Haplotype-resolved genome and population genomics of the threatened garden dormouse in Europe

Paige Byerly[1,2#*], Alina von Thaden[1,2#], Evgeny Leushkin[1,3], Leon Hilgers[1,3], Shenglin Liu[1,3], Sven Winter[4,5], Tilman Schell[1,3], Charlotte Gerheim[1,3], Alexander Ben Hamadou[1,3], Carola Greve[1,3], Christian Betz[6], Hanno J. Bolz[6], Sven Büchner[8], Johannes Lang[8], Holger Meinig[8], Eva Marie Famira-Parcsetich[8], Sarah P. Stubbe[8], Alice Mouton[9], Sandro Bertolino[10], Goedele Verbeylen[11], Thomas Briner[12], Lídia Freixas-Mora[13], Lorenzo Vinciguerra[14], Sarah A. Mueller[15], Carsten Nowak[1,2], and Michael Hiller[1,3,7*]

[1] LOEWE Centre for Translational Biodiversity Genomics, Senckenberganlage 25, 60325 Frankfurt, Germany
[2] Conservation Genetics Group, Senckenberg Research Institute and Natural History Museum Frankfurt, Clamecystraße 12, 63571 Gelnhausen, Germany
[3] Senckenberg Research Institute, Senckenberganlage 25, 60325 Frankfurt, Germany
[4] Senckenberg Biodiversity and Climate Research Centre, Senckenberganlage 25, 60325 Frankfurt am Main, Germany
[5] Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, Savoyenstr. 1, 1160 Vienna, Austria
[6] Bioscientia Human Genetics, Institute for Medical Diagnostics GmbH, Ingelheim, Germany
[7] Institute of Cell Biology and Neuroscience, Faculty of Biosciences, Goethe University Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt, Germany
[8] Justus-Liebig-University Giessen, Clinic for Birds, Reptiles, Amphibians and Fish, Working Group for Wildlife Research, Frankfurter Strasse 114, 35392 Giessen, Germany
[9] Socio-économie, Environnement et Développement (SEED), University of Liege (Arlon Campus Environment), Belgium
[10] Department of Life Sciences and Systems Biology, University of Turin, 10123 Torino, Italy
[11] Natuurpunt Studie vzw, Mammal Working Group, Michiel Coxiestraat 11, 2800 Mechelen, Belgium
[12] Naturmuseum Solothurn, Klosterplatz 2, 4500 Solothurn, Switzerland
[13] BiBio Research Group, Natural Sciences Museum of Granollers, C/ Francesc Macià 51, 08402 Granollers, Catalonia, Spain.
[14] Naturmuseum St.Gallen, Rorschacher Strasse 263, 9016 St. Gallen, Switzerland
[15] Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Germany

# contributed equally to the manuscript
* To whom correspondence should be addressed: paige.byerly@senckenberg.de, michael.hiller@senckenberg.de

Abstract

Genomic resources are important for evaluating genetic diversity and supporting conservation efforts. The garden dormouse (*Eliomys quercinus*) is a small rodent that has experienced one of the most severe modern population declines in Europe. We present a high-quality haplotype-resolved reference genome for the garden dormouse, and combine comprehensive short and long-read transcriptomics datasets with homology-based methods to generate a highly complete gene annotation. Demographic history analysis of the genome revealed a sharp population decline since the last interglacial, indicating that colder climates caused severe population declines prior to anthropogenic influence. Using our genome and genetic data from 100 individuals, largely sampled in a citizen-science project across the contemporary range, we conducted the first population genomic analysis for this species to investigate patterns of connectivity between regions and factors explaining population declines. We found clear evidence for population structure across the species' core Central European range. Notably, our data provide strong evidence that the Alpine population, characterized by strong differentiation likely due to habitat isolation, represents a differentiated evolutionary significant unit (ESU). Our data also show that the predominantly declining Eastern European populations show signs of recent isolation, a pattern consistent with a range expansion from Western to Eastern Europe during the Holocene, leaving relict populations now facing local extinction. Overall, our findings suggest that garden dormouse conservation may be enhanced in Europe through designation of ESUs.

Introduction

As genomic sequencing technology advances and costs decrease, high-quality reference genomes are becoming increasingly recognized as an essential resource for conservation genomics (Formenti et al. 2022; Brandies et al. 2019; Theissinger et al. 2023; Paez et al. 2022). When aligned to a reference genome assembly, high throughput genomic data can contribute to the understanding of past and contemporary population demographics (Gautier et al. 2016; du Plessis et al. 2023; Luo et al. 2023; Campana et al. 2016), reveal evolutionary patterns such as adaptive differentiation (Szarmach et al. 2021; Martchenko and Shafer 2023), and provide information about the current conservation status of wildlife species of conservation concern (Talla et al. 2023; Viluma et al. 2022). While reduced-representation sequencing (RRS) methodologies continue to be the more cost-effective and efficient means of generating genome-wide single nucleotide polymorphism (SNP) datasets for large numbers of samples (Wright et al. 2020; Peterson et al. 2012), alignment of RRS reads to a high-quality reference genome improves both the precision of SNP calls and the quantity of SNPs recovered when compared to *de novo* read alignment without a genome assembly (Rochette et al. 2019; Brandies et al. 2019). Reference genomes are also useful for data of closely-related species, with cross-species alignment shown to be highly effective (Takach et al. 2023; DeSaix et al. 2019; Burri et al. 2015; Nieto-Blázquez et al. 2022). Given the value of reference genome assemblies for population and conservation genomic analyses, sequencing of reference genomes for wildlife groups should be a focus of ongoing conservation efforts.

Highly informative sources of data such as reference genomes can have particular value for investigating understudied wildlife groups, as genomic data can contribute to baseline knowledge necessary for conservation planning. Within mammals, such foundational knowledge is lacking for many small-bodied mammal taxa such as rodents and insectivores, which have a known deficit of research (Verde Arregoitia 2016; Kennerley et al. 2021) that can be seen even in well-studied regions such as Central Europe (Pérez-Espona 2017). Indeed, general population trends of small mammals are not well understood in Europe (Lang et al. 2022; Pérez-Espona 2017; Gippoliti and Amori 2007; Bertolino et al. 2015), even though many species appear to be declining (Rammou et al. 2022; Gippoliti and Amori 2007; Lang et al. 2022; Reiners et al. 2014). Population genomic analysis could expand our understanding of population structure and the factors contributing to declines, but with the exception of a few flagship species such as the Eurasian beaver (*Castor fiber*; (Halley et al. 2021) it remains a limited tool in European rodent conservation.

The garden dormouse (*Eliomys quercinus*) is a small rodent species that exemplifies the small mammal conservation crisis in Europe. Once distributed across the

continent, the garden dormouse is now recognized to have experienced one of the most extensive modern population declines on the European continent, with an approximate 51% range contraction since the 1970s (Bertolino 2017). Currently, it is only considered common in five of the 26 countries that once comprised its historical range, and even within these refuge countries its distribution is patchy and highly localized (Bertolino 2017). Despite its known imperiled status, research on the garden dormouse has been limited, even in comparison to the three other European species in its family Gliridae (Lang et al. 2022)). In particular, the causes of its population decline are still unclear. Changing climate and intensification of land use have been suggested as possible influences; however, the demographic shift appears to have started prior to major 20th century landscape-level changes (Meinig and Büchner 2012; Bertolino 2017). Despite this possible influence of habitat loss, the garden dormouse also shows signs of being a habitat generalist, and this forest species has also adapted well to urban areas in some parts of its range while essentially vanishing from others (Meinig and Büchner 2012), making its decline an ongoing mystery.

Population genetic analysis has been identified as a major research need in garden dormouse conservation (Meinig and Büchner 2012), as genetic data could aid in understanding patterns of connectivity and gene flow between regions and uncover past population processes that may help explain current declines. Prior karyotyping and mitochondrial DNA analyses suggested the existence of four genetically distinct clades within Europe (Perez et al. 2013; Libois et al. 2012), although the existence of hybrid individuals indicated gene flow between clades (Perez et al. 2013). A better understanding of the population structure and genetic diversity within the hypothesized clades could help guide future management priorities and inform both regional and European-wide conservation efforts. Here, we present a high-quality reference genome for the garden dormouse, representing the first such complete genome for both the species and its genus. We then demonstrate the utility of the genome by conducting the first genome-wide analysis of population differentiation across the contemporary range of the garden dormouse.

Results

Haplotype-resolved genome assembly

To generate a reference-quality assembly for the garden dormouse, we used long-read sequencing on two PacBio Revio SMRT cells, producing a total of 124.4 Gb of High-Fidelity (HiFi) reads with a HiFi N50 read length of 14 kb. We used the Arima HiC protocol to generate 117 Gbp in long-range read pairs for scaffolding. Assuming a genome size

similar to other dormice (Zoonomia Consortium 2020; Böhne et al. 2023) of 2.5 Gb, the HiFi and HiC data have a coverage of ~50X and ~47X, respectively.

We used hifiasm (Cheng et al. 2021) in HiC mode to obtain two haplotype-resolved contig assemblies (Fig. S1). The contig N50 values of the two haplotypes are 51.7 and 45.4 Mb, respectively. This contiguity is at least 2.8-fold higher compared with other Sciuromorpha genome assemblies (Fig. 1A,B, Table S1). Scaffold N50 values are 108.1 and 107.5 Mb. Haplotype 1 contains the X chromosome, which was assembled as a chromosome-level scaffold. We identified 24 autosomes, inferring that the sequenced individual has a karyotype of 50 chromosomes (2*24 + 2 sex chromosomes) (Fig. 1C,D), which matches the range of 48 to 54 chromosomes identified for *Eliomys* so far (Perez et al. 2013). Overall, 99.4 and 99.6% of haplotype 1 and 2 are contained in chromosome-level scaffolds. These metrics exceed the standards set by the Vertebrate Genome Project (Rhie et al. 2021).

