PNAS RESEARCH ARTICLE | GENETICS OPEN ACCESS

# Genome analyses suggest recent speciation and postglacial isolation in the Norwegian lemming

Edana Lord[a,b,c,1], Isabelle S. Feinauer[a,b,c], André E. R. Soares[d], Vendela K. Lagerholm[a,e], Karin Näsvall[f], Erik Ersmark[a,e], Remi-André Olsen[g], Stefan Prost[h,i,j,k], Elena A. Kuzmina[l], Nickolay G. Smirnov[l], John R. Stewart[m], Monika V. Knul[n], Pierre Noiret[o,2], Mietje Germonpré[p], Dorothee Ehrich[q], Ivan Pokrovsky[r], Vadim B. Fedorov[s], Anna V. Goropashnaya[s], Love Dalén[a,b,c], and David Díez-del-Molino[a,b,c,1]

Affiliations are included on p. 8.

Edited by Marcus Feldman, Stanford University, Stanford, CA; received November 22, 2024; accepted April 23, 2025

**The Norwegian lemming (*Lemmus lemmus*) is a small rodent distributed across the Fennoscandian mountain tundra and the Kola Peninsula. The Norwegian lemming likely evolved during the Late Pleistocene and inhabited Fennoscandia shortly prior to the Last Glacial Maximum. However, the exact timing and origins of the species, and its phylogenetic position relative to the closely related Siberian lemming (*Lemmus sibiricus*) remain disputed. Moreover, the presence of ancient or contemporary gene flow between both species is largely untested. The Norwegian lemming displays characteristic phenotypic and behavioral adaptations (e.g., coat color, aggression) that are not present in other *Lemmus* species. We generated a de novo genome assembly for the Norwegian lemming and resequenced nine modern and two ancient *Lemmus* spp. genomes. We show that all *Lemmus* species form distinct monophyletic clades, with concordant topology between the mitochondrial and nuclear genome phylogenies. The Siberian lemming is divided into two distinct but paraphyletic clades, one in the east and one in the west, where the western clade represents a sister taxon to the Norwegian lemming. We estimate that the Norwegian and western Siberian lemming diverged shortly before the Last Glacial Maximum, making the Norwegian lemming one of the youngest known mammalian species. We did not find any indication of gene flow between *L. lemmus* and *L. sibiricus*, suggesting postglacial isolation of *L. lemmus*. Furthermore, we identify species-specific genomic differences in genes related to coat color and fat transport, which are likely associated with the distinctive coloration and overwintering behavior observed in the Norwegian lemming.**

genomes | evolution | gene flow | Last Glacial Maximum

## Significance

Speciation is often driven by climatic fluctuations. Recent studies using whole genome data are shedding light on the process of speciation and the occurrence of postspeciation gene flow in mammalian taxa. Here, we present a genome assembly and whole genome data for the Norwegian lemming (*Lemmus lemmus*), and some of its relatives. We confirm that the Norwegian lemming has a young divergence time for a mammalian species. Unlike other taxa with such recent speciation times, the Norwegian lemming does not show any indication of postspeciation gene flow with other *Lemmus* species. We further identify unique mutations in the Norwegian lemming genome that are likely related to coat coloration and fat transport.

Climatic fluctuations were a driver of speciation during the Pleistocene (1). Consequently, several taxa, such as lions, bears, voles, and boreal birds, have relatively young speciation times (2–6). In fact, isolation in interglacial refugia has been suggested as the mechanism for speciation in a range of Arctic taxa (7). However, glacial periods also provided conditions for divergence in cold-adapted taxa through the colonization of previously unoccupied areas and vicariance driven by glacial ice sheets (8, 9).

With the rise of high-throughput sequencing technologies in the last decade, the number of studies using whole-genome data to investigate the evolutionary history of wild species has increased dramatically (10–12). In these studies, the presence of postspeciation gene flow among species emerges as a prevalent evolutionary component of the speciation process for many taxa. For example, gene flow among bear species has been common throughout the Late Pleistocene, including gene flow between polar and brown bears (12), cave and brown bears (13), as well as between short-faced and spectacled bears (14). Similarly, complex patterns of interspecific hybridization seem to have played an important role in the adaptive evolution process of big cats (11) and birds (15). Climatic fluctuations leading to range expansions likely facilitated gene flow in these species.

One taxon that underwent large range expansions and contractions during the Pleistocene are true lemmings (genus: *Lemmus*), a group that consists of several species distributed across Eurasia and North America. The current phylogenetic knowledge of *Lemmus*, based on short mtDNA, suggests that there are at least four different species: the Norwegian lemming (*Lemmus lemmus*), the Siberian brown lemming (*Lemmus sibiricus*), the Amur lemming (*Lemmus amurensis*), and the Nearctic brown lemming (*Lemmus trimucronatus*) which spans both sides of Beringia (16–20). The Siberian lemming is further divided into two unique lineages—the western lineage hereafter referred to as *L. sibiricus* West and the eastern lineage hereafter referred to as *L. sibiricus* East (20). It is unclear

Author contributions: E.L., L.D., and D.D.-d.-M. designed research; E.L., I.S.F., A.E.R.S., V.K.L., K.N., E.E., R.-A.O., S.P., E.A.K., N.G.S., J.R.S., M.V.K., P.N., M.G., D.E., I.P., V.B.F., A.V.G., L.D., and D.D.-d.-M. performed research; V.K.L., K.N., E.E., V.B.F., and A.V.G. produced data; E.L., I.S.F., A.E.R.S., K.N., R.-A.O., and D.D.-d.-M. analyzed data; and E.L., I.S.F., A.E.R.S., L.D., and D.D.-d.-M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: edana.lord@zoologi.su.se or diez.molino@gmail.com.

[2]Deceased April 18, 2025.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2424333122/-/DCSupplemental.

Published June 30, 2025.

whether these two lineages represent distinct species. Divergence estimates based on fossil-calibrated mitochondrial data indicate that the diversification of *Lemmus* into different species took place during the Early to Middle Pleistocene (20).

The Norwegian lemming is the only endemic mammal in Fennoscandia. It has been suggested to have one of the youngest divergence times among mammals based on short mtDNA estimates, diverging from its sister taxon, *L. sibiricus* West, prior to the Last Glacial Maximum (LGM) (17, 21, 22). *L. lemmus* displays unique phenotypic adaptations relative to other lemmings, including conspicuous antipredatory behavior as well as bright coat color, which have been hypothesized to be aposematic traits (23, 24). Given these and other morphological differences, *L. lemmus* and *L. sibiricus* West are considered to be distinct species despite their close genetic relationship. Further conflicting the species delimitation is the possibility of gene flow between *L. lemmus* and *L. sibiricus* West. For example, in ref. 21 mitochondrial analyses identified a *L. sibiricus* West haplotype in Sirijorda Cave, Norway ~8,000 y ago (ka BP), suggesting that gene flow might have occurred between these two species during the Holocene.

In this study, we investigate the origins and phylogenomic relationships among the Norwegian and Siberian lemmings, examine the presence and extent of gene flow among species, and identify how specific genomic mutations may be related to the unique Norwegian lemming phenotype. To do this, we generated a de novo reference genome assembly for *L. lemmus*, and resequenced nine high-coverage genomes from four different *Lemmus* species/lineages, as well as low-coverage genomic and mitochondrial data for two ancient lemming specimens dating from the Late Pleistocene.

## Results

**Genome Assembly and Resequencing Data.** We generated a de novo assembly for the Norwegian lemming, which had a final size of 2.42 Gb and was composed of 11,293 scaffolds. The scaffold N50 was 1.2 Mb with 95% of the genome in scaffolds >144 kb long (*SI Appendix*, Fig. S1). BUSCO showed that the assembly contains 3,867 (94.2%) single-copy genes, 100 (2.4%) missing, 106 (2.6%) fragmented, and 31 (0.8%) duplicated single-copy orthologs of the 4,104 in the mammalian database. Using their relative coverage in males and females, we identified 140 scaffolds linked to chromosome X (*SI Appendix*, Fig. S2). These represent a total size of 28.8 Mb or ~1.2% of the size of the assembly.

