

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

Paige A. Byerly,<sup>1,2,16</sup> Alina von Thaden,<sup>1,2,16</sup> Evgeny Leushkin,<sup>1,3</sup> Leon Hilgers,<sup>1,3</sup> Shenglin Liu,<sup>1,3</sup> Sven Winter,<sup>4,5</sup> Tilman Schell,<sup>1,3</sup> Charlotte Gerheim,<sup>1,3</sup> Alexander Ben Hamadou,<sup>1,3</sup> Carola Greve,<sup>1,3</sup> Christian Betz,<sup>6</sup> Hanno J. Bolz,<sup>6</sup> Sven Büchner,<sup>7</sup> Johannes Lang,<sup>7</sup> Holger Meinig,<sup>7</sup> Evax Marie Famira-Parcsetich,<sup>7</sup> Sarah P. Stubbe,<sup>7</sup> Alice Mouton,<sup>8</sup> Sandro Bertolino,<sup>9</sup> Goedele Verbeylen,<sup>10</sup> Thomas Briner,<sup>11</sup> Lúdia Freixas,<sup>12</sup> Lorenzo Vinciguerra,<sup>13</sup> Sarah A. Mueller,<sup>14</sup> Carsten Nowak,<sup>1,2</sup> and Michael Hiller<sup>1,3,15</sup>

<sup>1</sup>LOEWE Centre for Translational Biodiversity Genomics, 60325 Frankfurt, Germany; <sup>2</sup>Conservation Genetics Group, Senckenberg Research Institute and Natural History Museum Frankfurt, 63571 Gelnhausen, Germany; <sup>3</sup>Senckenberg Research Institute, 60325 Frankfurt, Germany; <sup>4</sup>Senckenberg Biodiversity and Climate Research Centre, 60325 Frankfurt am Main, Germany; <sup>5</sup>Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna, 1160 Vienna, Austria; <sup>6</sup>Bioscientia Human Genetics, Institute for Medical Diagnostics GmbH, 55218 Ingelheim, Germany; <sup>7</sup>Justus-Liebig-University Giessen, Clinic for Birds, Reptiles, Amphibians and Fish, Working Group for Wildlife Research, 35392 Giessen, Germany; <sup>8</sup>Socio-économie, Environnement et Développement (SEED), University of Liege (Arlon Campus Environment), 81001 Arlon, Belgium; <sup>9</sup>Department of Life Sciences and Systems Biology, University of Turin, 10123 Torino, Italy; <sup>10</sup>Natuurpunt Studie vzw, Mammal Working Group, 2800 Mechelen, Belgium; <sup>11</sup>Naturmuseum Solothurn, 4500 Solothurn, Switzerland; <sup>12</sup>BiBio Research Group, Natural Sciences Museum of Granollers, 08402 Granollers, Catalonia, Spain; <sup>13</sup>Naturmuseum St. Gallen, 9016 St. Gallen, Switzerland; <sup>14</sup>Division of Evolutionary Biology, Faculty of Biology, LMU Munich, 80539 Munich, Germany; <sup>15</sup>Institute of Cell Biology and Neuroscience, Faculty of Biosciences, Goethe University Frankfurt, 60438 Frankfurt, Germany

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 data sets with homology-based methods to generate a highly complete gene annotation. Demographic history analysis of the genome reveal a sharp population decline since the last interglacial, indicating an association between colder climates and population declines before anthropogenic influence. Using our genome and genetic data from 100 individuals, largely sampled in a citizen-science project across the contemporary range, we conduct the first population genomic analysis for this species. We find clear evidence for population structure across the species' core Central European range. Notably, our data show that the Alpine population, characterized by strong differentiation and reduced genetic diversity, is reproductively isolated from other regions and likely represents a differentiated evolutionary significant unit (ESU). The predominantly declining Eastern European populations also 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 the designation of ESUs.

[Supplemental material is available for this article.]

As genome sequencing technology advances and costs decrease, high-quality reference genomes are becoming increasingly recognized as an essential resource for conservation genomics (Brandies et al. 2019; Formenti et al. 2022; Paez et al. 2022; Theissinger et al. 2023). When aligned to a reference genome assembly, informative markers such as genome-wide single-nucleo-

tide polymorphisms (SNPs) generated via high-throughput genomic data can contribute to the understanding of past and contemporary population demographics (Campana et al. 2016; Gautier et al. 2016; du Plessis et al. 2023; Luo et al. 2023), reveal evolutionary patterns such as adaptive differentiation (Szarmach et al. 2021; Martchenko and Shafer 2023), and inform the current status of wildlife species of conservation concern (Viluma et al.

<sup>16</sup>These authors contributed equally to this work.

Corresponding authors: [paige.byerly@senckenberg.de](mailto:paige.byerly@senckenberg.de), [michael.hiller@senckenberg.de](mailto:michael.hiller@senckenberg.de)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.279066.124>.

© 2024 Byerly et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

2022; Talla et al. 2023). While reduced-representation sequencing (RRS) methodologies continue to be the more cost-effective and efficient means of generating SNP data sets for large numbers of samples (Peterson et al. 2012; Wright et al. 2020), 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 (Brandies et al. 2019; Rochette et al. 2019). Reference genomes are also useful for improving data recovery from closely related species, with cross-species alignment shown to be highly effective (Burri et al. 2015; DeSaix et al. 2019; Nieto-Blázquez et al. 2022; Takach et al. 2023). Until recently, reference genomes for nonmodel organisms were often generated using short-read sequencing. However, short-read reference genomes are prone to errors such as sequence gaps, false duplicates, and misassembled reads, all of which can lead to inaccuracies in downstream data processing (Rhie et al. 2021). In contrast, genome assemblies generated from long-read sequencing are largely free from these sources of error. This higher accuracy combined with greater contiguity makes long-read sequencing more suitable for genomic analyses such as linking phenotype to genotype, detection of large structural variants, and gene expression analyses (Jebb et al. 2020; Rhie et al. 2021; Wold et al. 2021; Karimi et al. 2022). Given the value of high-quality reference genome assemblies for population and conservation genomic analyses, long-read sequencing of reference genomes for wildlife groups should be a focus of ongoing conservation efforts.

Highly informative data sources such as reference genomes derived from long-read sequencing can have particular value for investigating understudied wildlife groups, as genomic data can contribute to baseline knowledge needed for conservation planning such as population size estimates and historical changes in demography. Within mammals, such foundational knowledge is lacking for many small-bodied taxa such as Rodentia and Eulipotyphla, which have a known deficit of research (Verde Arregoitia 2016; Kennerley et al. 2021) 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 (Gippoliti and Amori 2007; Bertolino et al. 2015; Pérez-Espona 2017; Lang et al. 2022), though many species appear to be declining (Gippoliti and Amori 2007; Reiners et al. 2014; Lang et al. 2022; Rammou et al. 2022). Population genomic analysis could expand our understanding of population structure and the factors contributing to declines in Europe, 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 ~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 before 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; Büchner et al. 2024), 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; Büchner et al. 2024), 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 (Libois et al. 2012; Perez et al. 2013), 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 long-read-based 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