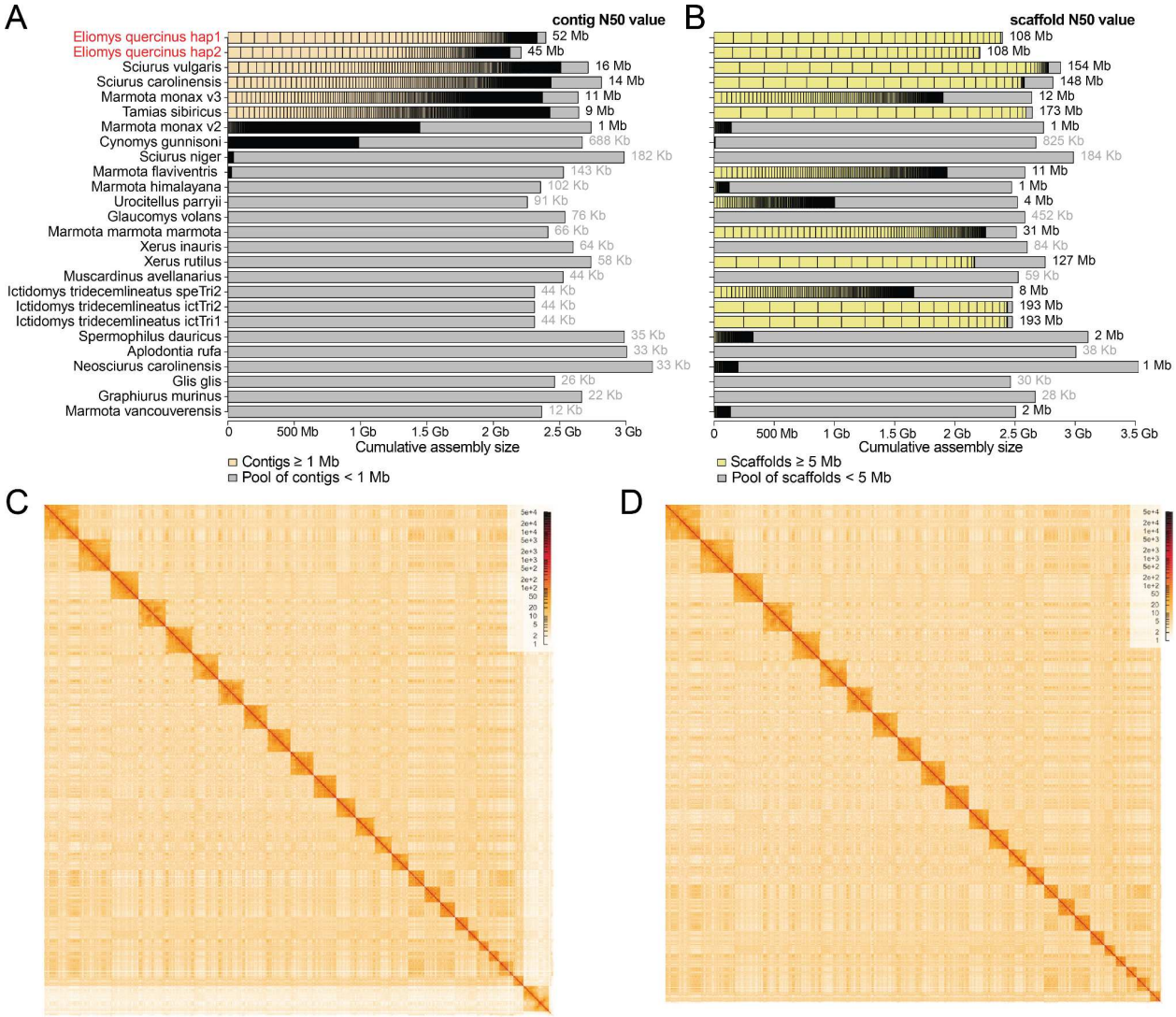

Figure 1: Contiguity and HiC maps of both haplotype assemblies. (A, B) Visualization of contig (A) and scaffold (B) sizes of the garden dormouse haplotype 1 and 2 (red font) and other Sciuromorpha genome assemblies. The N50 values are given on the right side. Contigs shorter than 1 Mb and scaffolds shorter than 5 Mb are not visualized individually, but shown as the grey portion of each bar.
(C, D) HiC density maps for haplotype 1 (C) and haplotype 2 (D).

Assembly quality and completeness

To estimate the assembly base accuracy, we used Merqury (Rhie et al. 2020) with the HiFi reads. Haplotype 1 and 2 have a QV (consensus quality) value of 60.46 and 60.24, indicating only one error per megabase. While these QV estimates represent an upper bound as they are based on the HiFi reads used for assembly, these values indicate a very high base accuracy in our assemblies.

To assess gene completeness, we first used compleasm 0.2.2 (Huang and Li 2023) with the odb10 set of 9,226 near-universally conserved mammalian genes. Counting completely detected genes present in a single copy, our assemblies contained 98.3% such genes for haplotype 1 and 90.16% for haplotype 2 that lacks the X chromosome. In comparison to other Sciuromorpha genome assemblies, our garden dormouse haplotype 1 is the second best in terms of this metric after the *Sciurus carolinensis* assembly (Mead et al. 2020) (Fig. 2A, Table S1). Second, we used TOGA (Kirilenko et al. 2023) to compare the status of 18,430 ancestral placental mammal coding genes across the garden dormouse and other Sciuromorpha assemblies. TOGA explicitly distinguishes between the two major assembly issues (incompleteness and base errors) by classifying genes into those that have an intact reading frame, those that have gene-inactivating mutations (frameshifts, stop codons, splice site mutations, exon or gene deletions) and those that have missing sequences due to assembly incompleteness or fragmentation. We found that 94.1% of the ancestral genes are intact and only 1.1% have missing exons (Fig. 2B, Table S1). In comparison to other assemblies in the order Sciuoromorpha, our garden dormouse haplotype 1 is the third best in terms of intact genes. Despite being top-ranked in the compleasm metric, the TOGA analysis reveals that the *S. carolinensis* assembly has substantially more genes with inactivating mutations, indicating a lower base accuracy (Fig. 2B).

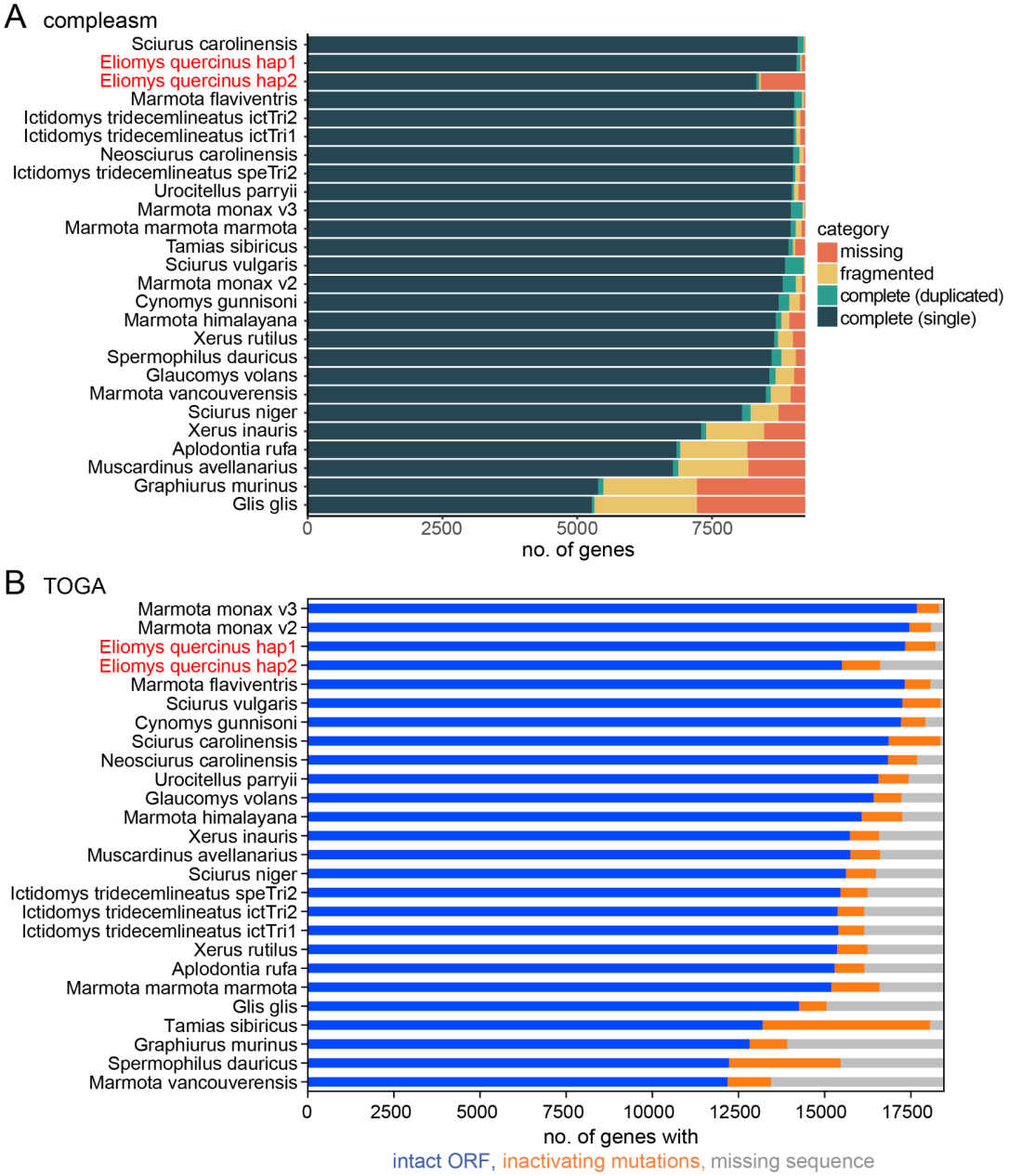

Figure 2: Comparison of gene completeness between our haplotype-resolved garden dormouse and other Sciuromorpha genomes.

(A) Fractions of detected, fragmented, and missing BUSCO genes (mammalian odb10 dataset, 9,226 genes), as classified by compleasm v.0.2.2 (Huang and Li 2023) across Sciuromorpha assemblies.

(B) TOGA classification of 18,430 ancestral placental mammal genes across Sciuromorpha assemblies.

Assemblies are ranked by the number of completely-detected (A) and by the number of intact (B) genes. For the garden dormouse, we show both haplotypes in red font and next to each other, but the X chromosome is only contained in haplotype 1. Species and assembly accessions are listed in Table S1.

Gene Annotation

To comprehensively annotate genes of the garden dormouse genome, we generated both short-read (RNA-Seq) and long-read (Iso-Seq) transcriptomics data. For RNA-Seq, we sequenced ~4.2 billion reads from eight different organs (heart, testis, liver, lung, spleen, muscle, kidney, and bladder) and for Iso-Seq, we sequenced ~1.6 million reads from six organs (heart, testis, liver, lung, kidney, and bladder) that often cover entire transcripts. The two types of transcriptome data were combined (see Methods) to produce transcript models. In addition to evidence from transcriptomics data, we generated genome alignments and used TOGA (Kirilenko et al. 2023) with the human GENCODE 38 and the mouse GENCODE M25 annotation (Frankish et al. 2021) as the reference to provide a homology-based annotation of the garden dormouse. Both types of transcript evidence were joined, which resulted in an annotation comprising 19,374 coding genes. Using compleasm in protein mode, we found that our gene annotation completely contains 99.91% of the mammalian BUSCO genes, indicating a very high completeness.