We then reconstructed nuclear genomes of nine individuals across four lemming species from Fennoscandia and the Holarctic [*L. lemmus* (n = 5), *L. sibiricus* West (n = 2), *L. sibiricus* East (n = 1), and *L. trimucronatus* (n = 1) to an average of 18.4× depth (Fig. 1 and *SI Appendix*, Table S1)]. We additionally recovered low-coverage genomic data from two ancient specimens: one from Bridged Pot Cave (England) dated to 12,573 BP (0.11×), and another one from Pymva Shor (Russia) dated to 9,655 BP (0.03×). We also reconstructed complete mitogenomes from the nine modern lemming samples (average depth of 143.3×) as well as the two ancient samples (average depth 53.1×). Finally, we obtained short mitochondrial data (522 bp) from 55 additional ancient lemming specimens from across their Late Pleistocene range (*SI Appendix*).

**Phylogenetic Relationships, Divergence Times, and Demography.** We examined phylogenetic relationships among our *Lemmus* samples at three levels: short mitochondrial DNA data (n = 131), complete mitogenomes (n = 11), and genomic nuclear data (n = 11). We found that all three phylogenies show concordant trees with

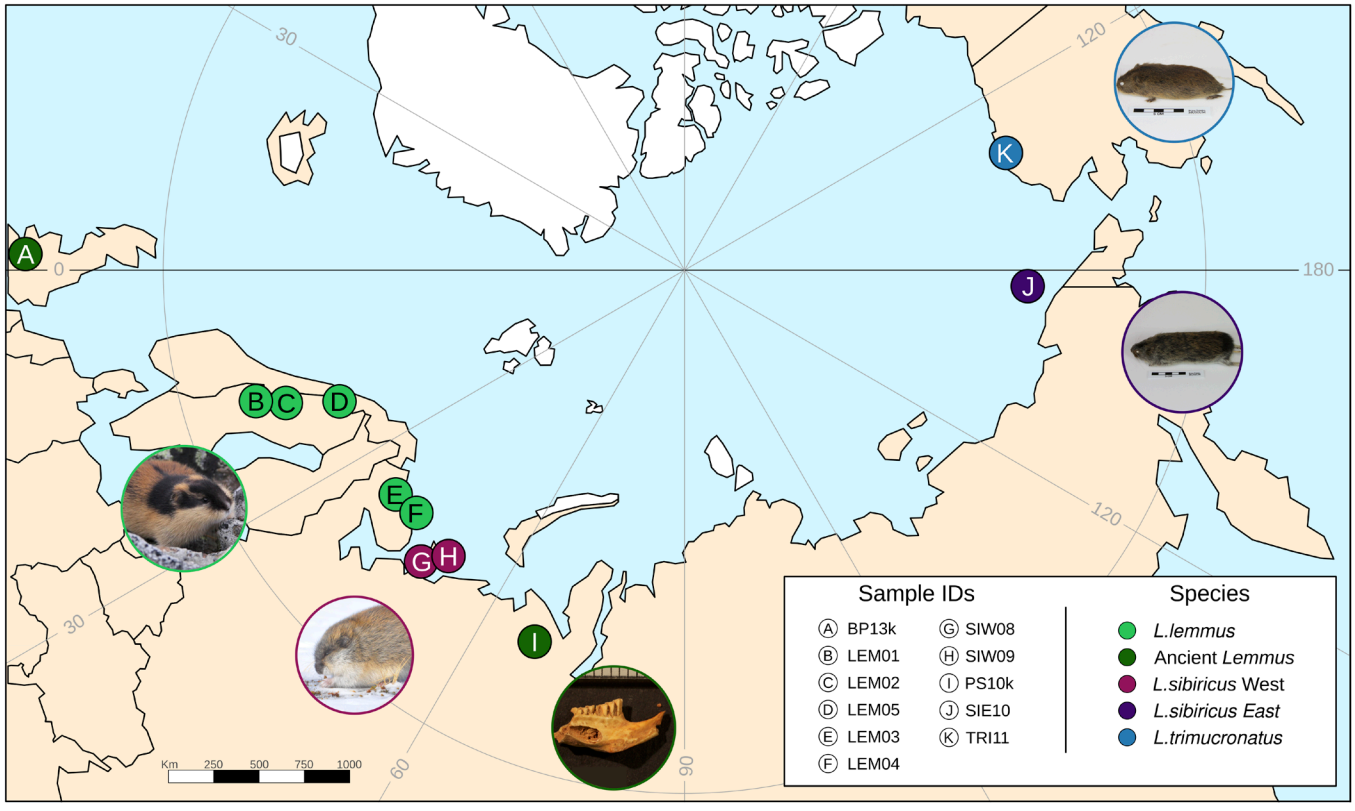


**Fig. 1.** Geographical sampling locations. Only samples for which we obtained genomic data are shown: the Norwegian lemming (*L. lemmus*; light green), ancient *Lemmus* (dark green), Western Siberian lemming (*L. sibiricus* West; pink), Eastern Siberian lemming (*L. sibiricus* East; purple), and the Nearctic Brown lemming (*L. trimucronatus*; blue). Sample locations (A to K) are based on geographic coordinates listed in *SI Appendix*, Table S1, or approximate locations.

well-supported monophyletic groups for each species/lineage (Fig. 2 and *SI Appendix*, Fig. S3). Interestingly, all phylogenies confirm the separation of *L. sibiricus* East and *L. sibiricus* West as distinct monophyletic lineages with *L. lemmus* samples forming a sister clade to *L. sibiricus* West. We also found this topology was consistent across the whole genome using nonoverlapping 1 Mb windows, with 99.95% of the trees showing the monophyletic structure observed in the mitochondrial and nuclear phylogenies (Fig. 3).

In both the full mitogenome and nuclear phylogenies, the ancient samples from Pymva Shor (PS10k) and Bridged Pot Cave (BP13k) are basal to and form a monophyletic clade with the modern *L. lemmus* samples (Fig. 2). This is consistent with the short mtDNA phylogeny which included a much more extensive sampling of ancient individuals (*SI Appendix*, Fig. S3). Within *L. lemmus*, principal component analyses separate the samples according to sampling location: Central Sweden, North Sweden, and Kola peninsula (*SI Appendix*, Fig. S4).

The species divergence estimates based on the full mitogenome phylogeny and pseudodiploid X chromosome coalescence are largely congruent (Table 1 and Fig. 4, and *SI Appendix*, Table S2). *L. trimucronatus* diverged from the other *Lemmus* species ~308 to 247 ka BP. Following this, *L. sibiricus* East diverged from the lineage leading to *L. sibiricus* West and *L. lemmus* around ~149 to 124 ka BP. Finally, we estimated a very shallow divergence time between *L. lemmus* and *L. sibiricus* West, occurring ~36.4 to 34 ka BP, just prior to the LGM. The time of the most recent common ancestor for the modern *L. lemmus* was ~6.4 ka BP based on our mitogenome estimates. However, when exploring other available mutation rates for rodents (mouse and microtine vole) the divergence times are younger, suggesting that the results presented here are likely an upper limit for the divergence times (*SI Appendix*, Fig. S5 and Table S3).

Finally, all species showed similar demographic trajectories with PSMC (*SI Appendix*, Fig. S6). The effective population size for all species peaked during the beginning of the Last Glacial Period, and decreased up to the onset of the Holocene, then remained stable until present day.