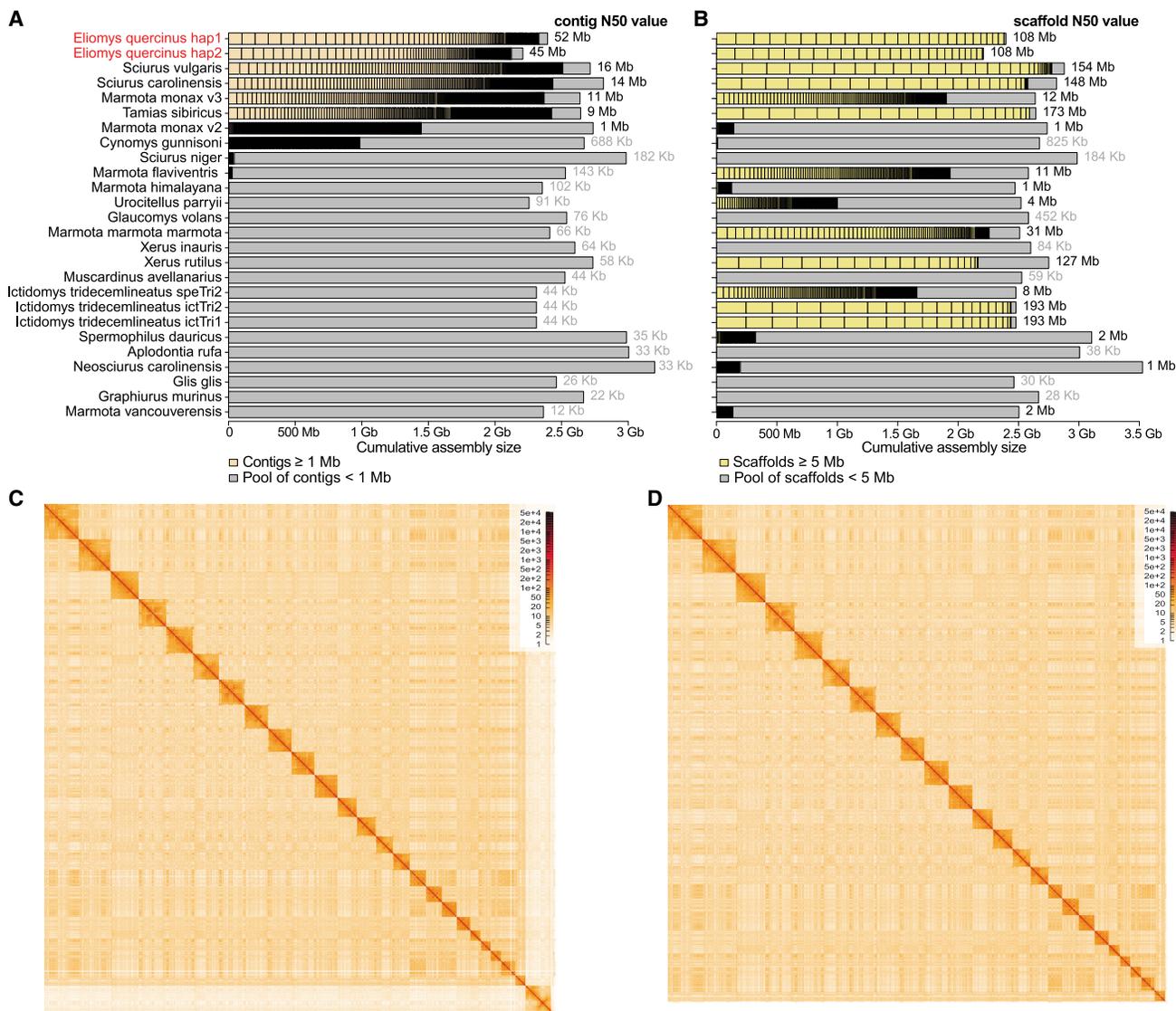
### Haplotype-resolved genome assembly

To generate a reference-quality assembly for the garden dormouse, we collected tissue from a male individual sourced from Mainz, Germany in 2023. We used DNA from this individual to generate 124.4 Gb of Pacific Biosciences (PacBio) high-fidelity (HiFi) long-read data, with 117 Gbp in long-range Hi-C read pairs. Assuming a 2.5 Gb genome size, similar to other dormice (*Zoosonia* Consortium 2020; Böhne et al. 2023), the HiFi and Hi-C data have an estimated coverage of ~50× and ~47×, respectively. With these data, we obtained two haplotype-resolved genome assemblies (Supplemental 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; Supplemental 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. The Y Chromosome could not be assembled well. We identified 24 autosomes, implying that the sequenced individual has a karyotype of 50 chromosomes ( $2 \times 24 + 2$  sex chromosomes) (Fig. 1C,D), which matches the range of 48–54 chromosomes identified for *Eliomys* so far and corresponds to the expected karyotype of the Northern clade where the specimen was sampled (Perez et al. 2013). Overall, 99.4% and 99.6% of haplotypes 1 and 2 are contained in chromosome-level scaffolds. These metrics exceed the standards set by the Vertebrate Genome Project (Rhie et al. 2021).

### Assembly quality and completeness

Using the HiFi reads, we estimate that our haplotypes 1 and 2 assemblies have a high base accuracy with a consensus quality value (QV) of 60.46 and 60.24, indicating only one error per megabase.



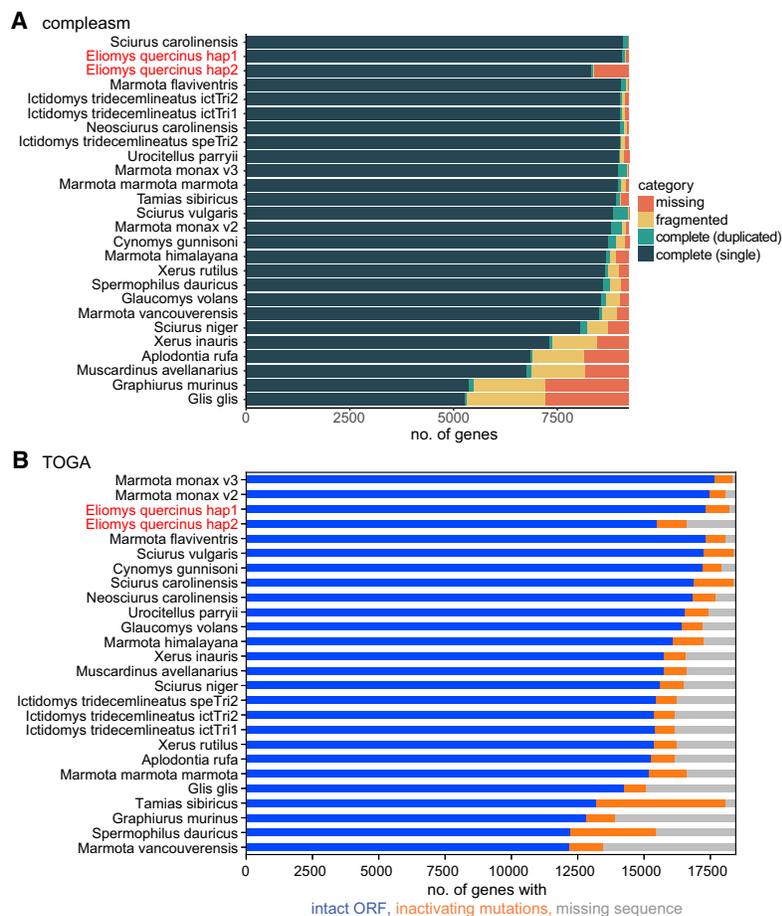
**Figure 1.** Contiguity and Hi-C maps of both haplotype assemblies. (A,B) Visualization of contig (A) and scaffold (B) sizes of the garden dormouse haplotypes 1 and 2 (red font) and other Sciuromorpha genome assemblies. The N50 values are given on the right side. Contigs shorter than 1 Mb and scaffolds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

To assess gene completeness, we first used compleasm (Huang and Li 2023) with the BUSCO odb10 set of 9226 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 ranks second after the *Sciurus carolinensis* assembly (Fig. 2A; Supplemental Table S1; Mead et al. 2020). 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 miss-

ing 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; Supplemental Table S1). In comparison to other assemblies in the order Sciuromorpha, our garden dormouse haplotype 1 ranks third in terms of intact genes. Despite being top-ranked in the compleasm metric, TOGA reveals that the *S. carolinensis* assembly has substantially more genes with inactivating mutations, indicating a lower base accuracy (Fig. 2B).

### Gene annotation

To comprehensively annotate genes of the garden dormouse genome, we generated transcriptomics data from eight different organs (heart, testis, liver, lung, spleen, muscle, kidney, and bladder). We combined ~4.2 billion short RNA-seq and ~1.6 million long Iso-Seq reads to produce transcript models. Transcript models



**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 data set, 9226 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 Supplemental Table S1.

were joined with a comprehensive homology-based annotation, obtained by first generating genome alignments and then applying TOGA with human and mouse as the reference. The resulting annotation comprised 19,374 coding genes and as estimated by compleasm in protein mode, completely contains 99.91% of the mammalian BUSCO genes, which indicates a very high completeness.

### Long-term demographic history of the garden dormouse

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

Our results indicate relatively high  $N_e$  (>130,000) and consistent population growth from 800 kya toward the last interglacial (Fig. 3). Population sizes reached their maximum of >200,000 individuals around the last interglacial. Subsequently,  $N_e$  gradually de-

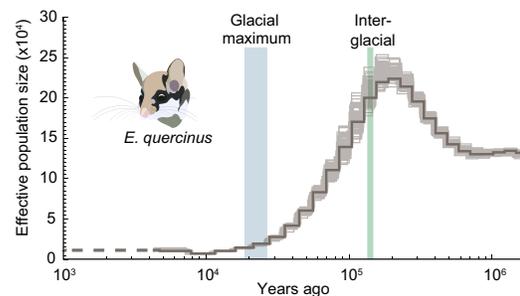
clined toward the last glacial maximum, reaching an estimated minimum  $N_e$  of less than 10,000 individuals around 10 kya. Together, this indicates that the garden dormouse thrived in warm periods, while a cooling climate with longer winters caused severe population declines. Consistent with this result, the Gliridae family thrived during earlier warm periods in the Miocene (23–5.3 Mya) (Nadachowski and Daoud 1995).