Demographic history of the garden dormouse

To gain insight into the demographic history of the garden dormouse and investigate potential effects of climatic changes in the Pleistocene, we inferred the effective population size (Ne) using pairwise sequentially Markovian coalescence (PSMC) (Li and Durbin 2011). We used a generation time of 1.5 years and a mutation rate of $5.7*10^{-9}$ per generation (see Methods).

Our results indicate relatively high Ne (> 130,000) and consistent population growth from 800 kya towards the last interglacial (Fig. 3). Population sizes reached their maximum of > 200,000 individuals around the last interglacial. Subsequently, Ne rapidly declined towards the last glacial maximum, reaching an estimated minimum Ne of less than 10,000 individuals around 10 kya. Together, this indicates that the garden dormouse thrived in warm periods, while a cooling climate with longer winters caused severe population declines. Consistent with this result, the Gliridae family thrived during earlier warm periods in the Miocene (23 to 5.3 Mya) (Nadachoswki and Daoud 1995)

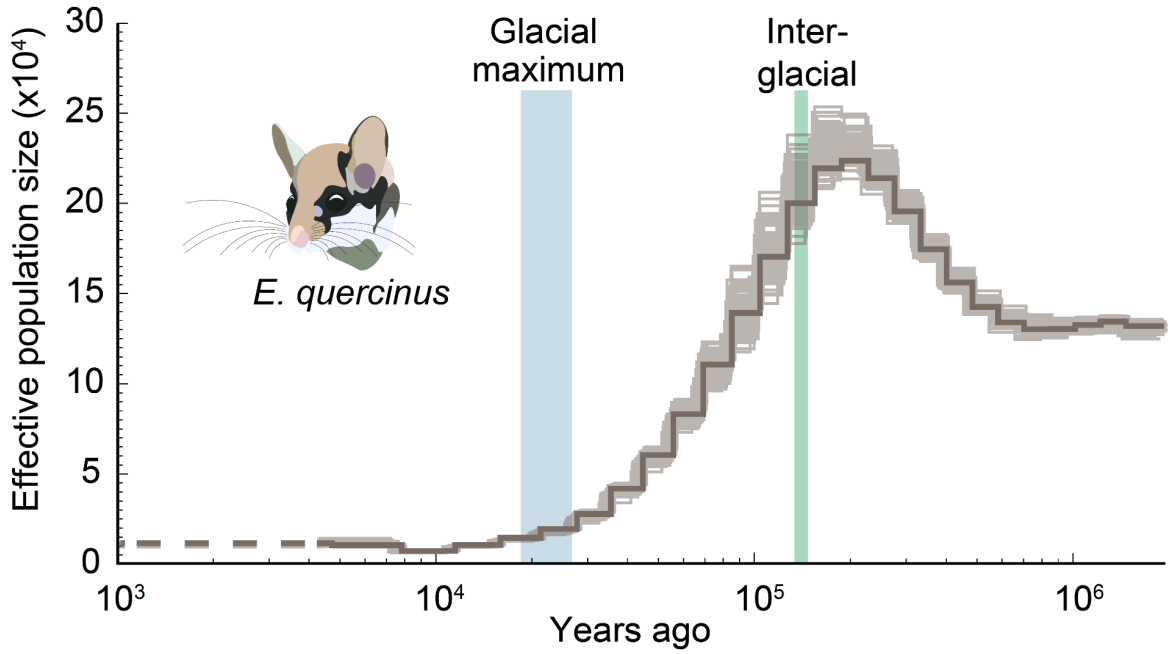

Figure 3: Demographic history of the garden dormouse.
Effective population size (Ne) of *E. quercinus* as estimated with PSMC. Ne in ten thousands is shown on the y-axis and time in years ago on the x-axis. Faint lines indicate uncertainty of inferred Ne based on 100 bootstrap replicates. Blue and green bars show the timing of the last glacial maximum and last interglacial, respectively. The most recent Ne estimate with reduced accuracy is displayed by a dotted line on the left.

Population genomics

To investigate population structure across the European range of the garden dormouse, we obtained *n* = 100 samples collected opportunistically between 1991–2020 for population genomic analyses and performed normalized Genotyping-by-Sequencing (nGBS). After sequencing, reads were aligned to the garden dormouse reference genome, and the resulting alignments were used to call single-nucleotide polymorphisms (SNPs). We assigned *a priori* population identity following in part the clades identified by (Perez et al. 2013) (Fig. 4A).

After removal of one individual with 98% missing data and one duplicate sample, we obtained data for a total of 41,175 SNPs from *n* = 98 samples with a mean per site depth of 14.75 X and mean frequency of missingness of 0.11 (Table S2). For analyses dependent upon allele frequencies, such as relatedness and PCA-based analyses, we pruned SNPs based on linkage disequilibrium, resulting in a dataset of 9,131 SNPs.

We first inferred relatedness among individual samples using the pruned SNP set and found six pairwise groups of first-order relations in the Harz population. Because

inclusion of related individuals can bias population genomic analyses, we randomly removed one individual from each pair above a kinship threshold > 0.35, which corresponds to a first-order familial relationship, for subsequent analyses (Table S2). All Russian samples were identified as second order relations (kinship > 0.20), indicating half-sibling or grandparent-offspring relationships. Therefore we removed all but one individual from Russia, retaining the sample with the lowest missing data (Table S2). As all individuals with kinship greater than the threshold were collected on the same day from the same locations, these kinship estimates likely reflect accurate estimates of relatedness between individuals. This left us with a final dataset of n = 86 samples (Table S2).

Phylogenetic analysis using RAxML showed that individuals from Northeastern Central Europe (Harz Mountains, Central, Northeast, Russia) form a single clade, separated from the Northwest, Alps, Italy, and Iberian Peninsula regions (Fig. S2). This tree conformed with the PCA of genetic distance among individuals and sampling regions. PCA showed a strong separation between the Alpine and other sampling regions, with this differentiation explaining 27.52% of the variation on PC1 (Fig. 5). Further spatial differentiation between the Northwest region and the other Central European sampling locations explained 5.20% of the variation on PC2. These results were consistent with findings of relatively high F_{ST} between most regional pairwise groupings (Table 1), suggesting that dispersal between spatially disjunct populations may be limited.

	NW	**ALP**	**HAR**	**CE**	**IB**	**NE**	**IT**
ALP	0.54						
HAR	0.15	0.69					
CE	0.08	0.68	0.21				
IB	0.28	0.62	0.44	0.32			
NE	0.13	0.69	0.16	0.11	0.39		
IT	0.43	0.64	0.61	0.53	0.36	0.58	
RUS	0.14	0.79	0.27	0.16	0.40	0.79	0.86

Table 1. Pairwise F_{ST} between eight garden dormouse sampling regions, corrected for sample size. Sampling regions are defined and abbreviated as Northwest (NW), the Alps (ALP), Harz Mountains (HAR), Central (CE), the Iberian Peninsula (IB), Northeast (NE), Italy (IT), and Russia (RUS).

However, the highest overall F_{ST} values were consistently recovered for pairwise comparisons of the Alpine group and all other populations, indicating that differentiation

between regions is not only driven by geographic distance, as the Alpine region is relatively central to other sampling sites (Fig. 4A).

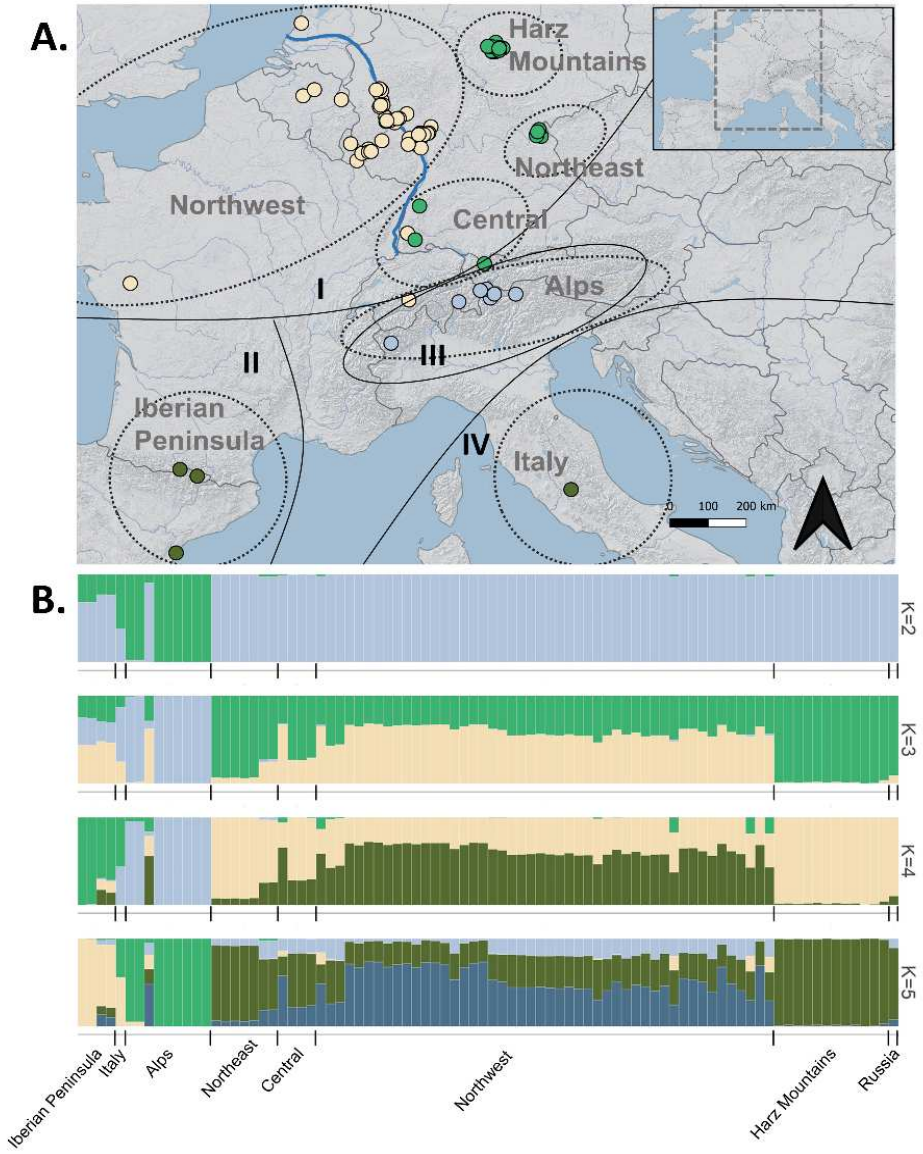

Figure 4. Garden dormouse population structure inferred from 47,115 nuclear SNP loci. (A) Group assignment via 10 PCs obtained by a discriminant analysis of principal components. (B) The predicted assignment of genetic clusters (K) 1–4 is shown by four different colors, the *a priori* sampling region identity is highlighted by dotted gray lines and text. The clades I-IV that were previously identified based on mitochondrial DNA (Perez et al. 2013) are outlined in black. One sample from Russia is not pictured. The Rhine River is outlined in blue, and the boundary of the study area is outlined in grey in the inset map.
(B) Genetic clustering inferred by Bayesian structure analysis for K ranging from 2 to 5 in STRUCTURE, with x-axes representing geographic sampling location and y-axes the proportion of group membership.