**Gene Flow Among Species.** We performed formal gene flow tests and did not detect evidence of excess allele sharing among any of the *Lemmus* species analyzed (Fig. 5). D-statistics for all comparisons involving *L. lemmus* as H1 and H2, and *L. sibiricus* West or East as H3, were not statistically different from zero (i.e., |Z|-score < 3), suggesting that no recent gene flow has occurred between *L. lemmus* and either of the *L. sibiricus*. In addition, all D-statistics were not significantly different from zero for the comparisons where H1 was *L. lemmus*, H2 was *L. sibiricus* West and H3 was *L. sibiricus* East, indicating no excess allele sharing between *L. sibiricus* West and *L. sibiricus* East.

**Unique Gene Variants in *L. lemmus*.** We investigated mutations unique to the *L. lemmus* lineage by identifying sites that are fixed for the derived allele in all five *L. lemmus* samples and fixed for the ancestral allele in all other *Lemmus* genomes, the collared lemming (*Dicrostonyx torquatus*), and the prairie vole (*Microtus ochrogaster*). We found 571 fixed-derived missense mutations and eight fixed-derived loss-of-function (LoF) mutations in our modern *L. lemmus* genomes. Both missense and LoF mutations are predicted to cause changes to protein structure and function. These mutations affected a total of 490 different genes (Dataset S1). Gene ontology enrichment analyses indicated an overrepresentation of genes involved in actomyosin structure organization, actin-filament processes, and cytoskeleton structure organization (*SI Appendix*, Table S4). We further explored the 17 genes with three or more nonsynonymous fixed-derived mutations in *L. lemmus* as well as all those containing LoF mutations (n = 8, Dataset S1 and *SI Appendix*, Fig. S7). These included genes involved in pigmentation (*LYST*), eye development and vision (*ATF6*, *NES*), fat transport (*APOB*, *BCO1*, *RELCH*), reproduction (*FSIP2*, *SPAG17, AOC1L3, ATF6*), neurological development (*NES*, *PIGN*, *SSPO, EPB41L3, AKNA*), and DNA repair (*TEX15*, *TREX1*, *TNKS1BP1*). Of all the fixed-derived nonsynonymous mutations in *L. lemmus*, only 57 sites in BP13k and nine sites in PS10k were covered by at least one read (*SI Appendix*, Table S5). We found that 12 (21%) and two (22%) sites in BP13k and PS10k, respectively, were also fixed for the *L. lemmus* derived allele.

We also investigated the presence of recent or ongoing selective sweeps across the *L. lemmus* genome using RAiSD (25). We identified 434 genes with evidence for a selective sweep (*SI Appendix*, Table S6 and Fig. S8). From the 255 genes that had IDs recognizable by gene ontology analysis with DAVID (26), we found

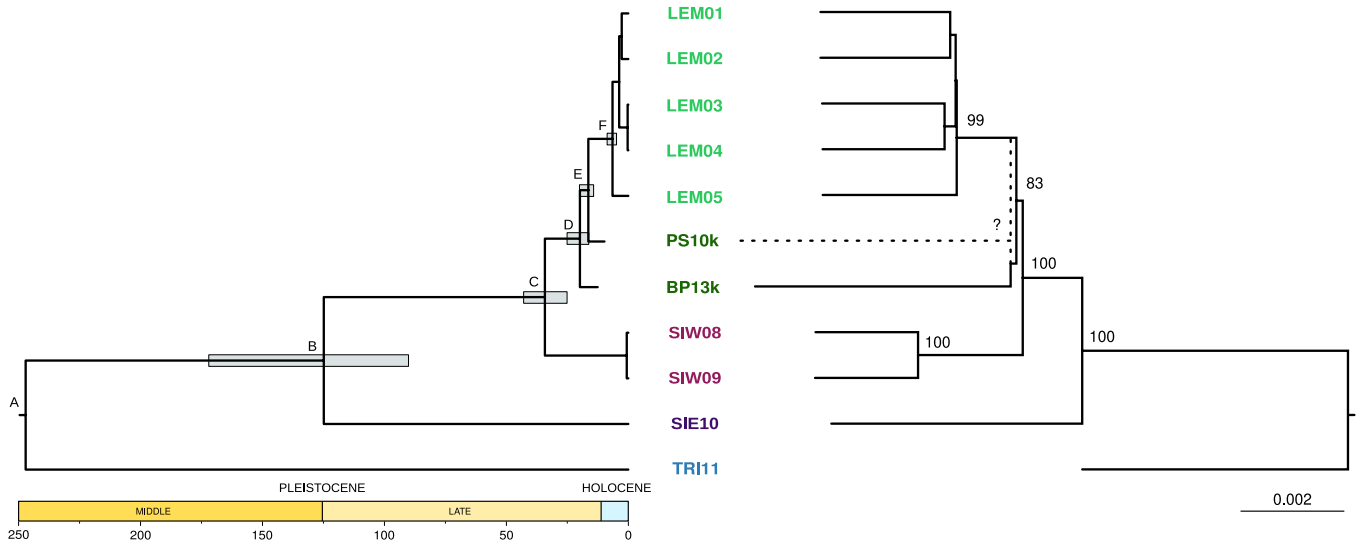


**Fig. 2.** Mitochondrial and nuclear phylogenetic trees for *Lemmus*. The mitochondrial tree (*Left*) is based on the complete mitogenomes of the 11 samples and obtained using BEAST. The nuclear phylogenetic tree (*Right*) was built using balanced minimum evolution (ME) based on Identity-By-State (IBS) matrices from the genomic data. Samples are as follows: *L. lemmus* (LEM01-05; light green), ancient *Lemmus* (PS10k, BP13k; dark green), *L. sibiricus* West (SIW08-09; pink), *L. sibiricus* East (SIE10; purple), *L. trimucronatus* (TRI11; blue). PS10k is represented by a dashed line in the nuclear phylogeny as there were not enough overlapping sites between the two ancient samples to include them in the same nuclear phylogeny.

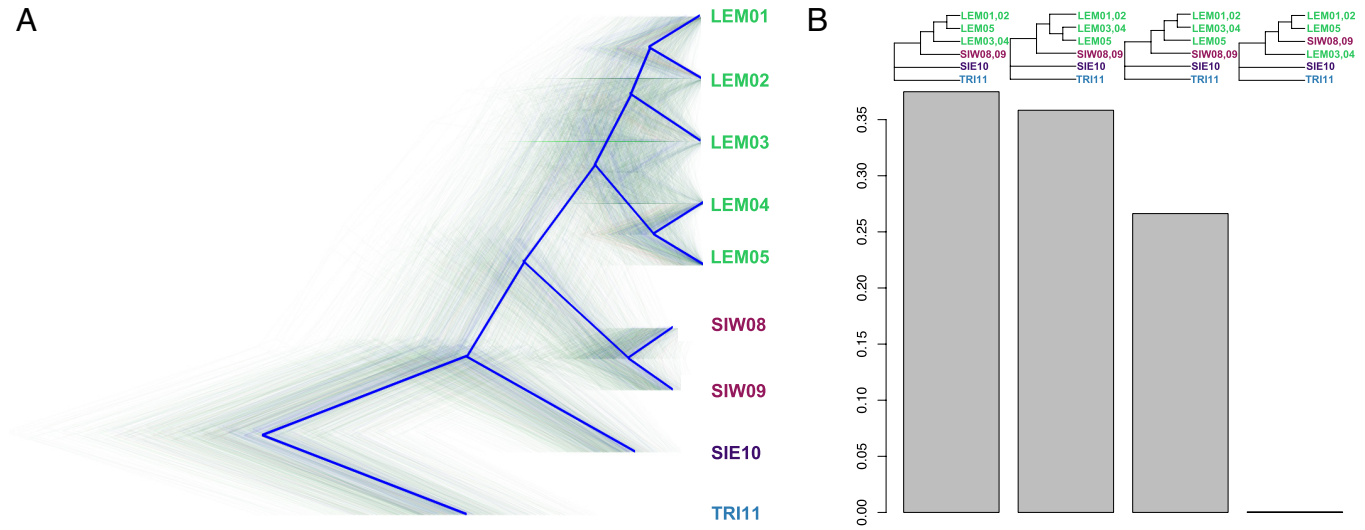


**Fig. 3.** Phylogenetic topology testing across the genome using 1 Mb windows. (*A*) Densitree representation of all trees, with the consensus tree drawn in blue. Samples are as follows: *L. lemmus* (LEM01-05; light green), *L. sibiricus* West (SIW08-09; pink), *L. sibiricus* East (SIE10; purple), *L. trimucronatus* (TRI11; blue). (*B*) Topology weighting for the top four topologies. Each bar represents the proportion of the given topology (drawn above) supported by the dataset. Each topology is drawn above the weighting, with coloration following *A*.

two main clusters of positively enriched genes, including olfactory processes (15 genes) and olfactory receptors (OR) (13 genes).

