vole populations and clusters in the studied area was higher than that in other populations and SNP clusters in this region of Europe (compare this study and Marková et al. 2020).

The genetic diversity parameters calculated for the SNP clusters originating from the Carpathian refugium were relatively higher than calculated for the specimens representing the recolonisation wave from the Eastern refugia. Furthermore, the genetic diversity of the Carpathian mtDNA lineage was higher than the Eastern mtDNA lineage in the bank vole population in the same study area (Tarnowska et al. 2016). Such a pattern can be explained by the closer spatial distance to the Carpathian than to the eastern LGM refugial area, which was localised in the Urals or in the area around the Black Sea (Deffontaine et al. 2005; Fløjgaard et al. 2009). On the contrary, this pattern was less clear when we compared the diversity of transcriptomic SNPs among individuals representing three different sampling localities (BIA and MI in the range of the Eastern mtDNA lineage and WL occurring in the range of the Carpathian mtDNA lineage), but this could be blurred by the low number of sequenced samples or by selection operating on functionally important regions in the genome.

Comparison of the spatial distribution of mtDNA haplotypes of bank voles found in north-eastern Poland and those identified in the Scandinavian populations of the species indicated that representatives of only some of the genetic clusters or lineages of the species (e.g. from the Carpathian but not from the Eastern LGM refugia) inhabiting the mainland were able to reach the northernmost parts of Europe such as southern Sweden and Norway in the postglacial period, probably before the opening of the Danish Straits about 8000 years ago (Björck 1995) (compare Table S3 and Fig. S7 in this study, Marková et al. 2020). The same

pattern was also noticed for other mammal species, e.g. moose (Niedziałkowska et al. 2014; 2016).

Adaptation to different environmental conditions

Associations between functional genomic variation and postglacial recolonisation history have not often been documented in European mammals. On the contrary, several studies indicated that the genetic structure and phylogenetic pattern of several mammal species in Europe were best explained by environmental conditions such as climatic factors (e.g., minimum January temperature most influenced the genetic structure of common and field *Microtus agrestis* voles and distribution of mtDNA haplogroups of weasel *Mustela nivalis*). This can be an effect of the adaptation of populations originating from distinct LGM refugial areas to different environmental conditions (McDevitt et al. 2012; Stojak et al. 2019, Stojak and Tarnowska 2019, and references therein). The study by Tarnowska et al. (2016) showed that the spatial distribution of the Carpathian mtDNA lineage of bank vole was positively correlated with the number of plant species originating from the Carpathian LGM refugium and mean temperature of July, whereas the distribution of the Eastern mtDNA lineage was negatively correlated. This indicated that during the recolonisation, the two lineages tracked environmental conditions similar to those occurring in their place of origin, where they survived glacial periods and were well adapted. Such a pattern was recently evidenced in the study of red deer (Niedziałkowska et al. 2021b).

Interestingly, the study of Kotlík et al. (2014, 2018) showed that bank voles inhabiting Britain originated from two recolonisation waves, which belonged to separate mtDNA clades and SNP clusters, and also differed in variants of haemoglobins. Individuals of the second wave, partly replacing the first one, possessed one of the haemoglobin variants which increased resistance of erythrocytes to oxidative stress (Kotlík et al. 2014). Among the analysed transcriptome SNPs in our study, one of the most differentiated genes between the studied populations ($F_{ST} = 0.88$) included the haemoglobin subunit beta, which is responsible for the structure of proteins in haemoglobin (Rouillard et al. 2016). Further analyses are needed to test whether different variants of this and other functional genes (e.g., complement factor H, spectrin beta chain) revealed in the bank voles studied play an important adaptive role. However, we did not find any significant differences ($F_{ST} \ll 1$) between individuals representing the two different populations in functional genes coding for protein-forming complexes, which are involved in the process of cell

respiration in mitochondria. Although individuals of the two populations belong to different mtDNA lineages (Tarnowska et al. 2016), the lack of significant differences between them in important functional genes of adaptive meaning (such as those involved in cell respiration) suggests that there were no genetic barriers that maintained the contact zone between these lineages and could have caused any incompatibility between mtDNA and nuclear DNA in hybrid individuals.

However, we cannot exclude the possibility that there are some significant differences in other functional genes between these two mtDNA lineages of the species, which we have not detected owing to the limitations of the F_{ST} outlier approach (e.g., only highly differentiated outliers can be detected). Furthermore, in this study, we focused only on looking for differences in functional genes involved in cell respiration, as we know that they play an important adaptive role. Further studies are needed to confirm whether other detected outlier genes could have an impact on the shape and width of the bank vole contact zone in north-eastern Poland.