Genetic diversity was highest and inbreeding depression was lowest for the Iberian Peninsula (Table 2). Within the Central European region, genetic diversity was lower for population clusters situated east of the Rhine River (Northeast, Alps, Harz Mountains) compared to those located west of or near the river valley (Northwest, Central). This finding is consistent with the overall lower population sizes of garden dormouse populations located east of the Rhine River in Central Europe (Bertolino 2017), and suggests that eastern populations may be more susceptible to negative effects of recent population declines and range contractions, including isolation and loss of genetic diversity.

Region	*n*	A_R	H_{obs}	H_{exp}	F_{IS}
Alps	9	1.11	0.04	0.12	0.62
Central	4	1.20	0.16	0.21	0.15
Harz Mountains	12	1.14	0.13	0.14	0.08
Iberian Peninsula	4	1.24	0.22	0.25	0.06
Italy	1	1.03	0.03	–	–
Northeast	7	1.16	0.13	0.16	0.18
Northwest	48	1.22	0.17	0.22	0.2
Russia	1	1.07	0.07	–	–

Table 2. Sample size (*n*) and genetic diversity metrics among eight garden dormouse sampling regions, with observed (H_{obs}) and expected (H_{exp}) heterozygosity, allelic richness rarified by sample size (A_R), and the inbreeding coefficient F_{IS}.

We used DAPC and STRUCTURE to infer the number of genetic clusters within our samples. Both methods identified $K = 4$ as the most probable number of genetic clusters, although results were clearer for DAPC (Fig. S3A) than for STRUCTURE, which also suggested $K = 2$ and $K = 3$ as possible cluster numbers (Fig. S4). Cluster membership for DAPC was suggestive of relatively strong differentiation between the Northwest region versus other regions in Germany (Fig. 4A, Fig. S3B,C), while STRUCTURE indicated greater admixture among these regions at all values of *K* (Fig. 4B; Fig. S5). Both methods were consistent in differentiating the Alpine samples from all other regions (Fig. 4B; Fig. S5). Patterns observed for the full dataset were comparable for the STRUCTURE analysis subset by sample size (Fig. S6), indicating that our results were not influenced by unequal sample size among our sampling regions for regions where n > 1.

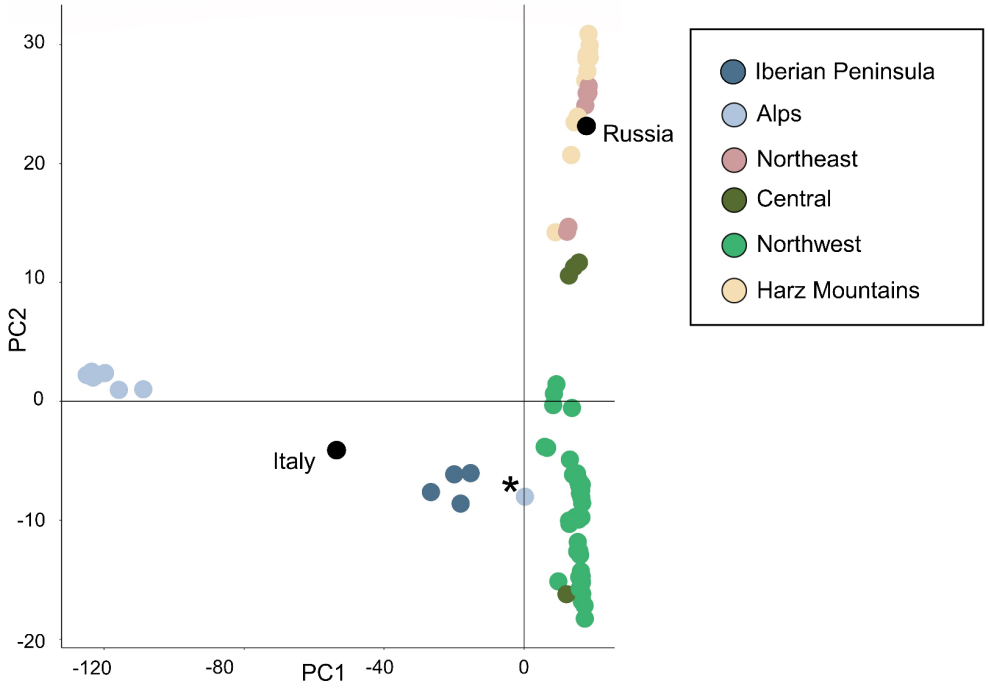

Figure 5. Principal component analysis (PCA) for the garden dormouse.
The PCA is based on 9,130 nuclear SNP loci, with colors representing the geographical origin. Results are plotted on the first two principal components, with PC1 explaining 27.52% and PC2 5.20% of variation within the dataset. Regions where *n* = 1 (Italy and Russia) are indicated by black circles, with region names labeled within the plot. One individual from the Alps clustered separately from the Alpine region and is marked with a *.

Overall, population genomic analyses consistently revealed little evidence for contemporary gene flow between the Alpine region and other populations, and the high degree of differentiation in this region may indicate long-term isolation of the Alpine group. Despite the clear isolation of the Alpine group, all analyses were consistent in separating one individual from other samples collected from the Alpine region (G190088CH). This individual, which was collected north of the main Alpine ridge, instead showed strong evidence for admixture with the Northwest and Harz Mountains regions (Fig. 4; Fig. 5).

Discussion

This study makes a valuable contribution to the conservation of small mammals in Europe by providing a high-quality reference genome for the garden dormouse, the first for its genus. This study was performed in the frame of the citizen-science focused project *Spurensuche Gartenschläfer* ('In Search of the Garden Dormouse'), which is the first

over-regional conservation and research project for this species in Germany (Meinig et al. 2023). The project aims at restoring garden dormouse populations, as the species continues to decline across much of its range (Lang et al. 2022). Using this assembly and population genomics data from samples across the species' range, we were able to investigate garden dormouse demographic history, as well as contemporary population structure and genetic diversity parameters.

The inferred demographic history indicates that garden dormice underwent a substantial population bottleneck already prior to anthropogenic influence, with rapid decline and long-term low Ne following the last interglacial. Our data suggest that warm periods around the last interglacial period supported high effective population sizes, and that hibernation did not save garden dormouse populations from declining during the last ice age. Previous paleontological data suggested that Gliridae diversified during periods of high glaciation (Lu et al. 2021). This observation has been interpreted as a sign that this family thrived during cold periods, likely because hibernation was an evolutionary advantage to survive cooling climates (Lu et al. 2021). In contrast, our results provide evidence that garden dormouse populations declined by more than 90% as the climate cooled down as it approached the last glacial maximum, raising the possibility that the observed lineage diversification during glaciation might rather have resulted from population fragmentation. This finding is in line with longer winters leading to reduced fitness following later emergence from hibernation (Lane et al. 2012). Similar results of declining populations after the last interglacial period have further been described for other Sciuromorpha, including the hibernating Gunnison's Prairie Dog *Cynomys gunnisoni* (Tsuchiya et al. 2020) and the Southern Flying Squirrel *Glaucomys volans* (Wolf et al. 2022)

Our population genomics analysis found clear evidence for population structure across the species' core Central European range, with strong differentiation in particular for the Alpine region. Our findings of high genetic differentiation and four potential genetic clusters confirm and extend prior genetic research, which also identified four garden dormouse lineages within a similar sampling area based on mitochondrial DNA haplotypes and chromosome number differences (Perez et al. 2013). We found the Alpine region south of the main Alpine ridge to be strongly differentiated from all other sampling regions, as evidenced by both model-based (STRUCTURE), and non-model-based (PCA) approaches. We also found differentiation of regions south of the Alpine range, which supports that this mountain range may act as a barrier to garden dormouse dispersal (Perez et al. 2013). One specimen collected within our Alps boundary showed admixture with more Northern regions, but, as this specimen was collected north of the main Alpine ridge, it is probable that this region is connected to the more northern populations, rather than this individual accurately representing admixture within the highly differentiated Alpine population. In contrast to prior research, with our more

comprehensive datasets, we were also able to detect additional moderate population structure within the Central European clade (Fig. 4), showing a clear pattern of differentiation between east and west Central Europe. Given the varying degrees of differences between these regions, it is likely that these patterns of genetic structure have arisen through different forces, with the Alpine population representing a potential evolutionary significant unit (ESU), while structure within Northern Europe may reflect modern population pressures resulting from habitat fragmentations and population isolation.

The range contraction of the garden dormouse is known to be centered in the eastern part of its range, with the species becoming more common and populations more contiguous as one moves further west (Bertolino 2017; Meinig and Büchner 2012). Fossil evidence indicates that the garden dormouse originated in the Iberian Peninsula and spread out into Central Europe during the Holocene, which may mean that remaining garden dormice in Eastern Europe represent small, relict populations on the edge of extirpation (Anděra 1986). It is possible, then, that the differentiation of these more eastern populations may therefore reflect genetic drift resulting from isolation of these populations in the 20th century. Support for this possibility is provided by our preliminary findings of lower genetic diversity and greater rates of inbreeding for the more easterly populations in Germany (Harz mountains and Northeast), while the Central population shows evidence of admixture with the larger and more contiguous Northwest population, thereby potentially increasing its genetic diversity through interbreeding. Larger sampling efforts are required to substantiate these findings that are based on rather low sample numbers. A similar east-west pattern of differentiation has also been observed in the garden dormouses' spatially overlapping relative, the hazel dormouse (*Muscardinus avellanarius*), which is distributed as two separate lineages or 'partial ESUs', delineated along a similar geographic divide as the garden dormouse (Leyhausen et al. 2022; Mouton et al. 2017). Unlike the hazel dormouse, however, this east-west pattern of differentiation was not reflected by mtDNA in the garden dormouse (Perez et al. 2013), indicating that the division among lineages may be a more recent phenomenon resulting from habitat fragmentation and/or demographic changes such as population declines.