Finally, we compared the number of fixed-derived coding mutations in each species to determine whether *L. lemmus* displays an excess of nonsynonymous mutations. Our results indicate that *L. lemmus* has a similar rate of accumulation of nonsynonymous fixed-derived variants compared to the other three species (*SI Appendix*, Fig. S9).

## Discussion

**Origin and Isolation of the Norwegian Lemming.** The Norwegian lemming has a distinct coloration and behavioral patterns compared to the other *Lemmus* species, and has long been regarded as a distinct species from its Siberian counterpart. Our results from both the complete mitogenome and nuclear phylogenies are highly concordant and place *L. lemmus* and *L. sibiricus* West as two independent, monophyletic sister taxa (Figs. 2 and 3 and *SI Appendix*, Fig. S3). This is in agreement with previous studies based on short mitochondrial regions (17, 20, 21).

The divergence estimates for these two species (36.4 and 34 ka BP, based on the X chromosome and mitochondrial analyses, respectively; Table 1) suggest that the origin of *L. lemmus* happened just prior to the LGM, thus confirming that the Norwegian lemming has one of the most recent speciation times in mammals. Although we note that due to differences in the available mutation rates for rodents, the X chromosome estimate may represent an upper limit to the divergence time. Furthermore, as our analyses are based on mitochondrial and X chromosomes, future studies on the autosomal divergence times may differ due to sex-biased demographic factors (27), thus this will be important to test with a more extensive genomic dataset. As Norwegian lemmings can have one or two generations per year, depending on the length of the breeding season (28, 29), this results in a divergence time of ~34 to 68,000 generations ago. Other taxa suggested to be recently speciated include polar and brown bears, and Himalayan and Chinese takins, diverging ~34 to 48,000 and ~25,000 generations ago, respectively (7, 30, 31). Thus, *L. lemmus* are among the most recently speciated mammals. Even though our divergence estimates are younger than those suggested previously from short mitochondrial data (~75 ka BP) (21), all studies thus far confirm a divergence prior to the LGM (i.e., in Marine Isotope Stage 3) of *L. lemmus* and *L. sibiricus* West, suggesting the Last Glacial period and/or the LGM itself impacted the evolutionary history of *L. lemmus*. Importantly, this divergence estimate

**Table 1. Divergence estimates among lemming species/lineages**

| | Mitogenome | | Pseudodiploid X chromosome | |
|---|---|---|---|---|
| Node | Median node age (ka BP) | 95% HPD | Median node age (ka BP) | Range |
| A | 247 | 343 to 175 | 308 | 335 to 280 |
| B | 124 | 172 to 89 | 149 | 163 to 143 |
| C | 34 | 45 to 25 | 36.4 | 47 to 31 |
| D | 19 | 24 to 16 | | |
| E | 16 | 20 to 13 | | |
| F | 6.4 | 8.1 to 4.4 | | |
| **Mutation rate** | $2.75 \times 10^{-7}$ mutations/site/y | $3.63 \times 10^{-7}$ $1.85 \times 10^{-7}$ | $2.63 \times 10^{-9}$ mutations/site/generation | |

Node labels correspond to those from the mitogenomic phylogenetic tree in Fig. 2. For the mitogenome divergence, the median node age and 95% highest posterior density (HPD) were estimated in BEAST. The pseudodiploid X chromosome median node ages were estimated using the pairwise sequential Markovian coalescent (PSMC). The range for the pseudodiploid X chromosome divergence is taken as the minimum and maximum across all comparisons. Analyses were conducted using a generation time of two generations per year and the X chromosome scaled mutation rate ($2.63 \times 10^{-9}$ mutations/site/generation).

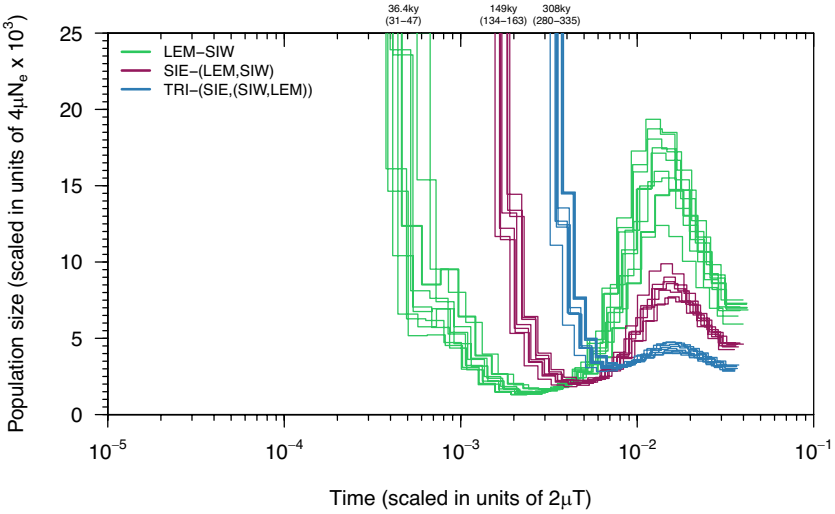


**Fig. 4.** Divergence estimates based on pseudodiploid X chromosomes and the PSMC. The divergence time of each pair of individuals is taken where the PSMC curve trends toward infinity (i.e., where the coalescence rate becomes zero). We estimated the divergence time between three sets of taxa: *L. lemmus* and *L. sibiricus* West (green), *L. sibiricus* East and *L. lemmus* or *L. sibiricus* West (pink), and *L. trimucronatus* and each of the other *Lemmus* lineages (purple). The lower x-axis shows time scaled by the mutation rate in generations, and generation time, with present to the left. The median and range for each set of the three estimated divergences are indicated on the upper x-axis and in Table 1. The y-axis shows the effective population scaled by the mutation rate.

between *L. lemmus* and *L. sibiricus* West is considerably younger than species-level splits in other small rodents, that occur in Marine Isotope Stage 5 (32).

Lemmings are known for migrating during population peaks (33), and they can travel long distances across ice with ease (34). It is thus possible that postspeciation gene flow could have occurred in areas where the distribution of *L. lemmus* and *L. sibiricus* West overlaps or comes in close contact. In fact, gene flow between these species was proposed to explain the occurrence of a specimen from Sirijorda Cave in Norway dating from 8 ka BP and having a *L. sibiricus* West mitochondrial haplotype (21). Currently, the regions where these species are in close contact are the Kola and Kanin peninsulas, where we sampled both *L. lemmus* and *L. sibiricus* West specimens for this study. However, none of our results, either using mitochondrial or nuclear phylogenies or explicit gene flow analyses (D-statistics), suggests the existence of gene flow between *L. sibiricus* West and *L. lemmus* (Figs. 2 and 4). This is surprising since gene flow between species with shallow divergence times is relatively common (11–14) and given the fact that *L. lemmus* and *L. sibiricus* West are known to be able to produce fertile offspring (35, 36).