### Genotyping and variant calling for population genomics

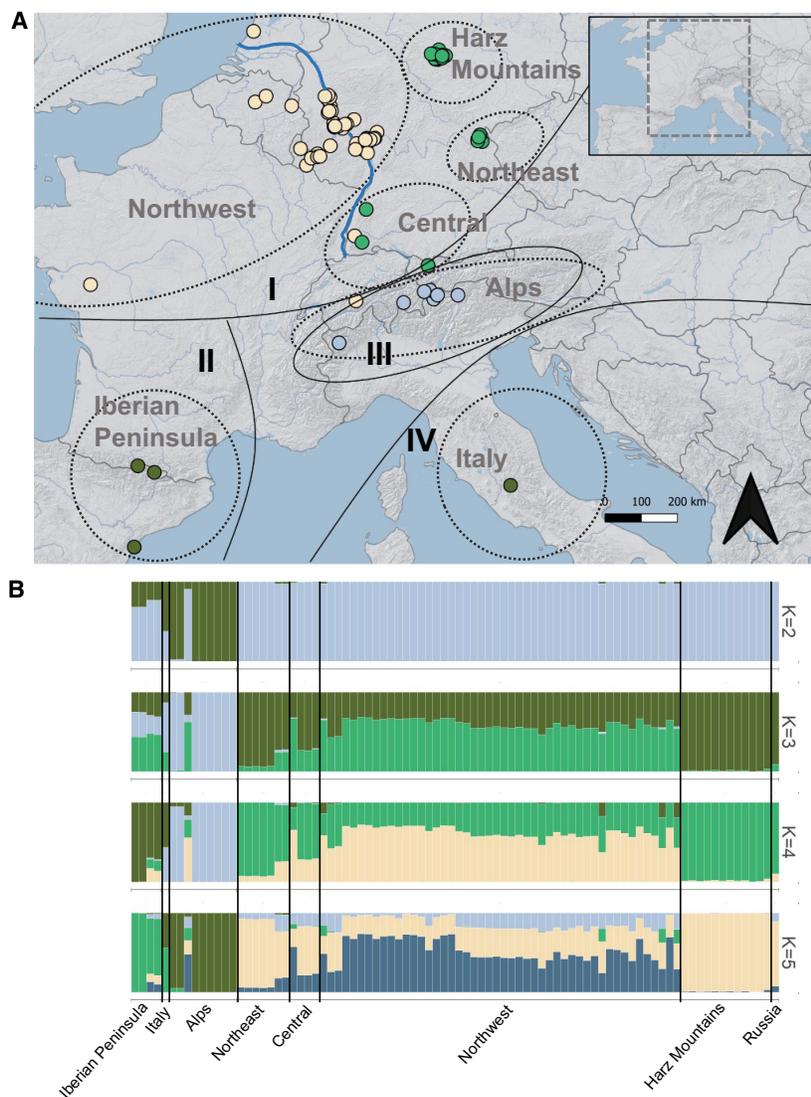
To investigate population structure across the European range of the garden dormouse, we obtained  $n = 100$  samples collected opportunistically between 1991 and 2020 for population genomic analyses and performed normalized genotyping-by-sequencing (nGBS). After sequencing, reads were aligned to the garden dormouse reference genome, and the resulting alignments were used to call 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 $\times$  and mean frequency of missingness of 0.11 (Supplemental Table S2). For analyses dependent on allele frequencies, such as principal component analysis (PCA)-based analyses, we pruned SNPs based on linkage disequilibrium (LD), resulting in a data set of 19,022 SNPs.

We first inferred relatedness among individual samples and found six pairwise groups of first-order relations in the Harz population. Because the inclusion of related individuals can bias population



**Figure 3.** Demographic history of the garden dormouse. Effective population size ( $N_e$ ) of *E. quercinus* as estimated with PSMC.  $N_e$  in 10,000 is shown on the y-axis and time in years ago on the x-axis. Faint lines indicate uncertainty of inferred  $N_e$  based on 100 bootstrap replicates. Blue and green bars show the timing of the last glacial maximum and last interglacial, respectively. The most recent  $N_e$  estimate with reduced accuracy is displayed by a dotted line on the left.



**Figure 4.** Garden dormouse population structure inferred from 47,115 nuclear SNP loci. (A) Group assignment via five PCs obtained by a discriminant analysis of principal components. The predicted assignment of genetic clusters ( $K$ ) 1–3 is shown by four different colors and 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 gray 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.

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 (Supplemental 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 (Supplemental 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 data set of  $n=86$  samples (Supplemental Table S2). Because we had only one sample for Italy and Russia, we omitted

these samples from further population genetic analyses, retaining these samples only for investigating population structure.

#### Assessing genetic diversity

We quantified genetic diversity as both SNP-based (polymorphisms only) and genome-wide (both fixed and variable sites) heterozygosity and found the highest genetic diversity for the Iberian Peninsula and Central regions (Table 1). Within the Central European region, genetic diversity was lower for population clusters east of the Rhine River (Northeast, Alps, Harz Mountains) than those located west of or near the river valley (Northwest, Central). This finding is consistent with the overall lower sizes of garden dormouse populations located east of the Rhine River in Central Europe (Bertolino 2017).

$F_{ST}$  was significantly high (confidence intervals did not overlap 0) between most regional pairwise groupings (Table 2), suggesting that dispersal between spatially disjunct populations may be limited. However, the highest overall  $F_{ST}$  values were consistently recovered for pairwise comparisons of the Alpine group and all other populations.

#### Assessing population structure

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 (Supplemental Fig. S2) except for a single Central individual (G190084BW), which clustered with the Northwest group. 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 14.48% of the variation on PC1 (Fig. 5). Further spatial differentiation

between the Northwest region and the other Central European sampling locations explained 8.19% of the variation on PC2. These results were consistent with findings of relatively high  $F_{ST}$  among regions.

We used discriminant analysis of principal components (DAPC) and STRUCTURE to infer the number of genetic clusters within our samples. DAPC identified  $K=4$  as the most probable number of genetic clusters (Supplemental Fig. S3A), while the results for STRUCTURE were less clear, suggesting  $K=2-4$  as possible cluster numbers (Supplemental Fig. S4). Cluster membership for DAPC was suggestive of relatively strong differentiation between the Northwest region versus other regions in Germany (Fig. 4A; Supplemental Fig. S3B,C), while STRUCTURE indicated greater

**Table 1.** Sample size ( $n$ ) and genetic diversity metrics among eight garden dormouse sampling regions with number of SNP-based polymorphic sites, observed ( $H_{obs}$ ) and expected ( $H_{exp}$ ) heterozygosity, allelic richness rarified by sample size ( $A_R$ ), the inbreeding coefficient  $F_{IS}$ , and nucleotide diversity  $\pi$  estimated from 19,022 SNP loci, and observed autosomal heterozygosity ( $H_{obs\_A}$ ) estimated from ~90,412 variant and ~58,406,409 fixed sites

| Region            | $n$ | Polymorphic sites | $H_{obs}$ | $H_{exp}$ | $F_{IS}$ | $\pi$ | $H_{obs\_A}$ |
|-------------------|-----|-------------------|-----------|-----------|----------|-------|--------------|
| Alps              | 9   | 5975              | 0.036     | 0.068     | 0.149    | 0.075 | 0.00028      |
| Central           | 4   | 13,552            | 0.257     | 0.273     | 0.123    | 0.321 | 0.00017      |
| Harz Mountains    | 12  | 12,992            | 0.223     | 0.233     | 0.053    | 0.245 | 0.00034      |
| Iberian Peninsula | 4   | 11,389            | 0.249     | 0.230     | 0.056    | 0.280 | 0.00067      |
| Northeast         | 7   | 13,173            | 0.209     | 0.244     | 0.124    | 0.264 | 0.00036      |
| Northwest         | 48  | 18,062            | 0.286     | 0.344     | 0.180    | 0.360 | 0.00042      |

Values for the Northwest region were generated by running four separate analyses of  $n=12$  and then averaging the output to avoid biases resulting from the uneven sample size between the Northwest ( $n=48$ ) and the other regions.

admixture among these regions at all values of  $K$  (Fig. 4B; Supplemental Fig. S5). Both methods were consistent in differentiating the Alpine samples from all other regions (Fig. 4B; Supplemental Fig. S5). Patterns observed for the full data set were comparable for the STRUCTURE analysis subset by sample size (Supplemental Fig. S6), indicating that our results were not influenced by unequal sample size among our sampling regions for regions where  $n > 1$ .

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 (Figs. 4, 5).