In contrast to the Central European regions, the Alpine population shows a strong signal of genetic differentiation, which confirms prior mtDNA genetic distance estimates (Perez et al. 2013). Within the Alpine region, the garden dormouse is found in both subalpine and montane habitats, where it occupies coniferous as well as deciduous forest habitat (Bertolino 2017). The strong genetic differentiation and the highly differentiated habitat may indicate the potential for longer-term genetic separation, possibly arising from adaptation to local habitat conditions or climate. Evidence for the Alpine population as a potential ESU is further provided by the lower rates of genetic diversity and high estimate of inbreeding indicative of reproductive isolation of the population. Perez et al. (2013) also

found lower nucleotide diversity for this region, and surmised that this may have resulted from either population bottlenecks or recent population expansion. By contrast, the relatively high genetic diversity and low inbreeding of the Iberian population likely reflect a rather large population size and contiguous breeding range of the species in the warmer Mediterranean regions, where it is still considered relatively common (Bertolino 2017). Given that we only have one sample from the Italian mainland, our data are not sufficient to infer whether the Iberian and Italian populations are differentiated, and further sampling of this region is needed to establish patterns of relatedness of garden dormouse within the Mediterranean.

Although the causes of the garden dormouse's east-to-west pattern of range contraction aren't fully understood, landscape level changes driven by land use and climate change have been partially implicated (Bertolino 2017). However, it is unclear what specific role climate might play in this contraction. For the ecologically-similar hazel dormouse, warmer and wetter winters have been found to have a negative effect on adult survivorship, possibly because warmer temperatures cause animals to wake during hibernation, thereby expending energy and causing a reduction in fat reserves (Combe et al. 2023). Garden dormice hibernate in regions of harsh winter, and their life histories appear strongly tied to winter temperatures, as well as the onset and duration of the season (Bennett and Richard 2021; Mahlert et al. 2018). Timing and duration of activity periods and hibernation vary widely based on local climate conditions, with Mediterranean populations hibernating only 1–2 months per year, if at all, while Alpine populations may hibernate for 7 months per year (Bertolino et al. 2001). While populations appear to be limited by extreme cold (Schaub and Vaterlaus-Schlegel 2001; Bertolino et al. 2001), warm weather during hibernation may also cause stressors that could inhibit population growth in regions where dormouse are more adapted to cold, particularly where winter food resources are not available (Giroud et al. 2023). Further investigation of the role of climate in shaping the garden dormouse's population structure via local adaptation may be useful in predicting the future viability of the species with ongoing climate change.

Conclusion

The garden dormouse is currently listed as "Near Threatened" in the IUCN Red List despite compelling evidence that recent range reductions necessitate relisting the species as "Vulnerable" (Bertolino 2017). Our findings provide strong evidence for the designation of the Alpine subpopulation of the garden dormouse as a separate ESU, given that it fulfills the criteria set by (Moritz 1994) of reciprocal mtDNA monophyly (Perez et al. 2013) and significant nuclear DNA divergence. Overall, findings from this study suggest that garden dormouse conservation may be enhanced in Europe through the designation of differentiated populations as ESUs. Further research is needed to

472 determine if these ESUs are experiencing local adaptation to environmental factors such
473 as temperature, and how these potential adaptations may influence future population
474 viability in the face of a changing climate. Our annotated genome will enable comparative
475 analyses and addressing open evolutionary questions of the *Eliomys* species complex,
476 as well as to inform specific conservation planning and management for the garden
477 dormouse in Europe. It is our hope that the findings outlined in this study can help guide
478 the direction of future research for the garden dormouse, while also highlighting the need
479 for further protection within regions of population isolation and/or low genetic diversity.
480
481
482 # Materials and Methods
483
484 ## Sample collection
485 Tissue samples for nGBS were collected from across the European distribution in
486 the frame of a larger conservation project on *Eliomys quercinus* in Germany. Samples
487 originated from Germany (*n* = 66), other Central European countries (*n* = 25), and Russia
488 (*n* = 7) and were collected between 1991-2020 (see Table S2). All samples were taken
489 opportunistically from dead-found individuals brought to museums or animal shelters,
490 during monitoring, or during citizen-science activities, and all in compliance with
491 respective local and national laws.
492
493 ## Genome sequencing
494 To sequence a high-quality genome, we used a male individual collected in March
495 2023 in Mainz that was euthanized by a veterinarian due to injuries.
496 High molecular weight genomic DNA was extracted from heart tissue according to
497 the protocol of (Sambrook and Russell) for the first library and from lung tissue using the
498 PacBio Nanobind Tissue Kit (PacBio, Menlo Park, CA, USA) for the second library. DNA
499 concentration and DNA fragment length were assessed using the Qubit dsDNA BR Assay
500 kit on the Qubit Fluorometer (Thermo Fisher Scientific) and the Genomic DNA Screen
501 Tape on the Agilent 4150 TapeStation system (Agilent Technologies), respectively. A
502 SMRTbell library was prepared for each tissue according to the instructions of the
503 SMRTbell Express Prep Kit v3.0. Total input DNA was approximately 4 µg per library.
504 Quality control of the final libraries was performed on an Agilent Femto Pulse system
505 (Agilent, Waldbronn Germany). Annealing of sequencing primers, binding of sequencing
506 polymerase, and purification of polymerase-bound SMRTbell complexes were performed
507 using the Revio polymerase kit (PacBio, Menlo Park, CA, USA). Loading concentration
508 for sequencing was 225 pM. Long-read whole-genome sequencing was performed on a
509 PacBio Revio instrument (PacBio, Menlo Park, CA, USA), 24 h HiFi sequencing, 15 kb

median read length, Q30 median read quality. HiFi read BAM files (128 Gb in total) were used for genome assembly.

To generate long-range data for scaffolding, we used the Arima High Coverage Hi-C Kit v01 (Arima Genomics) according to the Animal Tissue User Guide for proximity ligation using approximately 110 mg of heart tissue from the same individual. The proximally-ligated DNA was then converted into an Arima High Coverage HiC library according to the protocol of the Swift Biosciences® Accel-NGS® 2S Plus DNA Library Kit. The fragment size distribution and concentration of the Arima High Coverage HiC library was assessed using the TapeStation 4150 (Agilent Technologies) and the Qubit Fluorometer and Qubit dsDNA HS reagents Assay kit (Thermo Fisher Scientific, Waltham, MA), respectively. The library was sequenced on the NovaSeq 6000 platform at Novogene (UK) using a 150 paired-end sequencing strategy, resulting in an output of 180 Gb.

Transcriptome sequencing

Total RNA was isolated from bladder, heart, kidney, liver, lung, muscle, spleen, and testis from the same individual using TRIzol reagent (Invitrogen) according to the manufacturer's instructions. The quality and concentration of each extraction was assessed using the TapeStation 4150 (Agilent Technologies) and the Qubit Fluorometer with the RNA BR Reagents Assay Kit (Thermo Fisher Scientific, Waltham, MA). The RNA extractions were then sent to Novogene (UK) for Illumina paired-end 150 bp RNA-seq of a cDNA library (insert size: 350 bp) with a respective expected output of 9 Gb.

For the preparation of Iso-seq libraries using the SMRTbell Prep Kit 3.0, only RNA extractions with an RNA integrity number (RIN) > 7 were used, as recommended by the manufacturer's protocol. Two libraries were prepared by pooling RNA from kidney, liver, lung or bladder, heart and testis at approximately 90 ng each, with the exception of lung RNA at approximately 60 ng, resulting in two pooled libraries. These two Iso-Seq libraries were then each loaded onto an SMRT cell sequenced in CCS mode using the Sequel System IIe with the Sequel II Binding Kit 3.1 (Pacific Biosciences, Menlo Park, CA) at the Genome Technology Center (RGTC) at Radboudumc (Nijmegen, The Netherlands). The libraries were loaded to an on-plate concentration of 80 pM each by diffusion loading.

Genotyping-by-Sequencing

Tissue samples for nGBS were extracted using the DNeasy Blood & Tissue Kit (Qiagen) with slight protocol modifications to increase yield of RNA-free genomic DNA as described in (Leyhausen et al. 2022). Extracts were quantified with fluorometry (Qubit) and spectrophotometry (NanoDrop) and screened for DNA integrity using gel electrophoresis to identify samples of sufficient quality for NGS. Selected samples (*n* =

100) with ≥ 300 ng of high molecular weight DNA were sent for nGBS at LGC Genomics GmbH (www.biosearchtech.com). Details on enzymatic digestion, library preparation and nGBS protocols can be found in (Leyhausen et al. 2022; Arvidsson et al. 2016).

Genome assembly

We produced haplotype-specific assemblies from HiFi reads using hifiasm (v.0.19.5), integrating HiC reads for phasing (Cheng et al. 2021). Both haplotypes were subject to filtering procedure using Foreign Contamination Screen Tool (FCS) (Astashyn et al. 2023). Contamination free assemblies were then polished to remove unambiguous heterozygous sites. Specifically, we mapped HiFi reads back to each assembly with minimap2 v.2.26 (Li 2018), removed duplicates using picard MarkDuplicates tool v.3.1.0 (http://broadinstitute.github.io/picard), and called variants using DeepVariant v1.5.0 (Poplin et al. 2018). We then filtered for sites with genotype 1/1 and a 'PASS' filter value, meaning that all or nearly all reads support an alternative sequence at this position and passed DeepVariants internal filters. Finally, we corrected corresponding nucleotide sites in the assembly using bcftools consensus v.1.13 (Li 2011).