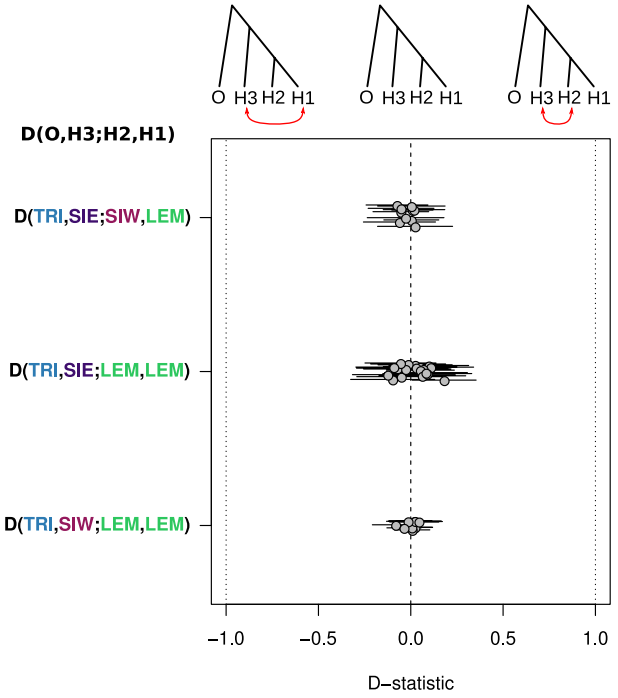


**Fig. 5.** D-statistic tests for excess allele sharing in *Lemmus*. The Y-axis shows the three sets of comparisons performed following the configuration D(O,H3;H2,H1). The red arrows indicate excess allele sharing between two samples in the test, i.e., a negative D-statistic indicates allele sharing between H1 and H3, whereas a positive D-statistic indicates allele sharing between H2 and H3. D-statistics of 0 indicate no excess of allele sharing.

The short mtDNA phylogeny indicates that ancient *Lemmus* samples from the Urals (Pymva Shor, 26 to 9 ka BP) are distributed between *L. sibiricus* West and clades basal to modern *L. lemmus* (*SI Appendix*, Fig. S3). This suggests that both species coexisted in the same area for a large portion of the Last Pleistocene, which might have provided opportunities for gene flow. However, these ancient *Lemmus* lineages seem to have not left descendants in the modern-day *L. lemmus* populations analyzed here. Additionally, we note that due to a lack of overlapping regions in the low coverage ancient samples, our D-statistic tests cannot rule out the possibility of gene flow between *L. sibiricus* West and the ancestral lineage of all the *L. lemmus* samples of this study. Since the most recent common ancestor of our *L. lemmus* samples estimated from the mitogenomic phylogeny was dated to 6.4 ka BP, we cannot exclude the possibility of gene flow prior to this time. The retrieval of higher coverage ancient DNA data, especially from areas where the species' ranges may have overlapped in the Late Pleistocene (e.g., in the Pymva Shor region) will be key to further testing postspeciation gene flow.

**Unique Gene Variants in the Norwegian Lemming.** To investigate the genomic basis of the unique phenotypic traits of *L. lemmus*, we identified genes containing nonsynonymous mutations fixed for the derived allele in all our *L. lemmus* samples but fixed for the ancestral allele in all the other lemmings, the collared lemming, and the prairie vole (Dataset S1). We caution that the gene functions discussed below are based on model organisms alone and note that functional validation would be needed to confirm the specific role of such mutations in lemmings.

We found fixed nonsynonymous mutations in one gene linked to coat color, *LYST* (37, 38), which may be involved in the distinct black and yellow coloration in *L. lemmus* compared to the other

*Lemmus* species. Mutations in the gene *LYST* are related to the truncation of melanosome and lysosome function, causing the expression of beige or pale colors in both mice and mink (37). It is possible that these changes in fur color phenotype differentiated *L. lemmus* from their conspecifics, perhaps contributing to avoiding hybridization with *L. sibiricus* West.

We found that all *L. lemmus* have four nonsynonymous mutations in *ATF6*, which is a transcription factor that is involved in the unfolded protein response to endoplasmic reticulum stress (39). In humans, *ATF6* is associated with cone photoreceptor development, and mutations in this gene can lead to color blindness (40). However, *ATF6* can also be involved in response to stress (41) and reducing spermatogenesis (42). Through its many different functions, this transcription factor is undoubtedly important for lemmings, with possible roles in altering color vision, responding to stressful environmental conditions, and male fertility.

We also identified mutations in several genes related to fat transport (*APOB*, *RELCH*, *BCO1*). In mice, *BCO1* knockout leads to a higher incidence of nonalcoholic fatty liver disease and upregulation of cholesterol transport markers (43). Interestingly, lemmings that have been exposed to stress (i.e., higher temperatures, a lack of proper nutrition, or too high population density) have increased levels of fatty liver disease (44, 45), suggesting that the fat metabolism of lemmings may be affected by stress. Moreover, energy production from fat likely aids in winter survival, as *L. lemmus* stays active throughout the winter.

The gene variants we observe in *L. lemmus* may have become fixed immediately after the divergence from *L. sibiricus* West some 36.4 to 34 ka BP, during the subsequent hypothesized isolation in the LGM (21), or following postglacial expansion. The two ancient samples predominantly have the ancestral allele at the positions where modern *L. lemmus* are fixed for the derived allele (*SI Appendix*, Table S5). This suggests that the majority of mutations that gave rise to the unique characteristics of *L. lemmus*, such as coat color, have evolved more recently, e.g., during isolation in a small refugium during the LGM or during the Holocene. Either of these scenarios would represent a very short time period for such a suite of variations to become fixed. This may also be explained by a regional rather than temporal effect and therefore genomic data from additional ancient *Lemmus* samples would be required to test these hypotheses more robustly.

Finally, although preliminary, our results suggest that there may be selective sweeps occurring in genes related to OR and olfactory processes in *L. lemmus*. OR are used to detect odorants in the environment (46) and represent the largest gene superfamily in vertebrates (47). Orthologous gene families have been found in most species, but rodents seem to be among the taxa with larger numbers (48). Since OR genes are an example of a fast evolving gene group in mammals (49), it is not surprising to find evidence of selective sweeps in such genes in lemmings as they are known to use olfactory clues for multiple social behaviors, such as discriminating between different species, social status, identifying sex, and determining the reproductive stage of other lemmings (50). However, we do note that the low sample size could affect the robustness of calculating the site frequency spectrum and linkage disequilibrium (51–53), and the demographic history of *L. lemmus* may confound the signal of a sweep, thus affecting our inferences. Nevertheless, our results may provide a starting point for evaluating the presence of selective sweeps in lemmings.

**Taxonomical Uncertainties of *L. sibiricus*.** Both the mitochondrial and nuclear phylogenetic analyses confirm that *L. lemmus* and *L. sibiricus* West are monophyletic sister taxa to the exclusion of *L. sibiricus* East. This has implications for the taxonomic status of *L. sibiricus*. Previous studies have suggested that *L. sibiricus* encompasses two distinct lineages (17), or alternatively that these lineages represent distinct species—*L. sibiricus, Lemmus bungei* (16), and *Lemmus ognevi* (20). A study on cytochrome B data suggested that *L. sibiricus* East is a sister taxon to *Lemmus amurensis* (unsampled in this study), and paraphyletic to *L. sibiricus* West (20). Our results confirm the paraphyly of *L. sibiricus* West and *L. sibiricus* East, suggesting they diverged ~130 ka BP, over fourfold older than the divergence between *L. sibiricus* West and *L. lemmus*. Moreover, we find no evidence for gene flow between *L. sibiricus* West and *L. sibiricus* East. Overall, the phylogenetic placement and divergence between the two *L. sibiricus* lineages supports the distinction of *L. sibiricus* East as a separate species, *Lemmus paulus*, Allen, 1914. We note that *Lemmus paulus*, *Lemmus portenkoi, L. bungei,* and *L. ognevi* are all synonyms previously used to refer to Siberian lemmings found in the eastern part of their range, though Allen (54) was the first to describe the central Siberian lemmings as distinct. Our *L. sibiricus* East sample, the first genomic data from the species, was collected on Wrangel Island, and it is possible that this genome represents an isolated, divergent lineage within *L. sibiricus* East. As previously suggested (17), further sampling along the geographic boundary between *L. sibiricus* West and East (e.g. the Verkhoyansk Range) can help determine whether these species remain genetically distinct also in areas where their range distributions overlap.