### Recent demographic history inferred from the site-frequency spectrum

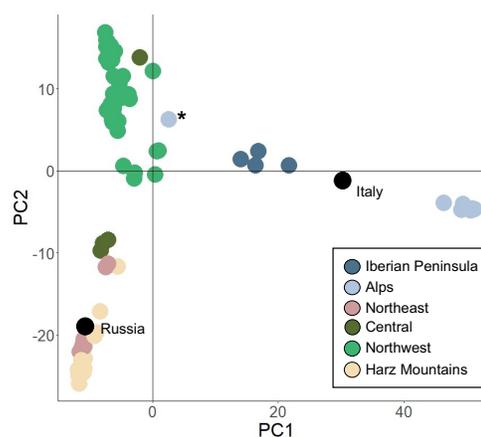
To infer recent demographic histories and  $N_e$  for each individual population, we used coalescent simulations and compared three models (Supplemental Fig. S7): constant steady population size (null model), population bottleneck (instantaneous size change model), and constant population growth or decline (incremental size change [ISC] model). For all populations, the instantaneous size change (bottleneck) model showed the strongest support based on the lowest Akaike information criterion (AIC) value, although likelihood differences between models were not great (Table 3). The timing of the bottlenecks was largely consistent among the Alps, the Harz, and the Northeast, indicating a bottleneck event ~3000 years ago. The Northwest model presented a slightly more distant bottleneck, at 7365 years ago, which aligns with the population minimum ~10,000 years ago inferred by PSMC. For the more northerly populations Harz, Northeast, and Northwest, differences between  $N_e$  before and during the bottleneck were not extensive (Table 3), indicating that the inferred population change may have been more gradual, which is in line with the steady decline suggested by PSMC. Despite this, all populations show contrastingly low contemporary  $N_e$ , with the lowest  $N_e$  inferred for the Harz and Northeast populations. In contrast, the top model indicated that the Alps population experienced a catastrophic decline during the bottleneck, with  $N_e$

reduced by 99.99% before recovering to over 2× the ancestral population size.

### Discussion

By generating a long-read-based high-quality reference genome for the garden dormouse, our study provides a genetic resource that can be used to aid conservation efforts of small mammals in Europe. 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 national conservation and research project for this species in Germany (Büchner et al. 2024). The project aims to restore garden dormouse populations, as the species continues to decline across much of its range (Lang et al. 2022). Using our long-read genome 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 long-term demographic history inferred with PSMC indicates that garden dormice underwent a substantial population



**Figure 5.** PCA for the garden dormouse. The PCA is based on 19,022 nuclear SNP loci, with colors representing the geographical origin. Results are plotted on the first two principal components, with PC1 explaining 14.48% and PC2 8.19% of variation within the data set. 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 an asterisk.

**Table 2.** Pairwise  $F_{ST}$  between six garden dormouse sampling regions below the diagonal line represented by dashes and 95% confidence intervals based on 1000 bootstraps above the diagonal, corrected for sample size

|      | NW    | ALP          | HARZ         | CE           | IB           | NE           |
|------|-------|--------------|--------------|--------------|--------------|--------------|
| NW   | —     | 0.336, 0.344 | 0.162, 0.168 | 0.063, 0.068 | 0.187, 0.195 | 0.133, 0.139 |
| ALP  | 0.340 | —            | 0.564, 0.576 | 0.556, 0.568 | 0.472, 0.488 | 0.560, 0.571 |
| HARZ | 0.165 | 0.570        | —            | 0.208, 0.219 | 0.377, 0.389 | 0.154, 0.163 |
| CE   | 0.065 | 0.562        | 0.213        | —            | 0.245, 0.259 | 0.083, 0.093 |
| IB   | 0.191 | 0.480        | 0.383        | 0.252        | —            | 0.331, 0.344 |
| NE   | 0.136 | 0.565        | 0.159        | 0.088        | 0.337        | —            |

Sampling regions are defined and abbreviated as Northwest (NW), the Alps (ALP), Harz Mountains (HAR), Central (CE), the Iberian Peninsula (IB), and Northeast (NE).

bottleneck before modern anthropogenic influence, with a sustained decline and long-term low  $N_e$  following the last interglacial. Our data suggest that the warm climate 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 toward the last glacial maximum, raising the possibility that the observed lineage diversification during glaciation may have been more likely due to 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).

More recent demographic history inferred with our SNP data set further supports a bottleneck history for the individual popula-

tions analyzed. While the populations of Harz, Northwest, and Northeast showed evidence for moderate reductions of their ancestral population sizes during the historical declines, estimates of contemporary population size indicated a significant reduction. It is conceivable that the more recent demographic trajectories of these regions were of a complex nature that our models were not able to approximate; however, the low contemporary population sizes as compared to ancestral do reflect what is known about the extensive garden dormouse declines in the 20th and 21st centuries. In contrast, the Alps population showed signs of a massive and severe bottleneck with a more than 99.99% population reduction, followed by a more than twofold modern recovery. While a severe bottleneck may in part explain the lower genetic diversity of this region, the magnitude of the decline seems too extreme and the estimated contemporary  $N_e$  is inappropriately high for the Alpine region, which may indicate that none of the models tested accurately represented the demography of the Alpine population. While the accuracy and precision of *fastsimcoal2* are known to be affected by the number of SNPs analyzed and the amount of missing data (Nunziata and Weisrock 2018), *fastsimcoal2* is also sensitive to population census size, and may not accurately detect population declines when census size is low (Nunziata and Weisrock 2018). It is unclear how these and other factors, such as sample size and temporal heterogeneity, may influence model

**Table 3.** Model selection results from demographic inference using the site-frequency spectrum (SFS) in *fastsimcoal2*

| Population | Model | $N_e$ CON | $N_e$ ANC | $N_e$ BOT | TBOT | $\Delta L$ | AIC       | $\Delta AIC$ |
|------------|-------|-----------|-----------|-----------|------|------------|-----------|--------------|
| Alps       | BOT   | 128,117   | 82,321    | 81        | 2193 | 37,862     | 249,495   | 0            |
| Harz       | BOT   | 355       | 89,848    | 64,104    | 1535 | 78,191     | 513,286   | 0            |
| Northeast  | BOT   | 749       | 65,464    | 48,177    | 2626 | 81,403     | 484,918   | 0            |
| Northwest  | BOT   | 1241      | 57,891    | 53,781    | 4910 | 141,472    | 1,160,715 | 0            |
| Alps       | ISC   | 9158      | 47,203    | NA        | NA   | 37,882     | 249,591   | 95           |
| Harz       | ISC   | 135       | 179,897   | NA        | NA   | 78,194     | 513,305   | 19           |
| Northeast  | ISC   | 2318      | 24,407    | NA        | NA   | 81,403     | 484,921   | 2            |
| Northwest  | ISC   | 5931      | 112,707   | NA        | NA   | 141,481    | 1,160,762 | 47           |
| Alps       | Null  | 26,348    | NA        | NA        | NA   | 37,961     | 249,947   | 452          |
| Harz       | Null  | 34,143    | NA        | NA        | NA   | 78,713     | 515,688   | 2402         |
| Northeast  | Null  | 50,657    | NA        | NA        | NA   | 81,587     | 485,760   | 841          |
| Northwest  | Null  | 90,174    | NA        | NA        | NA   | 142,192    | 1,164,026 | 3311         |

Three models were run for each regional grouping of Alps, Harz, Northeast, and Northwest (population). Model refers to one of three model types: Null model of constant population size (Null), instantaneous size change/bottleneck (BOT), and ISC. Parameters were estimated for contemporary  $N_e$  ( $N_e$ CON), ancestral  $N_e$  ( $N_e$ ANC),  $N_e$  during bottleneck ( $N_e$ BOT), and time of bottleneck in number of generations (TBOT). Model fit and selection criterion are shown as delta likelihood, calculated as maximum estimated likelihood maximum observed likelihood ( $\Delta L$ ), AIC, and  $\Delta AIC$ .

output (Marchi et al. 2024). The Alps sample set spans a collection period of 19 years, covers a wide geographic area, and exhibits extremely low genetic diversity. These factors may have influenced parameter output for this region and led to upwardly biased  $N_e$ . Future analysis with whole-genome resequencing is needed to better explore the demographic history of the species, and of the Alpine region in particular.

Our findings of high genetic differentiation across the garden dormouse's core Central European range and the existence of four 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; Forcina et al. 2022). As in 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 in our study by both model-based (STRUCTURE), and non-model-based (PCA) approaches. This strong signal of genetic differentiation displayed by the Alpine population further confirms prior mtDNA genetic distance estimates and karyotypic divergence between the Alps and other regions across the garden dormouse's range ( $2n = 52$  "eastern Alps" and 54 "western Alps") (Perez et al. 2013). One specimen collected within our Alps boundary showed admixture with more Northern regions, but this specimen was collected north of the main Alpine ridge and likely does not belong to the highly differentiated Alpine population. We additionally found the regions south of the Alpine range to be differentiated from those north of the range, which supports that this mountain range may act as a barrier to garden dormouse dispersal (Perez et al. 2013). In contrast to previous research, our more comprehensive data sets also allowed us to detect additional moderate population structure within the Central European clade (Fig. 4), showing a clear pattern of differentiation between eastern and western Central Europe.