To scaffold the polished haplotype assemblies, we first mapped the Arima HiC reads to each of the two assemblies independently using chromap v.0.2.5 (Zhang et al. 2021) and then processed mapped reads with yahs v1.1a (Zhou et al. 2023). Finally, we performed manual curation of the two scaffolded haplotypes jointly analyzing them in PretextView v.0.2.5 (https://github.com/wtsi-hpag/PretextView). To this end, we remapped the HiC reads to both haplotype assemblies concatenated together using again chromap but allowing for multi-mapping reads (-q 0) to avoid discarding information in regions identical between two haplotypes. We also identified telomeric sequences with tidk v.0.2.31 (https://github.com/tolkit/telomeric-identifier) and where necessary corrected wrong contigs orientations to have telomeres in the ends of resulting scaffolds. To finalize changes made via manual dual haplotype curation, as recently proposed by the Vertebrate Genome and Darwin Tree of Life Project, we used the rapid curation framework (https://gitlab.com/wtsi-grit/rapid-curation/-/tree/main) from the Genome Reference Informatics Team (Howe et al. 2021).

Repeat masking and whole-genome alignment

RepeatModeler version 2.0.4 (default parameters) was used to generate a *de novo* repeat library for the garden dormouse assembly. RepeatMasker 4.1.0 (parameters -engine ncbi) was used with the resulting library to soft-mask the genome.

We followed our previous workflow to align the human (hg38 assembly) and the mouse (mm10) assembly to both garden dormouse haplotype assemblies (Sharma and

Hiller 2017; Blumer et al. 2022). Briefly, we used LASTZ version 1.04.15 with sensitive parameters (K = 2400, L = 3000, Y = 9400, H = 2000, and the LASTZ default scoring matrix) (Harris 2007), axtChain (default parameters except linear-Gap=loose) to compute co-linear alignment chains (Kent et al. 2003), RepeatFiller (default parameters) to capture additional alignments between repetitive regions (Osipova et al. 2019), and chainCleaner (default parameters except minBrokenChainScore = 75,000 and -doPairs) to improve alignment specificity (Suarez et al. 2017).

Gene annotation

We used TOGA (Kirilenko et al. 2023) with the alignment chains and the human GENCODE 38 and the mouse GENCODE M25 annotation (Frankish et al. 2021) to annotate coding genes. We ran TOGA for both garden dormouse haplotypes.

Raw RNA-seq data were processed with fastP v0.23.4 (Chen 2023; Chen et al. 2018) to remove adapters and low quality (< Q15) bases. Processed reads were mapped to a first haplotype assembly with hisat2 v.2.2.1 (Kim et al. 2019). For Iso-Seq reads a circular consensus was generated from the subreads using ccs v.6.4.0 (https://github.com/PacificBiosciences/pbbioconda). Adapter trimming and base quality filtering was performed with lima v2.2.0 (https://github.com/PacificBiosciences/pbbioconda). Processed reads were cleaned from polyA tails and polymerase switching artifacts using 'refine' command from the isoseq package v.4.0.0 (https://github.com/PacificBiosciences/IsoSeq/blob/master/isoseq-clustering.md) and then clustered together and aligned to generate consensus reads using the 'cluster' command from the isoseq package. Additionally, reads were filtered for alignment artifacts using a custom perl script. To combine transcriptome information from RNA-seq and Iso-seq, we then used a hybrid option (--mix) in stringtie v.2.1.2 (Shumate et al. 2022).

To integrate transcriptomics and homology-based transcript evidence, we first merged human and mouse TOGA, keeping only transcripts with an intact reading frame. We then added transcripts classified as partially intact and uncertain loss, but only if they do not overlap intact transcripts. The RNA- and Iso-seq merged transcriptome data was then used to add UTRs to TOGA-annotated coding transcripts with a compatible exon-intron structure. Finally, we added transcriptome data for genomic loci that don't have a TOGA transcript prediction to incorporate lineage-specific and non-coding transcripts into the annotation.

Demographic history

We used PSMC (Li and Durbin 2011) to infer the demographic history of the garden dormouse across the past one million years. Minimap2 v.2.26 (Li 2018) was used to map the PacBio HiFi reads to haplotype 1. Heterozygous sites were called with DeepVariant

v1.2.0 (Poplin et al. 2018) using the PacBio model. A consensus genome sequence where the heterozygous sites were represented with IUPAC ambiguity codes was generated using bcftools v1.17 (Danecek et al. 2021). The consensus sequence was converted to the input format of PSMC by applying a bin size of 50 bp.

For PSMC, we included data from all chromosome-size scaffolds except for the X chromosome (scaffold 4). PSMC was run using the parameter settings *-N25 -t15 -r5 -p "4+25*2+4+6"*. We also explored including more parameters (*-p "4+45*2+4+6"* and *-p "4+65*2+4+6"*) as well as splitting the first parameter (e.g., *-p "2+2+25*2+4+6"*). None of these settings changed the general results, so we chose to use the default parameters. Finally, we calibrated PSMC results using a generation time of 1.5 years and a mutation rate of $5.7*10^{-9}$ per generation. This mutation rate is based on the assumption that garden dormice exhibit a similar mutation rate as mice, for which the rate is well established (Uchimura et al. 2015; Milholland et al. 2017). The mean generation time (1.57 years) was calculated based on a yearly reproductive cycle starting at age one, a maximum age of 5 years, and an average yearly survival rate of 0.38 (Schaub and Vaterlaus-Schlegel 2001).

Population genomics

The raw GBS data were trimmed using fastp v0.23.4 (Chen et al. 2018) with enabled base correction and low complexity filter to remove sequencing adaptors and polyG stretches at the end of reads. A 4 bp sliding window was employed to detect regions of poor quality (Phred score < 15). We removed reads if they fit into one of the following categories: read length below 36 bp, reads with > 40% low-quality bases, and reads with 5 or more undetermined bases (Ns). The trimmed reads were then mapped against haplotype 1 of our assembly using bwa-mem v0.7.17-r1188 (Li 2013). Mate coordinates in the resulting mapping files were filled in, the files were sorted by position, and indexed using samtools v.1.18 (Danecek et al. 2021). Variant calling was performed with the ref_map.pl pipeline for reference aligned reads in stacks v2.65 (Rochette et al. 2019) with additional filtering parameters given for the populations program: minor allele frequency of 5%, a minimum percentage of individuals per population of 80, and writing one random SNP per locus (-X "populations: -r 0.80 --min-maf 0.05 --write-random-snp"). Additional variant filtering was performed with VCFtools v0.1.16 (Danecek et al. 2011) to remove indels and only keep biallelic SNPs with a depth per individual between 6 and 50 (flags: '--remove-indels --min-alleles 2 --max-alleles 2 --min-meanDP 6 --max-meanDP 50 --minDP 6 --maxDP 50'). After filtering, we removed samples with > 75% missing data to generate our final VCF file.

Past mitochondrial DNA work has classified four distinct garden dormouse clades within its European range (Iberian, Italian, Western European, and Alpine; (Perez et al. 2013). We followed these clade designations in assigning *a priori* population identity but

expanded the 'Western European' category by delineating individuals within this clade as Northwest, the Harz Mountains, Northeast, Central, and Russia (Fig. 4A) as aligning to known garden dormouse distribution within this region (Meinig and Büchner 2012).

Because linkage disequilibrium (LD) can influence some analyses such as principal components-based analyses (PCAs) and estimating relatedness between individuals (Malomane et al. 2018), we created a pruned SNP set using the 'snpgds' function in the R package SNPRelate (Zheng et al. 2012) using an LD threshold of $r^2 < 0.2$ and a sliding window of 500bp, which was applied to analyses that rely on allele frequencies for population inferences. We hereafter refer to the 'full' and 'pruned' SNP sets.

Excessive numbers of related individuals can bias some population genomics analyses ((O'Connell et al. 2019); (Wang 2018)). Because our sample collection was not always random, we anticipated that several individuals within our sample set were related, and therefore estimated pairwise relatedness within sampling regions in the pruned SNP set using the KING-robust method ((Manichaikul et al. 2010)) in SNPRelate. We then removed 1 from each pair with a kinship estimate > 0.35, which corresponds to a first-order familial relationship (sibling pair, parent-offspring). This sample set was used for all subsequent analyses.

We investigated phylogenetic distance among samples and groups using the full SNP set by inferring a phylogeny in RAxML version 8.2.12 ((Stamatakis 2014)) using the ASC-GTRCAT model with Lewis correction for ascertainment bias with 100 bootstrap replicates. We plotted the tree in the R 4.3 (R Core Team 2022) package GGTREE ((Yu et al. 2017)) with a customized script formulated for SNP marker data ((Severn-Ellis et al. 2020)).

To explore population structure, we used the pruned SNP set to visualize population clustering via PCA and investigated genetic differentiation by calculating pairwise F_{ST} ((Weir and Cockerham 1984)) between sites with correction for sample size, both in SNPRelate. We compared genetic diversity among sampling regions by calculating observed (H_{obs}) and expected (H_{exp}) heterozygosity, allelic richness rarified by sample size (A_R), and the inbreeding coefficient F_{IS} in Hierfstat v 0.5-11 ((Goudet and Goudet 2014)).

We estimated the most probable number of genetic clusters K within our sample set using the full SNP set. First, we implemented a discriminant analysis of principal components (DAPC) using the 'find.clusters' function in the R package *adegenet* ((Jombart and Bateman 2008)) to estimate optimal number of K testing K = 1–8. The most probable number of clusters was identified as the one with the lowest Bayesian Information Criterion (BIC) value. To minimize overfitting, we first determined the optimal number of principal components (PCs) using cross-validation and then ran DAPC retaining the optimal number of PCs and using the identified optimal number of genetic clusters as K to predict group assignment for each sample without *a priori* population

information. Ten principal components (PCs) were retained to explain approximately 52% of the total retained variation without our dataset.

We also used STRUCTURE v2.3.4 ((Pritchard 2000)) within the program *Structure_threader* ((Pina-Martins et al. 2017)) to estimate the value of K. We ran STRUCTURE with the admixture model and no prior population information, with an optimal burn-in of 10^4 steps and 10^4 additional steps, and for values of K from 1 to 8 with 10 replicates per iteration of K. We used the R package pophelper ((Francis 2017)) to align runs, plot output, and determine the most likely value of K via both ΔK ((Evanno et al. 2005)) and mean probability (MeanLn*P(K)*). STRUCTURE is known to be sensitive to differences in sample size in populations ((Puechmaille 2016)), which was a factor in our dataset. To eliminate the potential effects of sampling bias on our results, we also ran STRUCTURE using the same conditions for a random subset of n = 5–6 individuals per population, omitting the Russian and Italian samples from the dataset, for values of K from 1 to 5.

Competing interests

The authors have no competing interests.