## Conclusion

In this study, we have generated a draft genome assembly for the Norwegian lemming, bringing true lemmings into the genomic era. Our data show that the Norwegian lemming forms a distinct monophyletic clade within the *Lemmus* phylogeny, that it has unique gene variants likely involved in pigmentation and fat transport, and that there is a lack of gene flow with the western Siberian lemming. Future studies generating higher coverage ancient genomes from *L. lemmus* across their Late Pleistocene range may determine whether any postspeciation gene flow occurred outside of Scandinavia in the Late Pleistocene, the geographic origin of the present-day *L. lemmus* lineage, and define when the unique gene variants arose along the *L. lemmus* lineage. Finally, the placement of *L. lemmus* and *L. sibiricus* West as sister species has important taxonomic implications for the distinction of the Eastern and Western Siberian lemmings as different species.

## Methods

**Sampling.** For whole genome analysis, nine tissue samples were collected from *Lemmus* specimens collected in Sweden, Russia, and the United States, corresponding to four different species/lineages: *L. lemmus*, *L. sibiricus* West clade, *L. sibiricus* East clade, and *L. trimucronatus* (Fig. 1 and *SI Appendix*, Table S1). In addition, bones from two ancient lemming specimens were collected from Russia (Pymva Shor) and the United Kingdom (Bridged Pot Cave). The ancient samples were dated at the Oxford Radiocarbon Dating Unit and dates were calibrated using OxCal v4.3 (55) with the IntCal20 curve (56). All ancient laboratory work was performed following standard ancient DNA procedures in a dedicated facility at the Swedish Museum of Natural History, Stockholm.

For a broader view of the *Lemmus* phylogeny, we also analyzed an additional 55 ancient samples for variation in a short mitochondrial DNA fragment. These samples were collected from across the Late Pleistocene range of *Lemmus* spp., namely from Trou Al'Wesse and Marie Jeanne Cave in Belgium as well as Pymva Shor in Russia. These samples were not directly radiocarbon dated and estimated ages were taken from the literature (*SI Appendix*, Table S7). Laboratory methods relating to these samples can be found in *SI Appendix*.

**DNA Extraction and Sequencing.** For the de novo assembly we extracted the DNA of a piece of muscle tissue from a sample (LEM05) collected in northernmost Sweden using a Thermo Scientific KingFisher Duo magnetic particle processor and the KingFisher Cell and Tissue DNA Kit (ThermoFisher Scientific, MA). We then constructed five different libraries: two libraries, one each with 180 and 650 bp inserts, using a TruSeq DNA kit (Illumina, CA) on an Agilent NGS workstation (Agilent, CA) and three mate-pair libraries with the Nextera gel-plus protocol (Illumina, CA) using a Blue Pippin (Sage Science Inc., MA) for size selecting target fragments: two 3 to 5 kb mate-pair short (MPS) libraries and one 5 to 8 kb mate-pair long (MPL) library using the 0.75% 1 to 10 kb Gel Cassette with Marker S1. We indexed these libraries and sequenced them on an Illumina HiSeq 2500 in high output mode with a paired-end 2x125 bp setup, with the exception of the MPL library which was run on an Illumina HiSeq X with a 2x150 bp setup. In total, the 180 bp and 650 bp insert libraries were sequenced on 1 lane each and the MPS and MPL libraries on 0.67 of a lane each. All sequencing was performed at the Swedish National Genomics Infrastructure (NGI Stockholm).

DNA extraction for the other eight modern samples was performed using the KingFisher Cell and Tissue DNA Kit on the Kingfisher DNA extraction robot (Thermo Fisher Scientific, MA). We then built TruSeq PCR-free libraries (Illumina, CA) and sequenced them on a total of four lanes of an Illumina HiSeqX v2.5 platform using a paired-end 2x150 bp setup at NGI Stockholm.

For two ancient samples (BP13k and PS10k), we first obtained ca. 50 mg of bone powder using a hand-held Dremel drill and then extracted the DNA using a silica-column method (57, 58). DNA extracts were treated with USER enzyme to cleave erroneous bases caused by postmortem damage (59, 60). We then built double-indexed genomic libraries (61) and sequenced them on two lanes of an Illumina HiSeqX v2.5 platform using paired-end 2x150bp setup at NGI Stockholm. All presequencing laboratory steps were carried out in a dedicated ancient DNA cleanroom laboratory.

**Bioinformatic Processing of Mitochondrial DNA Data.** We first reconstructed a reference mitogenome for *Lemmus* applying the iterative mapping process implemented in MITObim (62). We used the raw data from the sample used to generate the de novo genome assembly (LEM05), the mitogenome of a *D. torquatus* (NC_034646.1) as initial bait, and the –quick option with default values. The resulting mitogenome was used as reference to map the data from the rest of the samples.

For the nine modern samples, we first trimmed the sequencing adapters from the raw reads using Trimmomatic v0.32 and mapped them to the MITObim-reconstructed mitogenome using BWA mem (63). We then generated consensus sequences in Geneious (64) using a majority rule and a minimum depth threshold of 5. For the two ancient samples, we first trimmed sequencing adapters and merged paired-end reads using SeqPrep v1.1 (github.com/jstjohn/SeqPrep) with a minor modification to estimate the base qualities in the merged region (60). We subsequently mapped the merged reads to the MITObim-reconstructed mitogenome using BWA aln v0.7.17 (65) with settings specific for ancient DNA (66). We removed PCR duplicates using a python script that takes into account both the start and end coordinates of the alignments (github.com/pontussk/samremovedup). Finally, we generated consensus sequences in Geneious as above, but with a depth threshold of 3.

**Genome Assembly and Bioinformatic Processing of Nuclear DNA Data.** For the de novo genome assembly, we first used Trimmomatic (67) to remove adapters and low-quality sequences and then assembled the data using ALLPATHS-LG r.52485 (68) with the option "HAPLOIDIFY = True." The quality of the assembly was evaluated with QUAST v4.5.4 (69) and BUSCO v3.0.2 (70) using the "mammalia_odb9" dataset.

We processed the resequencing data from all modern samples, including the data generated in the paired-end short-insert (180 bp) libraries of LEM05, using the "modern track" of an in-development version of the GenErode pipeline (71), with the following modifications: trimming was performed using Trimmomatic. Briefly, we first trimmed the raw reads as above and mapped them to the de novo reference genome using BWA mem. We then used samtools v1.8 (72) to sort and index the alignments as well as to remove duplicates and GATK IndelRealigner v3.4.0 (73) to realign the reads mapped around indels. We additionally mapped the raw reads of our samples, plus the collared lemming [*D. torquatus*; (74)] to the prairie vole reference genome (*M. ochrogaster*, GCA_000317375) in order to perform selection analyses. Finally, for the two ancient samples (PS10k and BP13k), we mapped the raw data to the de novo generated reference *L. lemmus* genome using the "historical track" of GenErode, where trimming was performed using a modified version of SeqPrep (https://github.com/jstjohn/SeqPrep).

**Chromosome X Scaffolds.** For the modern samples, we identified the scaffolds potentially linked to the chromosome X of the *L. lemmus* de novo reference genome using their difference in relative coverage in males and females (75), which in turn, can be used to reliably estimate the sex of each sample from the genomic data. Briefly, we estimated which scaffolds larger than 10 kb displayed approximately half of the coverage than the genomic average in the male *Lemmus* samples but ca. the average in the female ones (*SI Appendix*, Fig. S2). For all samples, we also compared the number of reads confidently mapping to chromosome X in the prairie vole genome with the number of reads mapping to chromosome 16, which has a similar length (66).