CONCLUSIONS

The results of our study showed that the genetic structure of bank vole populations in NE Poland, indicated by the analyses of mtDNA, microsatellite DNA, whole genome and transcriptome SNPs analyses, was highly concordant. However, genome-wide SNP analyses have revealed more detailed genetic structure of the studied populations and have indicated more than two bank vole recolonisation waves in the study area. The lack of significant differences in functional genes of a confirmed adaptive role (such as involvement in cell respiration) between individuals representing two separate mtDNA lineages of the species allowed us to conclude that the contemporary genetic structure of bank vole populations and the width of their contact zone were not shaped by the genetic barriers but rather by climatic and environmental factors, as indicated by previous studies (Tarnowska et al. 2016, 2019). It is unlikely that the studied bank vole populations were isolated in separate LGM refugia for sufficient time for significant differentiation to evolve between the functional genes coding for protein-forming complexes, which are involved in the process of cell respiration in mitochondria, but also other genes, as revealed by low values of F_{ST} among different genomic clusters and functional genes. Thus, the correlation between the distribution of two mtDNA lineages of bank vole and environmental factors is the result of tracking of the similar climatic conditions and habitats by the populations recolonising Europe from separate refugial areas with different environmental characteristics.

Our study has demonstrated that a combination of mtDNA and genomic markers of different types (including both adaptive and neutral loci) can be used to reveal factors shaping genetic structure and phylogenetic patterns in species of interest. Only such complex genetic analyses will allow the phylogenetic relationship within the studied population to be reconstructed with confidence, which can have important implications, e.g. in delimitation of populations or conservation units, assessing the population fitness or impact of climatic oscillation on distribution and range of different genetic lineages of the species. Our findings contributed significantly to understanding the role of more northern (the Eastern and the Carpathian) LGM refugia in the evolutionary history of small mammals in central and eastern Europe. This study also provides some new data concerning the variability of functional genes in populations of European small mammal species originating from separate LGM refugia.

DATA AVAILABILITY

Data has been archived in the Dryad repository (reference transcriptome, annotation report, SNP data of bank voles: <https://doi.org/10.5061/dryad.jm63xsjg0>) and at NCBI (raw RNA sequence data of bank vole: BioProject ID PRJNA930868).

REFERENCES

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664
- Björck S (1995) A review of the history of the Baltic Sea, 13.0–8.0 ka BP. *Quat Int* 27:19–40
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120
- Catchen J, Amores A, Hohenlohe P, Cresko W, Postlethwait J (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genet* 1:171–182
- Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 11:3124–3140
- Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM et al. (2012) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a New Program, SnpSift. *Front Genet* 3:35–35
- Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B et al. (2009) The Last Glacial Maximum. *Science* 325:710–714
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4:1
- Corander J, Sirén J, Arjas E (2008) Bayesian spatial modeling of genetic population structure. *Computational Stat* 23(1):111–129
- Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinforma* 11:485
- Czarnomska SD, Jędrzejewska B, Borowik T, Niedziałkowska M, Stronen AV, Nowak S et al. (2013) Concordant mitochondrial and microsatellite DNA structuring between Polish lowland and Carpathian Mountain wolves. *Conserv Genet* 14(3):573–588
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158
- Deffontaine VLR, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB et al. (2005) Beyond the Mediterranean peninsulas: evidence for central European glacial refugia for a temperate mammal species, the bank vole (*Clethrionomys glareolus*). *Mol Ecol* 14:1727–1739
- Doan K, Niedziałkowska M, Stefaniak K, Sykut M, Jędrzejewska B, Ratajczak-Skrzatek U et al. (2022) Phylogenetics and phylogeography of red deer mtDNA lineages during the last 50 000 years in Eurasia. *Zool J Linn Soc* 194(2):431–456
- ESRI (Environmental Systems Resource Institute) (2012) ArcGIS 10.2. [computer program]. ESRI, Redlands
- Ellison CK, Burton RS (2006) Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution* 60:1382–1391
- Etter PD, Bassham S, Hohenlohe PA, Johnson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo V, Rockman MV editors. *Molecular methods for evolutionary genetics*. Springer, New York, p 157–78
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567
- Fløjgaard C, Normand S, Skov F, Svenning J-C (2009) Ice age distributions of European small mammals: insights from species distribution modelling. *J Biogeogr* 36(6):1152–1163
- Gonen S, Lowe NR, Cezard T, Gharbi K, Bishop SC, Houston RD (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *BMC Genom* 15(1):1
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Hirschfeld M, Kirschning CJ, Schwandner R, Wesche H, Weis JH, Wooten RM et al. (1999) Cutting edge: inflammatory signaling by *Borrelia burgdorferi* lipoproteins is mediated by toll-like receptor 2. *J Immunol* 163:2382–2386
- Horníková M, Markova S, Lanier H, Searle J, Kotlík P (2021) A dynamic history of admixture from Mediterranean and Carpathian glacial refugia drives genomic diversity in the bank vole. *Ecol Evolution* 11:8215–8225
- Hutterer R, Kryštufek B, Yigit N, Mitsainas G, Palomo L, Henttonen H et al. (2021) *Myodes glareolus* (amended version of 2016 assessment). IUCN Red List Threatened Species 2021:e.T4973A197520967. <https://doi.org/10.2305/IUCN.UK.20211.RLTS.T4973A197520967.en>. Downloaded on 15 October 2021
- Konczal M, Babik W, Radwan J, Sadowska ET, Koteja P (2015) Initial molecular-level response to artificial selection for increased aerobic metabolism occurs primarily through changes in gene expression. *Mol Biol Evol* 32(6):1461–1473
- Kondracki J (1994) *Geografia Polski. Mezoregiony fizycznogeograficzne*. Państwowe Wydawnictwo Naukowe, Warszawa
- 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