Acknowledgment

This work was supported by the LOEWE-Centre for Translational Biodiversity Genomics (TBG) funded by the Hessen State Ministry of Higher Education, Research and the Arts (HMWK). Funding was also provided by the German Federal Agency for Nature Conservation (BfN) with resources from the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV), as part of a German-wide project led by the Bund für Umwelt und Naturschutz BUND (Friends of the Earth Germany), Justus Liebig University Gießen, and the Senckenberg Gesellschaft für Naturforschung in the Federal Programme for Biological Diversity called ‘In Search of the Garden Dormouse’.

We are very grateful to many people who have helped with sample collection: Peter Adamik, Hermann Ansorge, Jonas Astrin, Simon Capt, Stefanie Jessolat, Jeroen van der Kooij, Ines Leonhard, Gernot Segelbacher, David Stille, Adria Viñals Domingo, Frank Zachos, Alain Frantz, Olga Grigoryeva, Maurice La Haye, Roel Baets, Kasper Van Acker, Caren Raditz, Lutz Nielen, Stefanie Erhardt, Jürg Paul Müller, Andrea Krug, Danilo Hartung, Christine Thiel-Bender, Ferry Böhme, Julia Hofmann, Sarah Beer, Jörg und Elli Hacker, Otfried Wüstemann, Sonja Klein, Thomas Doebel, Martina Klotz, Hartmut Schmitt, the Schwarzwald national park, the Museo di Scienze Naturali dell’Alto Adige

and Museo di Storia Naturale La Specola di Firenze, Antoni Arrizabalaga, Natural Sciences Museum of Granollers, as well as numerous students and citizen scientists without whom this work would have not been possible. We thank Sarah Stubbe at JLU Gießen for sample preparation.

Data and Code Availability

Genomic and transcriptomics read data and our annotated genome is in the process of being uploaded to NCBI. The haplotype assemblies, gene and transposable element annotations are available for download at http://genome.senckenberg.de/download/GardenDormouse/. TOGA annotations are available at https://genome.senckenberg.de/download/TOGA/. We also provide a genome browser showing our genomes and all annotations at https://genome.senckenberg.de/.

Most of our analyses relied on publicly available tools and methods. New scripts are available at https://github.com/pabyerly/TBG_GardenDormouse.

References

Anděra M. 1986. Dormice (Gliridae) in Czechoslovakia. Part I. : Glis glis, Eliomys quercinus (Rodentia: Mammalia). *Folia Musei Rerum Naturalium Bohemiae Occidentalis, Plzeň - Zoologica* **24**: 3–47.

Arvidsson S, Fartmann B, Winkler S, Zimmermann W. 2016. Technical note: Efficient high-throughput SNP discovery and genotyping using normalised genotyping by sequencing (nGBS). https://biosearch-cdn.azureedge.net/assetsv6/efficient-high-throughput-snp-discovery-genotyping-ngbs-app-note.pdf (Accessed January 6, 2024).

Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Mozes E, Strope PK, Sylla PM, Wagner L, et al. 2023. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *bioRxiv*. http://dx.doi.org/10.1101/2023.06.02.543519.

Bennett D, Richard FJ. 2021. Distribution modelling of the garden dormouse Eliomys quercinus (Linnaeus, 1766) with novel climate change indicators. *Mamm Biol* **101**: 589–599.

Bertolino S. 2017. Distribution and status of the declining garden dormouse Eliomys quercinus. *Mammal Review* **47**: 133–147. http://dx.doi.org/10.1111/mam.12087.

Bertolino S, Colangelo P, Mori E, Capizzi D. 2015. Good for management, not for conservation: an overview of research, conservation and management of Italian small mammals. *Hystrix*. http://www.italian-journal-of-mammalogy.it/pdf-77188-13341?filename=Good for management_not.pdf.

Bertolino S, Viano C, Currado I. 2001. Population dynamics, breeding patterns and spatial use of

779 the garden dormouse (Eliomys quercinus) in an Alpine habitat. *J Zool* **253**: 513–521.

780 Blumer M, Brown T, Freitas MB, Destro AL, Oliveira JA, Morales AE, Schell T, Greve C, Pippel M,
781 Jebb D, et al. 2022. Gene losses in the common vampire bat illuminate molecular adaptations to
782 blood feeding. *Sci Adv* **8**: eabm6494.

783 Böhne A, Thiel-Bender C, Kukowka S, Darwin Tree of Life Barcoding collective, Wellcome Sanger
784 Institute Tree of Life programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines
785 collective, Tree of Life Core Informatics collective, Darwin Tree of Life Consortium. 2023. The
786 genome sequence of the hazel dormouse, Muscardinus avellanarius (Linnaeus, 1758). *Wellcome*
787 *Open Research* **8**. https://wellcomeopenresearch.org/articles/8-514 (Accessed January 30, 2024).

788 Brandies P, Peel E, Hogg CJ, Belov K. 2019. The Value of Reference Genomes in the Conservation
789 of Threatened Species. *Genes* **10**. https://pubmed.ncbi.nlm.nih.gov/31717707/.

790 Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S,
791 Garamszegi LZ, et al. 2015. Linked selection and recombination rate variation drive the evolution of
792 the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers.
793 *Genome Res* **25**: 1656–1665.

794 Campana MG, Parker LD, Hawkins MTR, Young HS, Helgen KM, Szykman Gunther M, Woodroffe
795 R, Maldonado JE, Fleischer RC. 2016. Genome sequence, population history, and pelage genetics
796 of the endangered African wild dog (Lycaon pictus). *BMC Genomics* **17**: 1013.

797 Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly
798 using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170–175.

799 Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication
800 using fastp. *Imeta* **2**. https://onlinelibrary.wiley.com/doi/10.1002/imt2.107.

801 Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor.
802 *Bioinformatics* **34**: i884–i890.

803 Combe FJ, Juškaitis R, Trout RC, Bird S, Ellis JS, Norrey J, Al-Fulaij N, White I, Harris WE. 2023.
804 Density and climate effects on age-specific survival and population growth: consequences for
805 hibernating mammals. *Anim Conserv* **26**: 317–330.

806 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G,
807 Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–
808 2158.

809 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T,
810 McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**:
811 giab008.

812 DeSaix MG, Bulluck LP, Eckert AJ, Viverette CB, Boves TJ, Reese JA, Tonra CM, Dyer RJ. 2019.
813 Population assignment reveals low migratory connectivity in a weakly structured songbird. *Mol Ecol*
814 **28**: 2122–2135.

815 du Plessis SJ, Blaxter M, Koepfli K-P, Chadwick EA, Hailer F. 2023. Genomics reveals complex
816 population history and unexpected diversity of Eurasian otters (Lutra lutra) in Britain relative to
817 genetic methods. *Mol Biol Evol* **40**. https://dx.doi.org/10.1093/molbev/msad207.

818 Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the
819 software STRUCTURE: A simulation study. *Mol Ecol* **14**: 2611–2620.

Formenti G, Theissinger K, Fernandes C, Bista I, Bombarely A, Bleidorn C, Ciofi C, Crottini A, Godoy JA, Höglund J, et al. 2022. The era of reference genomes in conservation genomics. *Trends in Ecology and Evolution* **37**: 197–202. https://www.

Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour* **17**: 27–32.

Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, Sisu C, Wright JC, Armstrong J, Barnes I, et al. 2021. GENCODE 2021. *Nucleic Acids Res* **49**: D916–D923.

Gautier M, Moazami-Goudarzi K, Levéziel H, Parinello H, Grohs C, Rialle S, Kowalczyk R, Flori L. 2016. Deciphering the Wisent Demographic and Adaptive Histories from Individual Whole-Genome Sequences. *Mol Biol Evol* **33**: 2801–2814.

Gippoliti S, Amori G. 2007. Beyond threatened species and reintroduction: establishing priorities for conservation and breeding programmes for European rodents in zoos. *International Zoo Yearbook* **41**: 194–202.

Giroud S, Ragger MT, Baille A, Hoelzl F, Smith S, Nowack J, Ruf T. 2023. Food availability positively affects the survival and somatic maintenance of hibernating garden dormice (Eliomys quercinus). *Front Zool* **20**: 1–11.

Goudet AJ, Goudet MJ. 2014. Package " hierfstat ."

Halley DJ, Saveljev AP, Rosell F. 2021. Population and distribution of beavers Castor fiber and Castor canadensis in Eurasia. *Mamm Rev* **51**: 1–24.

Harris RS. 2007. Improved pairwise alignment of genomic DNA. The Pennsylvania State University, Ann Arbor, United States https://www.proquest.com/dissertations-theses/improved-pairwise-alignment-genomic-dna/docview/304835295/se-2.

Howe K, Chow W, Collins J, Pelan S, Pointon D-L, Sims Y, Torrance J, Tracey A, Wood J. 2021. Significantly improving the quality of genome assemblies through curation. *Gigascience* **10**. http://dx.doi.org/10.1093/gigascience/giaa153.

Huang N, Li H. 2023. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**. http://dx.doi.org/10.1093/bioinformatics/btad595.

Jombart T, Bateman A. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405.

Kennerley RJ, Lacher TE, Hudson MA, Long B, McCay SD, Roach NS, Turvey ST, Young RP. 2021. Global patterns of extinction risk and conservation needs for Rodentia and Eulipotyphla. *Diversity and Distributions* **27**: 1792–1806.

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. 2003. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A* **100**: 11484–11489.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915.

Kirilenko BM, Munegowda C, Osipova E, Jebb D, Sharma V, Blumer M, Morales AE, Ahmed A-W, Kontopoulos D-G, Hilgers L, et al. 2023. Integrating gene annotation with orthology inference at scale. *Science* **380**: eabn3107.

Lane JE, Kruuk LEB, Charmantier A, Murie JO, Dobson FS. 2012. Delayed phenology and reduced fitness associated with climate change in a wild hibernator. *Nature* **489**: 554–557.

Lang J, Büchner S, Meinig H, Bertolino S. 2022. Do We Look for the Right Ones? An Overview of Research Priorities and Conservation Status of Dormice (Gliridae) in Central Europe. *Sustainability (Switzerland)* **14**: 9327. https://www.mdpi.com/2071-1050/14/15/9327/htm.

Leyhausen J, Cocchiararo B, Nowak C, Ansorge H, Bertolino S, Büchner S, Fietz J, Foppen R, Juškaitis R, La Haye M, et al. 2022. Genotyping-by-sequencing based SNP discovery in a non-model rodent, the endangered hazel dormouse. *Conserv Genet Resour* **14**: 195–201.