**Repeat Content, Variant Calling, and Filtering.** We first used RepeatMasker and RepeatModeler to identify repetitive regions in the de novo assembly of *L. lemmus* (76, 77). We then genotyped variants on all the modern genomes using bcftools v1.8 (78). Variant calls were subsequently filtered by excluding those within 5 bp of indels, with genotype quality < 30 and depth < 10, as well as those heterozygote variants for which the ratio of reads with the reference allele divided by depth was below 0.2 or above 0.8. We additionally excluded all variants allocated in the previously identified repetitive regions, and in the scaffolds identified as part of chromosome X.

**Phylogenetic Analyses.** We built phylogenetic trees based on three different sets of data and samples: short ancient mitochondrial sequences (n = 131; *SI Appendix*), complete mitogenomes of modern and ancient specimens (n = 11), and nuclear genomic data from modern and ancient specimens (n = 11).

For the whole mitogenome phylogenetic analyses, we determined GTR+G as the best-fitting nucleotide substitution model using Jmodeltest v2.10.1 (79). We then constructed a mitochondrial phylogeny using BEAST v1.10.4 (80). Mean radiocarbon dates were specified as tip dates for the two ancient samples using a normal prior (3). We used a strict clock and a constant coalescent tree model. The analysis was run for 50 million generations, with sampling every 1,000 generations. The run was checked for convergence (ESS values > 200) in Tracer v1.7 (81), and 10% burn-in was removed with TreeAnnotator. The consensus tree was visualized in Figtree v4.4.4 (82).

We constructed nuclear phylogenies based on the whole genome Identity-By-State (IBS) matrices for all individuals with available genomic data. We used ANGSD –doIBS (83), which selects one allele per site for each sample based on a randomly sampled high-quality sequencing read. The phylogenetic trees were inferred using the balanced minimum evolution (ME) method in FastME v2 (84) and the robustness of each node was estimated from 100 resampling replicates based on 100,000 sites each. The two ancient samples for which there was genomic data available were run separately in the nuclear tree as there were not enough overlapping sites between them to include both at the same time.

We further investigated the robustness of the topology building phylogenetic trees based on 1 Mb genomic windows (85). Given the low number of overlapping sites in the ancient samples, these were excluded from this analysis. First, individual FASTA files were produced from the BAM file using ANGSD, with the following options: –dofasta 3, –docounts 1, –setmaxdepth 50. Each scaffold over 1 Mb in the reference genome was divided into 1 Mb nonoverlapping windows. We then used the getfasta –split option in Bedtools (86) to split the individual FASTA files into the 1 Mb windows. Next, the individual sample FASTA files for each window were merged and RAXML v8.2.12-gcc (87) was used to generate a tree for each window. All tree files were concatenated and visualized using Densitree (88). We used Twisst (89) to assess the fraction of the different topologies represented by the obtained trees. The analysis was repeated using 100 kb windows, with every 10th window selected (*SI Appendix*, Fig. S10).

**Estimating Divergence Through Demography.** In addition to divergence times estimated using the mitochondrial dataset, we assessed divergence between the nuclear genomes using pseudodiploid X chromosomes from the modern individuals (for both males and females) of each species, i.e., the five *L. lemmus*, the two *L. sibiricus* West, the one *L. sibiricus* East and the one *L. trimucronatus* individual. We first generated haploidized X chromosome sequences for

each individual using samtools view, samtools mpileup, and the pu2fa script from Chrom-Compare (github.com/Paleogenomics/Chrom-Compare), and filtered for minimum mapping and base quality of 30 and sites with less than 5X and more than 100X coverage. Next, we merged each pair of haploid sequences from the different species into one pseudodiploid X chromosome sequence using seqtk mergefa (github.com/lh3/seqtk) (60, 90). We ran PSMC on the pseudodiploid X chromosomes with default parameters (64 discrete intervals: "4 + 25 * 2 + 4 + 6") (27). We used a generation time of 0.5 y, as lemmings can produce two generations during the increasing phases of the population cycle (28, 29). We estimated the X chromosome mutation rate from the autosomal rate [$2.96 \times 10^{-9}$ substitutions per site per generation, (91)], using the calculation outlined in ref. 27, where alpha = 2, giving a rate of $2.631 \times 10^{-9}$ substitutions/site/generation. To account for the different population size of the X chromosome, we scaled population size by 0.75. The divergence time is taken where the PSMC plot trends toward infinity (i.e., $N_e$ = 1/coalescent rate, where the coalescence rate is zero). We repeated the analysis with X chromosome scaled mutation rates for the mouse (*Mus musculus*, X chromosome mutation rate = $4.79 \times 10^{-9}$ substitutions/site/y) (92) and common vole (*Microtus arvalis*, X chromosome mutation rate = $7.73 \times 10^{-9}$ substitutions/site/generation) (93) (*SI Appendix*, Table S3).

**Gene Flow.** We investigated the presence of gene flow between the different species of lemmings using D-statistics, also known as ABBA-BABA tests (94, 95), using popstats (96). We performed tests of the form D(O, H3; H2, H1), where O is an outgroup sample (TRI11); H3, a possible donor sample; and H2 and H1 two samples that could have received gene flow from H3. We tested all combinations of samples for four different tests: 1) gene flow from *L. sibiricus* East into *L. lemmus* [D(TRI11, SIE10; LEM, LEM)], 2) from *L. sibiricus* West into *L. lemmus* [D(TRI11, SIW; LEM, LEM)], and 3) from *L. sibiricus* East into *L. sibiricus* West [D(TRI11, SIE10; SIW, LEM)]. All tests were performed in the filtered variants by sampling one random allele per site. Significant deviations from D = 0 were assessed using Z-scores based on a jackknife of 100 Kb of block size. The two ancient samples (BP13k and PS10k) were excluded from these analyses due to having too few sites to perform robust jackknifing.

**Unique Gene Variants.** To look at species-specific mutations in *L. lemmus*, we first identified all variants that were fixed for the derived variant in our modern *L. lemmus* samples (n = 5), and fixed for the ancestral allele all other *Lemmus* samples (n = 4), the collared lemming (n = 1), and the prairie vole reference genome, using a custom python script. We used SnpEff (97) to annotate all mutations in coding regions into different impact categories based on predicted changes to protein structure. We then used bedtools intersect to find all the fixed derived mutations in the annotated VCF and used SnpSift (97) to produce a text file with the position, predicted impact (high, moderate, low, or modifier), and gene name. We counted the number of fixed-derived mutations in each SnpEff impact class and focused on genes with three or more mutations in the moderate class (variants that might change the protein function but are not disruptive [missense]), as well as all genes with mutations in the high impact class (variants that have disruptive effects on the protein causing truncation and loss of function [LoF], such as stop gain and frameshift variants). We assessed which functional gene ontology (GO) categories are enriched in these variants using GOrilla (98) and the annotation of *Rattus norvegicus* as a reference set. We then used bam-readcount (github.com/genome/bam-readcount) to determine the allele state in the two ancient samples (BP13k and PS10k) for each of the previously identified *L. lemmus* fixed-derived nonsynonymous mutations.

Finally, we evaluated whether any of the *Lemmus* species had an excess of nonsynonymous derived mutations relative to synonymous derived mutations and to the other species. We counted the number of homozygous nonsynonymous variants per sample which are homozygous for the ancestral allele in a random sample from each of the other three species and the prairie vole reference genome. We repeated this for all combinations of samples in the dataset (i.e., 10 times per sample) and normalized the values using the number of homozygous derived synonymous mutations calculated in the same way.