- Kotlík P, Marková S, Vojtek L, Stratil A, Šlechta V, Hyršl P et al. (2014) Adaptive phylogeography: functional divergence between haemoglobins derived from different glacial refugia in the bank vole. *Proc R Soc B: Biol Sci* 281:20140021
- Kotlík P, Marková S, Konczal M, Babik W, Searle JB (2018) Genomics of end-Pleistocene population replacement in a small mammal. *Proc R Soc B: Biol Sci* 285:20172624
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33:1870–1874
- Lane N (2009) On the origin of bar codes. *Nature* 462(7271):272–274
- Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* 135:1065–1073
- Li H, Durbin R (2010) Fast and accurate long-read alignment with burrows–wheeler transform. *Bioinformatics* 26(5):589–595
- Lima TG, Burton RS, Willett CS (2019) Genomic scans reveal multiple mito-nuclear incompatibilities in population crosses of the copepod *Tigriopus californicus*. *Evolution* 73:609–620
- Marková S, Horníková M, Lanier HC, Henttonen H, Searle JB, Weider LJ et al. (2020) High genomic diversity in the bank vole at the northern apex of a range expansion: The role of multiple colonisations and end-glacial refugia. *Mol Ecol* 29:1730–1744
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 17(1):10–12
- McDevitt AD, Zub K, Kawalko A, Oliver MK, Herman JS, Wójcik JM (2012) Climate and refugial origin influence the mitochondrial lineage distribution of weasels (*Mustela nivalis*) in a phylogeographic suture zone. *Biol J Linn Soc* 106(1):57–69
- Niedziałkowska M, Kończak J, Czarnomska S, Jędrzejewska B (2010) Species diversity and abundance of small mammals in relation to forest productivity in northeast Poland. *Écoscience* 17(1):109–119
- Niedziałkowska M, Hundertmark KJ, Jędrzejewska B, Niedziałkowski K, Sidorovich VE, Górny M et al. (2014) Spatial structure in European moose (*Alces alces*): genetic data reveal a complex population history. *J Biogeogr* 41:2173–2184
- Niedziałkowska M, Hundertmark KJ, Jędrzejewska B, Sidorovich VE, Zaleska H, Veeroja R et al. (2016) The contemporary genetic pattern of European moose is shaped by postglacial recolonization, bottlenecks, and the geographical barrier of the Baltic Sea. *Biol J Linn Soc* 117:879–894
- Niedziałkowska M, Tarnowska E, Ligmanowska J, Jędrzejewska B, Podgórski T, Radziszewska A et al. (2021a) Clear phylogeographic pattern and genetic structure of wild boar *Sus scrofa* population in Central and Eastern Europe. *Sci Rep*. 11:9680
- Niedziałkowska M, Doan K, Górny M, Sykut M, Stefaniak K, Piotrowska N et al. (2021b) Winter temperature and forest cover have shaped red deer distribution in Europe and the Ural Mountains since the Late Pleistocene. *J Biogeogr* 48(1):147–159
- Plis K, Niedziałkowska M, Borowik T, Lang J, Heddergott M, Tiainen J et al. (2022) Pan-European phylogeography of the European roe deer (*Capreolus capreolus*). *Ecol Evol* 12(5):e8931
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81:559–575
- Queirós J, Acevedo P, Santos JPV, Barasona J, Beltran-Beck B, González-Barrio D et al. (2019) Red deer in Iberia: Molecular ecological studies in a southern refugium and inferences on European postglacial colonisation history. *PLOS ONE* 14(1):e0210282
- R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, <http://www.R-project.org>
- Rouillard AD, Gunderson GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG et al. (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxf) 2016:ID baw100
- Searle JB, Kotlík P, Rambau RV, Marková S, Herman JS, McDevitt AD (2009) The Celtic fringe of Britain: insights from small mammal phylogeography. *Proc R Soc B: Biol Sci* 276(1677):4287–4294
- Singhal S, Moritz C (2012) Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a reforested lizard. *Evolution* 66:1474–1489
- Sommer RS, Nadachowski A (2006) Glacial refugia of mammals in Europe: evidence from fossil records. *Mammal Rev* 36:251–265