Libois R, Ramalhinho MG, Rosoux R. 2012. Evidence for a differentiated chromosomal race north of classical south European refuge areas in the garden dormouse Eliomys quercinus. *Acta Theriol* **57**: 313–320.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bioGN]*. http://arxiv.org/abs/1303.3997.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496.

Luo H, Jiang X, Li B, Wu J, Shen J, Xu Z, Zhou X, Hou M, Huang Z, Ou X, et al. 2023. A high-quality genome assembly highlights the evolutionary history of the great bustard (Otis tarda, Otidiformes). *Communications Biology* **6**: 1–11.

Lu X, Costeur L, Hugueney M, Maridet O. 2021. New data on early Oligocene dormice (Rodentia, Gliridae) from southern Europe: phylogeny and diversification of the family. *J Syst Palaeontol* **19**: 169–189.

Mahlert B, Gerritsmann H, Stalder G, Ruf T, Zahariev A, Blanc S, Giroud S. 2018. Implications of being born late in the active season for growth, fattening, torpor use, winter survival and fecundity. *Elife* **7**. http://dx.doi.org/10.7554/eLife.31225.

Malomane DK, Reimer C, Weigend S, Weigend A, Sharifi AR, Simianer H. 2018. Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies. *BMC Genomics* **19**: 22.

Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**: 2867–2873.

Martchenko D, Shafer ABA. 2023. Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. *Heredity* **131**: 273–281.

Mead D, Fingland K, Cripps R, Portela Miguez R, Smith M, Corton C, Oliver K, Skelton J, Betteridge E, Doulcan J, et al. 2020. The genome sequence of the eastern grey squirrel, Sciurus carolinensis Gmelin, 1788. *Wellcome Open Res* **5**: 27.

Meinig H, Büchner S. 2012. The current situation of the garden dormouse (Eliomys quercinus) in Germany. *Peckiana* **8**: 129–134.

900 Meinig H, Büchner S, Lang J, von Thaden A, Reiners TE, Nowak C, Nava TF, Parcsettich EM,
901 Brünner H, Andersen A, et al. 2023. „Spurensuche Gartenschläfer“–ein Citizen-Science-Projekt zum
902 Schutz einer gefährdeten Schlafmaus in Deutschland ” In search of the garden dormouse " - A
903 citizen science project to protect an endangered dormouse species in Germany. *Natur and*
904 *Landschaft* **8**: 382–390.

905 Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. 2017. Differences between germline and
906 somatic mutation rates in humans and mice. *Nat Commun* **8**: 15183.

907 Moritz C. 1994. Defining “ Evolutionarily Significant Units ” for conservation. *Trees* **9**: 373–375.

908 Mouton A, Mortelliti A, Grill A, Sara M, Kryštufek B, Juškaitis R, Latinne A, Amori G, Randi E,
909 Büchner S, et al. 2017. Evolutionary history and species delimitations: a case study of the hazel
910 dormouse, Muscardinus avellanarius. *Conserv Genet* **18**: 181–196.

911 Nadachoswki A, Daoud A. 1995. Patterns of myoxid evolution in the Pliocene and Pleistocene of
912 Europe. *Hystrix-italian Journal of Mammalogy* **6**.
913 https://doc.rero.ch/record/15906/files/PAL_E3309.pdf (Accessed February 14, 2024).

914 Nieto-Blázquez ME, Schreiber D, Mueller SA, Koch K, Nowak C, Pfenninger M. 2022. Human
915 impact on the recent population history of the elusive European wildcat inferred from whole genome
916 data. *BMC Genomics* **23**: 1–14.

917 O’Connell KA, Mulder KP, Maldonado J, Currie KL, Ferraro DM. 2019. Sampling related individuals
918 within ponds biases estimates of population structure in a pond-breeding amphibian. *Ecol Evol* **9**:
919 3620–3636.

920 Osipova E, Hecker N, Hiller M. 2019. RepeatFiller newly identifies megabases of aligning repetitive
921 sequences and improves annotations of conserved non-exonic elements. *Gigascience* **8**.
922 http://dx.doi.org/10.1093/gigascience/giz132.

923 Paez S, Kraus RHS, Shapiro B, Gilbert MTP, Jarvis ED. 2022. Reference genomes for
924 conservation. *Science* **377**: 364–366.

925 Pérez-Espona S. 2017. Conservation genetics in the European Union – Biases, gaps and future
926 directions. *Biol Conserv* **209**: 130–136.

927 Perez GCL, Libois R, Nieberding CM. 2013. Phylogeography of the garden dormouse Eliomys
928 quercinus in the western Palearctic region. *J Mammal* **94**: 202–217.

929 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: An
930 inexpensive method for de novo SNP discovery and genotyping in model and non-model species.
931 *PLoS One* **7**. http://dx.doi.org/10.1371/journal.pone.0037135.

932 Pina-Martins F, Silva DN, Fino J, Paulo OS. 2017. Structure_threader: An improved method for
933 automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU
934 systems. *Mol Ecol Resour* **17**: e268–e274.

935 Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J,
936 Nguyen N, Afshar PT, et al. 2018. A universal SNP and small-indel variant caller using deep neural
937 networks. *Nat Biotechnol* **36**: 983–987.

938 Pritchard JK. 2000. Inference of population structure using multi-locus genotypes. *Genetics* **155**:
939 945–959.

940 Puechmaille SJ. 2016. The program structure does not reliably recover the correct population
941 structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Mol*
942 *Ecol Resour* **16**: 608–627.

943 Rammou D-L, Astaras C, Migli D, Boutsis G, Galanaki A, Kominos T, Youlatos D. 2022. European
944 ground squirrels at the edge: Current distribution status and anticipated impact of climate on
945 Europe's southernmost population. *Land (Basel)* **11**: 301.

946 Reiners TE, Eidenschenk J, Neumann K, Nowak C. 2014. Preservation of genetic diversity in a wild
947 and captive population of a rapidly declining mammal, the Common hamster of the French Alsace
948 region. *Mamm Biol* **79**: 240–246.

949 Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W,
950 Fungtammasan A, Kim J, et al. 2021. Towards complete and error-free genome assemblies of all
951 vertebrate species. *Nature* **592**: 737–746.

952 Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness,
953 and phasing assessment for genome assemblies. *Genome Biol* **21**: 245.

954 Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: Analytical methods for paired-end
955 sequencing improve RADseq-based population genomics. *Mol Ecol* **28**: 4737–4754.

956 Sambrook J, Russell WD. Protocol: DNA isolation from mammalian tissue. *Molecular Cloning: A*
957 *Laboratory Manual (Sambrook J*.

958 Schaub M, Vaterlaus-Schlegel C. 2001. Annual and seasonal variation of survival rates in the
959 garden dormouse (Eliomys quercinus). *J Zool* **255**: 89–96.

960 Severn-Ellis AA, Scheben A, Neik TX, Saad NSM, Pradhan A, Batley J. 2020. Genotyping for
961 species identification and diversity assessment using double-digest restriction site-associated DNA
962 sequencing (ddRAD-Seq). In *Methods in Molecular Biology*, Vol. 2107 of, pp. 159–187, Humana
963 Press Inc.

964 Sharma V, Hiller M. 2017. Increased alignment sensitivity improves the usage of genome
965 alignments for comparative gene annotation. *Nucleic Acids Res* **45**: 8369–8377.

966 Shumate A, Wong B, Pertea G, Pertea M. 2022. Improved transcriptome assembly using a hybrid of
967 long and short reads with StringTie. *PLoS Comput Biol* **18**: e1009730.

968 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
969 phylogenies. *Bioinformatics* **30**: 1312–1313.

970 Suarez HG, Langer BE, Ladde P, Hiller M. 2017. chainCleaner improves genome alignment
971 specificity and sensitivity. *Bioinformatics* **33**: 1596–1603.

972 Szarmach SJ, Brelsford A, Witt CC, Toews DPL. 2021. Comparing divergence landscapes from
973 reduced-representation and whole genome resequencing in the yellow-rumped warbler (Setophaga
974 coronata) species complex. *Mol Ecol* **30**: 5994–6005.

975 Takach BV, Sargent H, Penton CE, Rick K, Murphy BP, Neave G. 2023. Population genomics and
976 conservation management of the threatened black-footed tree-rat (Mesembriomys gouldii) in
977 northern Australia. 1–11.

978 Talla V, Mrazek V, Höglund J, Backström N. 2023. Whole genome re-sequencing uncovers
979 significant population structure and low genetic diversity in the endangered clouded apollo

(Parnasssius mnemosyne) in Sweden. *Conserv Genet*. https://doi.org/10.1007/s10592-023-01502-9.

Theissinger K, Fernandes C, Formenti G, Bista I, Berg PR, Bleidorn C, Bombarely A, Crottini A, Gallo GR, Godoy JA, et al. 2023. How genomics can help biodiversity conservation. *Trends in Genetics* **39**: 545–559. https://doi.org/10.1016/j.tig.2023.01.005.

Tsuchiya MTN, Dikow RB, Cassin-Sackett L. 2020. First Genome Sequence of the Gunnison's Prairie Dog (Cynomys gunnisoni), a Keystone Species and Player in the Transmission of Sylvatic Plague. *Genome Biol Evol* **12**: 618–625.

Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S, Nishino J, Yagi T. 2015. Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome Res* **25**: 1125–1134.

Verde Arregoitia LD. 2016. Biases, gaps, and opportunities in mammalian extinction risk research. *Mamm Rev* **46**: 17–29.

Viluma A, Flagstad Ø, Åkesson M, Wikenros C, Sand H, Wabakken P, Ellegren H. 2022. Whole-genome resequencing of temporally stratified samples reveals substantial loss of haplotype diversity in the highly inbred Scandinavian Wolf population. *Genome Res* **32**: 449–458.

Wang J. 2018. Effects of sampling close relatives on some elementary population genetics analyses. *Mol Ecol Resour* **18**: 41–54.

Weir BS, Cockerham CC. 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**: 1358.

Wolf JF, Bowman J, Keobouasone S, Taylor RS, Wilson PJ. 2022. A de novo genome assembly and annotation of the southern flying squirrel (Glaucomys volans). *G3* **12**. http://dx.doi.org/10.1093/g3journal/jkab373.

Wright BR, Farquharson KA, McLennan EA, Belov K, Hogg CJ, Grueber CE. 2020. A demonstration of conservation genomics for threatened species management. *Mol Ecol Resour* **20**: 1526–1541.

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* **8**: 28–36.

Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, et al. 2021. Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat Commun* **12**: 6566.

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326–3328.

Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**. http://dx.doi.org/10.1093/bioinformatics/btac808.

Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**: 240–245.