Combined with previous findings of reciprocal mtDNA monophyly (Perez et al. 2013), the significant differentiation of nuclear loci found in this study supports the designation of the Alpine garden dormouse population as an evolutionarily significant unit (ESU) according to the classic conservation genetics definition (Moritz 1994). 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 distinct habitat suggest the potential for long-term genetic separation, possibly arising from adaptation to local habitat conditions or climate. Future research should focus on investigating potential adaptive differences with whole-genome resequencing. Although the garden dormouse has been largely reported as demographically stable in the Alps (Bertolino 2017), our findings of low genetic diversity and high rates of inbreeding are indicative of small population sizes, and may indicate that population declines in this area have been greater than previously thought. Hence, the status of this population should be reevaluated. Perez et al. (2013) also found lower nucleotide diversity for this region via mtDNA analyses, and surmised that reductions in genetic diversity may have resulted from population expansion following a genetic bottleneck, which our results also support. The proportional differences in genetic diversity parameters between the Alpine and other populations and other sampling regions reported in this study are comparable to those seen in comparisons between mainland and both oceanic and montane "habitat" island populations of other threatened small mammalian species (Ditto and Frey 2007; White and Searle 2007; Rick et al. 2023), providing further support for the Alpine

population as an isolated, genetically distinct unit that may require the implementation of separate conservation strategies.

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 (Meinig and Büchner 2012; Bertolino 2017). Fossil evidence indicates that the garden dormouse originated on the Iberian Peninsula and spread out into Central Europe during the Holocene (Mansino et al. 2014), which may mean that remaining garden dormice in Eastern Europe represent small, relict populations on the edge of extirpation (Andèra 1986). This is somewhat supported by the higher autosomal genetic diversity and lower inbreeding estimate we detected for the westernmost Iberian clade, which reflects a larger and more connected population. However, while the garden dormouse is still regarded as relatively common in some parts of the Iberian Peninsula (Bertolino 2017), extensive declines and local extirpations have been noted in some parts of Portugal and Spain, and the distribution has been characterized as patchy and scattered across the region (Santoro et al. 2017). A comparative analysis using the entire mitogenome found that the Iberian clade exhibits relatively lower genetic diversity as compared to that of Central Europe, with the most genetically diverse individuals instead centered in France, possibly resulting from admixture following the emergence from glacial refugia in Central Europe (Forcina et al. 2022). Given that samples for our study were collected from the northernmost boundary of the Iberian Peninsula, our characterization of the Iberian clade may, therefore, be more reflective of this contiguous core population rather than of the wider peninsula. This sampling distribution likely also explains the minimal genetic structuring we found between the Iberian Peninsula and the central European regions, which contrasts with the strong differentiation shown in mitochondrial studies with broader sampling distributions (Perez et al. 2013; Forcina et al. 2022).

Given the known contractions and declines in the eastern edge of the garden dormouse's range, it is likely that the lower genetic diversity of the eastern populations in our study is reflective of genetic drift resulting from the modern isolation of these populations. Support for this is provided by our observation of reduced genetic diversity and greater rates of inbreeding for the more easterly populations in Germany (Harz mountains and Northeast), whereas 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 preliminary findings. Notably, a similar east–west pattern of differentiation has also been observed for the hazel dormouse (*Muscardinus avellanarius*), which overlaps the garden dormouse range and is distributed as two separate lineages or "partial ESUs," delineated along a similar geographical divide to the garden dormouse (Mouton et al. 2017; Leyhausen et al. 2022). 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.

Although the causes of the garden dormouse's east-to-west pattern of range contraction are not fully understood, landscape-level changes driven by land use and climate change have been partly 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 reducing 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 (Mahlert et al. 2018; Bennett and Richard 2021). 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 (Viñals-Domingo et al. 2020), while Alpine populations may hibernate for 7 months per year (Bertolino et al. 2001). 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.

The garden dormouse is currently listed as “Near Threatened” in the International Union for Conservation of Nature (IUCN) Red List (Bertolino et al. 2008), but its status will be changed to “Vulnerable” (S Bertolino, pers. comm.) in the next assessment (Bertolino 2017) due to the observed continuous decline in the species' range. Our results support this change in status, and suggest that the garden dormouse in Europe is threatened due to population isolation and low regional genetic diversity, especially at the eastern edge of its range. Hence, intervention is needed to mitigate further population extirpations. We further provide strong evidence for the designation of the Alpine subpopulation of the garden dormouse as a separate ESU. Further research is needed to determine if local adaptation is driving the genetic differentiation of this and other isolated garden dormouse populations, and how these potential adaptations may influence future population viability in the face of a changing climate. Our annotated genome will enable comparative analyses addressing open evolutionary questions of the *Eliomys* species complex, as well as inform specific conservation planning and management for the garden dormouse in Europe. We hope that the findings outlined in this study can help guide future research for the garden dormouse, while highlighting the need for further protection within regions of population isolation and/or low genetic diversity.

## Methods

### Sample collection

Tissue samples for nGBS were collected from across the European distribution in the frame of a larger conservation project on *E. quercinus* in Germany. Samples originated from Germany ( $n = 66$ ), other Central European countries ( $n = 25$ ), and Russia ( $n = 7$ ) and were collected between 1991 and 2020 (see Supplemental Table S2). All samples were taken opportunistically from dead-found individuals brought to museums or animal shelters, during monitoring, or during citizen-science activities, and all in compliance with respective local and national laws.

### Genome sequencing

To sequence a high-quality genome, we used a male individual collected in March 2023 in Mainz that was euthanized by a veterinarian due to injuries. A voucher specimen for this individual was prepared and is housed at the Zoological Research Museum Alexander Koenig under catalog number ZFMK MAM 2024.0263. DNA extraction, PacBio, and Hi-C sequencing followed standard protocols that are described in the Supplemental Methods.

### Transcriptome sequencing

We generated RNA- and Iso-Seq transcriptomics data from the same individual, following standard protocols that are described in the Supplemental Methods.

### Genotyping-by-sequencing

Tissue samples for normalized genotyping-by-sequencing (nGBS) were extracted using the DNeasy Blood and Tissue Kit (Qiagen) with slight protocol modifications to increase the 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 high-throughput sequencing. Selected samples ( $n = 100$ ) with  $\geq 300$  ng of high molecular mass DNA were sent for nGBS (restriction associated DNA sequencing [RAD-seq]) at LGC Genomics GmbH ([www.biosearchtech.com](http://www.biosearchtech.com)). Briefly, nGBS methodologies included enzymatic digestion with the enzyme MspI; purification; barcoding, library preparation, and PCR amplification with Illumina TruSeq amplification primers; size selection for 300–500 bp with the BluePippin system; and sequencing on an Illumina NextSeq 500/550. Sequences were then demultiplexed in house by LGC Genomics using proprietary software. Further details on enzymatic digestion, library preparation, and nGBS protocols can be found in Arvidsson et al. (2016) and Leyhausen et al. (2022).

### Genome assembly

We produced haplotype-specific assemblies from HiFi reads using hifiasm (v.0.19.5), integrating Hi-C reads for phasing (Cheng et al. 2021). Both haplotypes were subject to filtering procedure using the Foreign Contamination Screen (FCS) Tool (Astashyn et al. 2024). 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 Hi-C 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 Hi-C reads to both haplotype assemblies concatenated together using again Chromap but allowing for multimapping 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 Genomes and Darwin Tree of Life Projects, 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).

To estimate the assembly base accuracy, we used Merqury (Rhie et al. 2020) with the HiFi reads. It should be noted that the

QV estimates represent an upper bound as they are based on the HiFi reads used for assembly. To assess gene completeness, we used compleasm 0.2.2 (Huang and Li 2023) with the BUSCO odb10 set of 9226 near-universally conserved mammalian genes and TOGA (below).

### 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 colinear 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 version 1.1.4 (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 et al. 2018; Chen 2023) 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 were 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 Iso-Seq 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 Iso-Seq package. Additionally, reads were filtered for alignment quality using a custom Perl script (Supplemental Code). 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 do not have a TOGA transcript prediction to incorporate lineage-specific and noncoding transcripts into the annotation.