- Stojak J, McDevitt AD, Herman JS, Kryštufek B, Uhlíková J, Purger JJ et al. (2016) Between the Balkans and the Baltic: Phylogeography of a common vole mitochondrial DNA lineage limited to Central Europe. *PLOS ONE* 11:e0168621
- Stojak J, Tarnowska E (2019) Polish suture zone as the goblet of truth in post-glacial history of mammals in Europe. *Mammal Res* 64:463–475
- Stuglik MT, Babik W, Prokop Z, Radwan J (2014) Alternative reproductive tactics and sex-biased gene expression: the study of the bulb mite transcriptome. *Ecol Evol* 4:623–632
- Stojak J, Borowik T, Górny M, McDevitt AD, Wójcik JM (2019) Climatic influences on the genetic structure and distribution of the common vole and field vole in Europe. *Mammal Res* 64(1):19–29
- Sunnucks P, Morales HE, Lamb AM, Pavlova A, Greening C (2017) Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. *Front Genet* 8:25
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* 7(4):453–464
- Tarnowska E, Niedziałkowska M, Gerc J, Korbut Z, Górny M, Jędrzejewska B (2016) Spatial distribution of the Carpathian and Eastern mtDNA lineages of the bank vole in their contact zone relates to environmental conditions. *Biol J Linn Soc* 119(3):732–744
- Tarnowska E, Niedziałkowska M, Jędrzejewska B (2019) Genetic structure of bank vole populations in the contact zone of two lineages in north-eastern Poland. *Mamm Biol* 96:93–101
- Tarnowska E, Niedziałkowska M, Stojak J, Jędrzejewska B (2020) Polymorphism of TLR2 in bank vole populations in North Eastern Poland is not associated with *Borrelia afzelii* infection prevalence. *Mammal Res* 65(4):779–791
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21:3907–3930
- Tschirren B, Andersson M, Scherman K, Wester Dahl H, Mittl PRE, Råberg L (2013) Polymorphisms at the innate immune receptor (TLR2) are associated with *Borrelia* infection in a wild rodent population. *Proc R Soc B: Biol Sci* 280:20130364
- Wallis GP, Cameron-Christie SR, Kennedy HL, Palmer G, Sanders TR, Winter DJ (2017) Interspecific hybridization causes long-term phylogenetic discordance between nuclear and mitochondrial genomes in freshwater fishes. *Mol Ecol* 26:3116–3127
- Wójcik JM, Kawalko A, Markova S, Searle JB, Kotlik P (2010) Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: a case study of bank voles using museum specimens. *J Zool* 281:249–262
- Veličković N, Ferreira E, Djan M, Ernst M, Obreht Vidaković D, Monaco A et al. (2016) Demographic history, current expansion and future management challenges of wild boar populations in the Balkans and Europe. *Heredity* 117:348–357

ACKNOWLEDGEMENTS

This study was financed by the Foundation for Polish Science, project POMOST/2010-2/11 to M.N. (in the frames of the Program Innovative Economy, one of six national programs under the National Strategy Reference Framework, which was co-financed from the European Regional Development Fund of the EU) and (to E.T.) the National Science Centre – UMO-2014/13/N/NZ8/02472. M.K. was supported by the National Science Centre – grant no 2019/34/A/NZ8/00231. Computations were performed at the Poznan Supercomputing and Networking Center. We thank Kamila Plis for her help in preparing Fig. S2 and Steve Jones for the revision of the English language.

AUTHOR CONTRIBUTIONS

MN, BJ and WB conceived the ideas, MN applied for financial support, ET collected the samples, performed the laboratory work and statistically analysed the RAD-tag and mtDNA data sets, MK conducted the transcriptome analyses, TC and KG designed and supervised RAD-tag library preparation, sequencing and initial SNP filtering, ET and MK prepared the tables and figures, MN wrote the main manuscript, and all authors reviewed it.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-023-00600-1>.

Correspondence and requests for materials should be addressed to Magdalena Niedziałkowska.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.