### Demographic history inferred from the reference genome

We used PSMC (Li and Durbin 2011) to infer the demographic history of the garden dormouse across the past 1 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 ge-

nome 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”`), as recommended by Hilgers et al. (2024). 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  $5.7 \times 10^{-9}$  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

#### Genotyping, filtering, and variant calling

Full methods for processing the raw GBS data, including trimming and alignment, are described in the Supplemental Methods. Variant calling was performed with the `ref_map.pl` pipeline for reference aligned reads in Stacks 2 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 to remove strong physical LD (-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 the X Chromosome 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 >80% 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 LD can influence some analyses such as PCAs, we created a pruned SNP set using the `+prune` plugin in BCftools (Danecek et al. 2021) using an LD threshold of  $r^2 < 0.5$  and a sliding window of 1000, 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 (Wang 2018; O’Connell et al. 2019). 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 with the full SNP set pruned by  $r^2 < 0.2$  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.

### Genetic diversity

Genetic diversity estimates can be biased by a number of factors, including allele frequencies, unbalanced sample size, and missing data (Schmidt et al. 2021). We calculated observed ( $H_{\text{obs}}$ ) and expected ( $H_{\text{exp}}$ ) heterozygosity, nucleotide diversity ( $\pi$ ), number of polymorphic loci, and the inbreeding coefficient  $F_{\text{IS}}$  with the pruned SNP set using the “population” function in stacks (Rochette et al. 2019). Because biases due to unbalanced sample sizes can confound comparisons among populations, we additionally report  $H_{\text{exp}}$  derived from genotype-wide autosomal data, which is not subject to the same bias (Schmidt et al. 2021; Sopniewski and Catullo 2024). Autosomal heterozygosity was generated from the aligned Stacks catalog using the “population” function with a minor allele count of 2 to omit singletons and with one random SNP selected per locus. To further minimize the effects of differences in sample size, we calculated SNP-based genetic diversity parameters for the Northwest population in four separate runs of  $n=12$  samples per run and report the mean values. We investigated genetic differentiation by calculating pairwise values of Weir and Cockerham’s unbiased estimator of  $F_{\text{ST}}$  (Weir and Cockerham 1984) in the R package hierfstat (Goudet 2005) using the pruned SNP set and with 95% confidence intervals generated with 1000 bootstrap iterations.

### Population structure

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 2018) 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 first used the pruned SNP set to visualize population clustering via PCA in the R package *adeigenet* (Jombart and Bateman 2008). We estimated the most probable number of genetic clusters  $K$  within our sample set using the full SNP set. First, we implemented a DAPC using the “find.clusters” function in *adeigenet* (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. Five PCs were retained to explain ~52% of the total retained variation without our data set.

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  $\Delta K$  (Evanno et al. 2005) and mean probability (MeanLnP( $K$ )). STRUCTURE is known to be sensitive to differences in sample size in populations (Puechmaile 2016), which was a factor in our data set. 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 data set, for values of  $K$  from 1 to 5.

### Recent demographic history inferred from the site-frequency spectrum

We inferred demographic history among our individual populations using a sequential Markovian coalescent approximation approach in *fastSimcoal2* v2.08 (Excoffier and Foll 2011; Excoffier et al. 2021). We omitted the Central and Iberian Peninsula populations and the Italian and Russian samples due to their low sample sizes. To generate the input site frequency spectrum (SFS) files, we used the distribution of minor allele frequencies in the full SNP set to create a folded SFS for each population with the R program *darTR* 2.9.7 (Gruber et al. 2018), using nearest-neighbor imputation to replace missing data. While *fastSimcoal2* can be used to infer complex demographic histories, we opted to test models for each population separately to avoid model overparameterization, given our small sample sizes (Alps,  $n=9$ ; Harz,  $n=12$ ; Northeast,  $n=7$ ; Northwest,  $n=48$ ). For each population, we compared three models: “null,” “instantaneous size change” (a.k.a., bottleneck), and “incremental size change.” The null model assumed a constant population size over time and inferred only one parameter, contemporary effective population size ( $N_e$ ). The instantaneous size change model assumed a rapid population decline and inferred four parameters: ancestral  $N_e$ , time of the bottleneck (in generations), population size during the bottleneck, and contemporary  $N_e$ . The ISC model assumed either steady population growth or decline over time and inferred five parameters: ancestral  $N_e$ , time of ancestral  $N_e$  (in generations), population growth rate (with a negative value indicating a positive growth rate), rate of change of  $N_e$ , and contemporary  $N_e$ . For all models, we assumed a generation time of 1.5 years and a mutation rate of  $5.7 \times 10^{-9}$ . Parameter ranges were uniformly distributed between 10 and  $10^5$ . We ran 100 independent iterations of each model per population with 70 EMC cycles and  $1 \times 10^6$  simulations per model run. We selected the best run within each model as the one with the highest likelihood (smallest difference between maximum likelihood and obtained likelihood) using the custom script *fsc-selectbestrun.sh* (<https://raw.githubusercontent.com/speciationgenomics/scripts/master/fsc-selectbestrun.sh>). Because maximum likelihood can be influenced by parameter size, model selection within each population was then performed using AIC comparisons, which account for the number of parameters in the selection process, with AIC calculated using the custom script *calculateAIC.sh* (<https://github.com/speciationgenomics/scripts/blob/master/calculateAIC.sh>).

### Software availability

Most of our analyses relied on publicly available tools and methods. New scripts are available at GitHub ([https://github.com/pabyerly/TBG\\_GardenDormouse](https://github.com/pabyerly/TBG_GardenDormouse)) and as Supplemental Code.

### Data access

Genomic and transcriptomics read data, our annotated genome and the trimmed ddRAD reads have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA1142780. The haplotype assemblies, gene, and transposable element annotations are available for download at <https://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/>.

## Competing interest statement

The authors declare no competing interests.

## Acknowledgments

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 (LOEWE/1/10/519/03/03.001(0014)/52). 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 Ansoorge, 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 Schmid, the Schwarzwald National Park, the Gran Paradiso 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.

**Author contributions:** A.v.T., C.N., and M.H. designed and supervised the study, labwork, and data analysis. Ch.G., A.B.H., Ca.G., C.B., and H.J.B. performed labwork. P.A.B., E.L., L.H., S.L., S.W., and T.S. conducted the statistical and bioinformatic analysis. Sv.B., J.L., H.M., E.M.F.P., S.P.S., A.M., S.B., G.V., T.B., L.F., L.V., and S.A.M. contributed samples, supervised sample collection, and/or provided expert species knowledge. P.A.B., A.v.T., E.L., L.H., and M.H. wrote the manuscript with input from all authors.

## References

- Anděra M. 1986. Dormice (Gliridae) in Czechoslovakia. Part I: *Glis glis*, *Eliomys quercinus* (Rodentia: Mammalia). *Folia Mus Rer Natur Bohem Occid 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/assets/v6/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, Strobe PK, Sylla PM, Wagner L, et al. 2024. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biol* **25**: 60. doi:10.1186/s13059-024-03198-7
- 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. doi:10.1007/s42991-021-00118-1
- Bertolino S. 2017. Distribution and status of the declining garden dormouse *Eliomys quercinus*. *Mamm Rev* **47**: 133–147. doi:10.1111/mam.12087
- Bertolino S, Viano C, Currado I. 2001. Population dynamics, breeding patterns and spatial use of the garden dormouse (*Eliomys quercinus*) in an Alpine habitat. *J Zool* **253**: 513–521. doi:10.1017/S0952836901000474
- Bertolino S, Amori G, Henttonen H, Zagorodnyuk I, Zima J, Juškaitis R, Meinig H, Kryštufek B. 2008. *Eliomys quercinus*. The IUCN Red List of Threatened Species 2008: e.T7618A12835766. <https://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T7618A12835766.en>. Accessed on 7 October 2024.
- 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* **26**: 25–35. doi:10.4404/hystrix-26.1-10263
- Blumer M, Brown T, Freitas MB, Destro AL, Oliveira JA, Morales AE, Schell T, Greve C, Pippel M, Jebb D, et al. 2022. Gene losses in the common vampire bat illuminate molecular adaptations to blood feeding. *Sci Adv* **8**: eabm6494. doi:10.1126/sciadv.abm6494
- Böhne A, Thiel-Bender C, Kukowka S, Darwin Tree of Life Barcoding Collective, Wellcome Sanger Institute Tree of Life Programme, Wellcome Sanger Institute Scientific Operations: DNA Pipelines Collective, Tree of Life Core Informatics Collective Darwin Tree of Life Consortium. 2023. The genome sequence of the hazel dormouse, *Muscardinus avellanarius* (Linnaeus, 1758). *Wellcome Open Res* **8**: 514. doi:10.12688/wellcomeopenres.20360.1
- Brandies P, Peel E, Hogg CJ, Belov K. 2019. The value of reference genomes in the conservation of threatened species. *Genes (Basel)* **10**: 846. doi:10.3390/genes10110846
- Büchner S, Lang J, Meinig HU, Nava TF, Zisl K. 2024. The garden dormouse as a research and conservation priority in a German large-scale citizen science project. *Acta Zool Bulg Suppl* **19**: 1–8.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. *Genome Res* **25**: 1656–1665. doi:10.1101/gr.196485.115
- Campana MG, Parker LD, Hawkins MTR, Young HS, Helgen KM, Szykman Gunther M, Woodroffe R, Maldonado JE, Fleischer RC. 2016. Genome sequence, population history, and pelage genetics of the endangered African wild dog (*Lycaon pictus*). *BMC Genomics* **17**: 1013. doi:10.1186/s12864-016-3368-9
- Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**: e107. doi:10.1002/imt2.107
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890. doi:10.1093/bioinformatics/bty560
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170–175. doi:10.1038/s41592-020-01056-5
- Combe FJ, Juškaitis R, Trout RC, Bird S, Ellis JS, Norrey J, Al-Fulaij N, White I, Harris WE. 2023. Density and climate effects on age-specific survival and population growth: consequences for hibernating mammals. *Anim Conserv* **26**: 317–330. doi:10.1111/acv.12843
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158. doi:10.1093/bioinformatics/btr330
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008. doi:10.1093/gigascience/giab008
- DeSaix MG, Bulluck LP, Eckert AJ, Viverette CB, Boves TJ, Reese JA, Tonra CM, Dyer RJ. 2019. Population assignment reveals low migratory connectivity in a weakly structured songbird. *Mol Ecol* **28**: 2122–2135. doi:10.1111/mec.15083
- Ditto AM, Frey JK. 2007. Effects of ecogeographic variables on genetic variation in montane mammals: implications for conservation in a global warming scenario. *J Biogeogr* **34**: 1136–1149. doi:10.1111/j.1365-2699.2007.01700.x
- du Plessis SJ, Blaxter M, Koepfli K-P, Chadwick EA, Hailer F. 2023. Genomics reveals complex population history and unexpected diversity of Eurasian otters (*Lutra lutra*) in Britain relative to genetic methods. *Mol Biol Evol* **40**: msad207. doi:10.1093/molbev/msad207
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* **14**: 2611–2620. doi:10.1111/j.1365-294X.2005.02553.x
- Excoffier L, Foll M. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**: 1332–1334. doi:10.1093/bioinformatics/btr124
- Excoffier L, Marchi N, Marques DA, Matthey-Doret R, Gouy A, Sousa VC. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics* **37**: 4882–4885. doi:10.1093/bioinformatics/btab468
- Forcina G, Camacho-Sanchez M, Cornellas A, Leonard JA. 2022. Complete mitogenomes reveal limited genetic variability in the garden dormouse *Eliomys quercinus* of the Iberian Peninsula. *Anim Biodivers Conserv* **45**: 107–122. doi:10.32800/abc.2022.45.0107
- 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 Ecol Evol* **37**: 197–202. doi:10.1016/j.tree.2021.11.008

- Francis RM. 2017. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour* **17**: 27–32. doi:10.1111/1755-0998.12509
- 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. doi:10.1093/nar/gkaa1087
- 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. doi:10.1093/molbev/msw144
- Gippoliti S, Amori G. 2007. Beyond threatened species and reintroduction: establishing priorities for conservation and breeding programmes for European rodents in zoos. *Int Zoo Yearb* **41**: 194–202. doi:10.1111/j.1748-1090.2007.00002.x
- Goudet J. 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Mol Ecol Notes* **5**: 184–186. doi:10.1111/j.1471-8286.2004.00828.x
- Gruber B, Unmack PJ, Berry OF, Georges A. 2018. dart: an R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Mol Ecol Resour* **18**: 691–699. doi:10.1111/1755-0998.12745
- Halley DJ, Saveljev AP, Rosell F. 2021. Population and distribution of beavers *Castor fiber* and *Castor canadensis* in Eurasia. *Mamm Rev* **51**: 1–24. doi:10.1111/mam.12216
- Harris RS. 2007. “Improved pairwise alignment of genomic DNA.” PhD thesis, The Pennsylvania State University, University Park, PA.
- Hilgers L, Liu S, Jensen A, Brown T, Cousins T, Schweiger R, Guschanski K, Hiller M. 2024. Avoidable false PSMC population size peaks occur across numerous studies. bioRxiv doi:10.1101/2024.06.17.599025
- 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**: gaa153. doi:10.1093/giga/science/gaa153
- Huang N, Li H. 2023. compleasm: a faster and more accurate reimplementation of BUSCO. *Bioinformatics* **39**: btad595. doi:10.1093/bioinformatics/btad595
- Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, Devanna P, Winkler S, Jermin LS, Skirmuntt EC, Katzourakis A, et al. 2020. Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**: 578–584. doi:10.1038/s41586-020-2486-3
- Jombart T, Bateman A. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403–1405. doi:10.1093/bioinformatics/btn129
- Karimi K, Do DN, Wang J, Easley J, Borzouie S, Sargolzaei M, Plastow G, Wang Z, Miar Y. 2022. A chromosome-level genome assembly reveals genomic characteristics of the American mink (*Neogale vison*). *Commun Biol* **5**: 1381. doi:10.1038/s42003-022-04341-5
- 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. *Divers Distrib* **27**: 1792–1806. doi:10.1111/ddi.13368
- 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* **100**: 11484–11489. doi:10.1073/pnas.1932072100
- 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. doi:10.1038/s41587-019-0201-4
- 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. doi:10.1126/science.abn3107
- Lane JE, Kruuk LEB, Charmanier A, Murie JO, Dobson FS. 2012. Delayed phenology and reduced fitness associated with climate change in a wild hibernator. *Nature* **489**: 554–557. doi:10.1038/nature11335
- 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. doi:10.3390/su14159327
- Leyhausen J, Cocchiario B, Nowak C, Ansoerge 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. doi:10.1007/s12686-022-01253-8
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi:10.1093/bioinformatics/bty191
- Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. *Bioinformatics* **27**: 2987–2993. doi:10.1093/bioinformatics/btr509
- Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**: 493–496. doi:10.1038/nature10231
- 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. doi:10.1007/s13364-012-0085-x
- Lu X, Costeur L, Huguency 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. doi:10.1080/14772019.2021.1888814
- 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). *Commun Biol* **6**: 746. doi:10.1038/s42003-023-05137-x
- 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**: e31225. doi:10.7554/eLife.31225
- 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. doi:10.1093/bioinformatics/btq559
- Mansino S, García-Alix A, Ruiz-Sánchez FJ, Montoya P. 2014. A new eliomys from the late miocene of Spain and its implications for the phylogeny of the genus. *Acta Palaeontol Pol* **60**: 577–588.
- Marchi N, Kapopoulou A, Excoffier L. 2024. Demogenomic inference from spatially and temporally heterogeneous samples. *Mol Ecol Resour* **24**: e13877. doi:10.1111/1755-0998.13877
- Martchenko D, Shafer ABA. 2023. Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. *Heredity (Edinb)* **131**: 273–281. doi:10.1038/s41437-023-00643-4
- Mead D, Fingland K, Cripps R, Portela Miguez R, Smith M, Corton C, Oliver K, Skelton J, Betteridge E, Doulan J, et al. 2020. The genome sequence of the eastern grey squirrel, *Sciurus carolinensis* Gmelin, 1788. *Wellcome Open Res* **5**: 27. doi:10.12688/wellcomeopenres.15721.1
- Meinig H, Büchner S. 2012. The current situation of the garden dormouse (*Eliomys quercinus*) in Germany. *Peckiana* **8**: 129–134.
- Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. 2017. Differences between germline and somatic mutation rates in humans and mice. *Nat Commun* **8**: 15183. doi:10.1038/ncomms15183
- Moritz C. 1994. Defining “evolutionarily significant units” for conservation. *Trends Ecol Evol* **9**: 373–375. doi:10.1016/0169-5347(94)90057-4
- Mouton A, Mortelliti A, Grill A, Sara M, Kryštufek B, Juškaitis R, Latienne A, Amori G, Randi E, Büchner S, et al. 2017. Evolutionary history and species delimitations: a case study of the hazel dormouse, *Muscardinus avelanarius*. *Conserv Genet* **18**: 181–196. doi:10.1007/s10592-016-0892-8
- Nadachowski A, Daoud A. 1995. Patterns of myoxid evolution in the Pliocene and Pleistocene of Europe. *Hystrix* **6**: 141–149. doi:10.4404/hystrix-6.1-2-4023
- Nieto-Blázquez ME, Schreiber D, Mueller SA, Koch K, Nowak C, Pfenninger M. 2022. Human impact on the recent population history of the elusive European wildcat inferred from whole genome data. *BMC Genomics* **23**: 709. doi:10.1186/s12864-022-08930-w
- Nunziata SO, Weisrock DW. 2018. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity (Edinb)* **120**: 196–207. doi:10.1038/s41437-017-0037-y
- O’Connell KA, Mulder KP, Maldonado J, Currie KL, Ferraro DM. 2019. Sampling related individuals within ponds biases estimates of population structure in a pond-breeding amphibian. *Ecol Evol* **9**: 3620–3636. doi:10.1002/ece3.4994
- Osipova E, Hecker N, Hiller M. 2019. Repeatfiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. *GigaScience* **8**: giz132. doi:10.1093/gigascience/giz132
- Paez S, Kraus RHS, Shapiro B, Gilbert MTP, Jarvis ED, Al-Ajli FO, Ceballos G, Crawford AJ, Fedrigo O, Johnson RN, et al. 2022. Reference genomes for conservation. *Science* **377**: 364–366. doi:10.1126/science.abm8127
- Perez GCL, Libois R, Nieberding CM. 2013. Phylogeography of the garden dormouse *Eliomys quercinus* in the western Palearctic region. *J Mammal* **94**: 202–217. doi:10.1644/11-MAMM-A-404.1
- Pérez-España S. 2017. Conservation genetics in the European union – biases, gaps and future directions. *Biol Conserv* **209**: 130–136. doi:10.1016/j.biocon.2017.01.020
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**: e37135. doi:10.1371/journal.pone.0037135
- Pina-Martins F, Silva DN, Fino J, Paulo OS. 2017. Structure\_threader: an improved method for automation and parallelization of programs

- structure, fastStructure and Maverick on multicore CPU systems. *Mol Ecol Resour* **17**: e268–e274. doi:10.1111/1755-0998.12702
- Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, Newburger D, Dijamco J, Nguyen N, Afshar PT, et al. 2018. A universal SNP and small-indel variant caller using deep neural networks. *Nat Biotechnol* **36**: 983–987. doi:10.1038/nbt.4235
- Pritchard JK. 2000. Inference of population structure using multilocus genotypes. *Genetics* **155**: 945–959. doi:10.1093/genetics/155.2.945
- Puechmaile SJ. 2016. The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. *Mol Ecol Resour* **16**: 608–627. doi:10.1111/1755-0998.12512
- Rammou D-L, Astaras C, Migli D, Boutsis G, Galanaki A, Kominos T, Youlatos D. 2022. European ground squirrels at the edge: current distribution status and anticipated impact of climate on Europe's southern-most population. *Land (Basel)* **11**: 301. doi:10.3390/land11020301
- R Core Team. 2018. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>.
- Reiners TE, Eidsenck J, Neumann K, Nowak C. 2014. Preservation of genetic diversity in a wild and captive population of a rapidly declining mammal, the Common hamster of the French Alsace region. *Mamm Biol* **79**: 240–246. doi:10.1016/j.mambio.2013.10.004
- Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**: 245. doi:10.1186/s13059-020-02134-9
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammanan A, Kim J, et al. 2021. Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**: 737–746. doi:10.1038/s41586-021-03451-0
- Rick K, Byrne M, Cameron S, Cooper SJB, Dunlop J, Hill B, Lohr C, Mitchell NJ, Moritz C, Travouillon KJ, et al. 2023. Population genomic diversity and structure in the golden bandicoot: a history of isolation, extirpation, and conservation. *Heredity (Edinb)* **131**: 374–386. doi:10.1038/s41437-023-00653-2
- Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: analytical methods for paired-end sequencing improve RADseq-based population genomics. *Mol Ecol* **28**: 4737–4754. doi:10.1111/mec.15253
- Santoro S, Sanchez-Suarez C, Rouco C, Palomo LJ, Fernández MC, Kufner MB, Moreno S. 2017. Long-term data from a small mammal community reveal loss of diversity and potential effects of local climate change. *Curr Zool* **63**: 515–523. doi:10.1093/cz/zow109
- Schaub M, Vaterlaus-Schlegel C. 2001. Annual and seasonal variation of survival rates in the garden dormouse (*Eliomys quercinus*). *J Zool* **255**: 89–96. doi:10.1017/S0952836901001133
- Schmidt TL, Jasper M-E, Weeks AR, Hoffmann AA. 2021. Unbiased population heterozygosity estimates from genome-wide sequence data. *Methods Ecol Evol* **12**: 1888–1898. doi:10.1111/2041-210X.13659
- Severn-Ellis AA, Scheben A, Neik TX, Saad NSM, Pradhan A, Batley J. 2020. Genotyping for species identification and diversity assessment using double-digest restriction site-associated DNA sequencing (ddRAD-Seq). In *Methods in molecular biology*, Vol. 2107, pp. 159–187. Humana Press Inc., Totowa, NJ.
- Sharma V, Hiller M. 2017. Increased alignment sensitivity improves the usage of genome alignments for comparative gene annotation. *Nucleic Acids Res* **45**: 8369–8377. doi:10.1093/nar/gkx554
- Shumate A, Wong B, Perteau G, Perteau M. 2022. Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Comput Biol* **18**: e1009730. doi:10.1371/journal.pcbi.1009730
- Sopniewski J, Catullo RA. 2024. Estimates of heterozygosity from single nucleotide polymorphism markers are context-dependent and often wrong. *Mol Ecol Resour* **24**: e13947. doi:10.1111/1755-0998.13947
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313. doi:10.1093/bioinformatics/btu033
- Suarez HG, Langer BE, Ladde P, Hiller M. 2017. Chaincleaner improves genome alignment specificity and sensitivity. *Bioinformatics* **33**: 1596–1603. doi:10.1093/bioinformatics/btx024
- Szarmach SJ, Brelford A, Witt CC, Toews DPL. 2021. Comparing divergence landscapes from reduced-representation and whole genome resequencing in the yellow-rumped warbler (*Setophaga coronata*) species complex. *Mol Ecol* **30**: 5994–6005. doi:10.1111/mec.15940
- Takach BV, Sargent H, Penton CE, Rick K, Murphy BP, Neave G. 2023. Population genomics and conservation management of the threatened black-footed tree-rat (*Mesembriomys gouldii*) in northern Australia. *Heredity (Edinb)* **130**: 278–288. doi:10.1038/s41437-023-00601-0
- Talla V, Mrazek V, Höglund J, Backström N. 2023. Whole genome re-sequencing uncovers significant population structure and low genetic diversity in the endangered clouded apollo (*Parnassius mnemosyne*) in Sweden. *Conserv Genet* **24**: 305–314. doi:10.1007/s10592-023-01502-9
- Theissingner 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 Genet* **39**: 545–559. doi:10.1016/j.tig.2023.01.005
- Tschiya 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. doi:10.1093/gbe/evaa069
- 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. doi:10.1101/gr.186148.114
- Verde Arregoitia LD. 2016. Biases, gaps, and opportunities in mammalian extinction risk research. *Mamm Rev* **46**: 17–29. doi:10.1111/mam.12049
- 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. doi:10.1101/gr.276070.121
- Viñals-Domingo A, Bertolino S, López-Iborra G, Gil-Delgado JA. 2020. Seasonal survival in a non-hibernating Mediterranean garden dormouse population. *Mamm Biol* **100**: 581–589. doi:10.1007/s42991-020-00067-1
- Wang J. 2018. Effects of sampling close relatives on some elementary population genetics analyses. *Mol Ecol Resour* **18**: 41–54. doi:10.1111/1755-0998.12708
- Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution (N Y)* **38**: 1358. doi:10.2307/2408641
- White TA, Searle JB. 2007. Genetic diversity and population size: island populations of the common shrew, *Sorex araneus*. *Mol Ecol* **16**: 2005–2016. doi:10.1111/j.1365-294X.2007.03296.x
- Wold J, Koepfli K-P, Galla SJ, Eccles D, Hogg CJ, Le Lec MF, Guhlin J, Santure AW, Steeves TE. 2021. Expanding the conservation genomics toolbox: incorporating structural variants to enhance genomic studies for species of conservation concern. *Mol Ecol* **30**: 5949–5965. doi:10.1111/mec.16141
- 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 (Bethesda)* **12**: jkab373. doi: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. doi:10.1111/1755-0998.13211
- 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. doi:10.1111/2041-210X.12628
- 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. doi:10.1038/s41467-021-26865-w
- Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**: btac808. doi:10.1093/bioinformatics/btac808
- Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**: 240–245. doi:10.1038/s41586-020-2876-6

Received February 21, 2024; accepted in revised form September 12, 2024.



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

Paige A. Byerly, Alina von Thaden, Evgeny Leushkin, et al.

*Genome Res.* published online November 14, 2024

Access the most recent version at doi:[10.1101/gr.279066.124](https://doi.org/10.1101/gr.279066.124)

---

**Supplemental Material** <http://genome.cshlp.org/content/suppl/2024/10/28/gr.279066.124.DC1>

**P<P** Published online November 14, 2024 in advance of the print journal.

**Creative Commons License** This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <https://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---