**Selective Sweep Analysis.** We performed a preliminary analysis to detect genome regions under recent selective pressure using RAiSD (25). It is based on a composite likelihood rate test that relies on multiple signatures of a selective sweep, capturing multiple sweep signatures at once. It includes changes in the site frequency spectrum, linkage disequilibrium patterns, and the spatial distribution of genetic diversity. This method is SNP-driven, and is performed over 50 bp sliding-windows. We used a 99.95% threshold to determine regions under positive selection. We identified all gene Ensembl gene IDs in these regions, and submitted the ID list to DAVID's Gene Functional Classification Tool (99). This tool generates a similarity matrix based on shared functional annotation. We ran the tool with a high stringency classification.

**Data, Materials, and Software Availability.** Data associated with this project is available on the European Nucleotide Archive Project: PRJEB87511 (100).

**ACKNOWLEDGMENTS.** D.D-d-M. acknowledges funding from the Carl Tryggers Foundation (CTS 17:109). L.D. acknowledges funding from Vetenskapsrådet (Projects: 2020-06174, 2020-04808, and 2021-00625), and FORMAS (2018-01640) and the European Union (ERC, PrimiGenomes, 101054984). L.D. and E.L. acknowledge funding from the Bolin Centre for Climate Research. V.K.L. acknowledges funding from the Swedish Royal Academy of Sciences, Riksmusei Vänner, and Längmanska Kulturföreningen. V.B.F. and A.V.G. acknowledge funding from the NIH (P20GM130443 and P20GM103395). We thank Lars Olsson from the County Administration Board of Norbotten for providing sample LEM05, and Håkan Berglund for their involvement in sample collection. A. Taylor (Archaeology and Anthropology Museum at Cambridge University, UK) is acknowledged for giving access to the samples from Bridged Pot. We acknowledge Becky Miller, the director of the Trou Al'Wesse excavation, and funding for the project provided annually by the Ministère de la Région wallonne, Service Public de Wallonie. We acknowledge the AWaP (Agence Wallonne du Patrimoine) as the main funding institution for the work at Trou Al'Wesse. This manuscript is dedicated in memory of Pierre Noiret who sadly passed away in April 2025. We acknowledge support from the Uppsala Multidisciplinary Centre for Advanced Computational Science for assistance with massively parallel sequencing and access to the UPPMAX computational infrastructure, and resources provided by Swedish National Infrastructure for Computing (SNIC) at Uppsala partially funded by the Swedish Research Council through Grant Agreement No. 2018-05973.

Author affiliations: [a]Centre for Palaeogenetics, Stockholm 10691, Sweden; [b]Department of Zoology, Stockholm University, Stockholm 10691, Sweden; [c]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm 11418, Sweden; [d]National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 752 37, Sweden; [e]Department of Archaeology and Classical Studies, Stockholm University, Stockholm 10691, Sweden; [f]Tree of Life Programme, Wellcome Sanger Institute, Hinxton CB10 1RQ, United Kingdom; [g]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Solna 17121, Sweden; [h]Natural History Museum Vienna, Central Research Laboratories, Vienna 1010, Austria; [i]Department of Cognitive Biology, University of Vienna, Vienna 1090, Austria; [j]Konrad Lorenz Institute of Ethology, Vienna 1160, Austria; [k]South African National Biodiversity Institute, National Zoological Garden, Pretoria 0184, South Africa; [l]Institute of Plant and Animal Ecology, Ural Branch of the Russian Academy of Sciences, Ekaterinburg 620144, Russia; [m]Faculty of Science and Technology, Bournemouth University, Talbot Campus, Poole BH12 5BB, United Kingdom; [n]Department of Archaeology, Anthropology and Geography, University of Winchester, Winchester SO22 4NR, United Kingdom; [o]Service de Préhistoire, Université de Liège, Liège 4000, Belgium; [p]Operational Direction "Earth and History of Life", Royal Belgian Institute of Natural Sciences, Brussels 1000, Belgium; [q]Department of Arctic and Marine Biology, Faculty of Biosciences, Fisheries and Economics University of Tromsø The Arctic University of Norway, Tromsø 9019, Norway; [r]Max Planck Institute of Animal Behaviour, Radolfzell am Bodensee 78315 Germany; and [s]Institute of Arctic Biology, University of Alaska, Fairbanks, AK 99775-7000

1. G. M. Hewitt, Some genetic consequences of ice ages, and their role in divergence and speciation. *Biol. J. Linn. Soc. Lond.* **58**, 247–276 (1996).
2. M. de Manuel *et al.*, The evolutionary history of extinct and living lions. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10927–10934 (2020).
3. D. W. G. Stanton *et al.*, Early pleistocene origin and extensive intra-species diversity of the extinct cave lion. *Sci. Rep.* **10**, 12621 (2020).
4. J. A. Cahill, R. E. Green, T. L. Fulton, M. Stiller, F. Jay, Genomic evidence for island population conversion resolves conflicting theories of polar bear evolution. *PLoS Genet.* **9**, 1003345 (2013).

5. M. Baca *et al.*, The Tien Shan vole (*Microtus ilaeus*; Rodentia: Cricetidae) as a new species in the Late Pleistocene of Europe. *Ecol. Evol.* **11**, 16113–16125 (2021).
6. J. T. Weir, D. Schluter, Ice sheets promote speciation in boreal birds. *Proc. Biol. Sci.* **271**, 1881–1887 (2004).
7. S. Liu *et al.*, Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell* **157**, 785–794 (2014).
8. J. R. Stewart, A. M. Lister, I. Barnes, L. Dalén, Refugia revisited: Individualistic responses of species in space and time. *Proc. Biol. Sci.* **277**, 661–671 (2010).
9. C. C. R. Hansen *et al.*, The muskox lost a substantial part of its genetic diversity on its long road to Greenland. *Curr. Biol.* **28**, 4022–4028.e5 (2018).
10. S. Liu *et al.*, Ancient and modern genomes unravel the evolutionary history of the rhinoceros family. *Cell* **184**, 4874–4885 (2021).
11. H. V. Figueiró *et al.*, Genome-wide signatures of complex introgression and adaptive evolution in the big cats. *Sci. Adv.* **3**, e1700299 (2017).
12. J. A. Cahill *et al.*, Genomic evidence of widespread admixture from polar bears into brown bears during the last ice age. *Mol. Biol. Evol.* **35**, 1120–1129 (2018).
13. A. Barlow *et al.*, Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**, 1563–1570 (2018).
14. A. T. Salis *et al.*, Ancient genomes reveal hybridisation between extinct short-faced bears and the extant spectacled bear (*Tremarctos ornatus*). bioRxiv [Preprint] (2021). https://doi.org/10.1101/2021.02.05.429853 (Accessed 12 May 2022).
15. J. Ottenburghs *et al.*, Avian introgression in the genomic era. *Avian Res.* **8**, 1–11 (2017).
16. K. Fredga, V. Fedorov, G. Jarrell, L. Jonsson, Genetic diversity arctic lemmings. *AMBIO* **28**, 261–269 (1999).
17. V. Fedorov, A. Goropashnaya, G. H. Jarrell, K. Fredga, Phylogeographic structure and mitochondrial DNA variation in true lemmings (*Lemmus*) from the Eurasian Arctic. *Biol. J. Linn. Soc. Lond.* **66**, 357–371 (1999).
18. G. H. Jarrell, K. Fredga, "How many kinds of lemmings? A taxonomic overview" in *The Biology of Lemmings,* N. C. Stenseth, R. A. Ims, Eds. (Academic Press, London, UK, 1993).
19. V. B. Fedorov, A. V. Goropashnaya, M. Jaarola, J. A. Cook, Phylogeography of lemmings (*Lemmus*): No evidence for postglacial colonization of Arctic from the Beringian refugium. *Mol. Ecol.* **12**, 725–731 (2003).
20. N. I. Abramson, T. V. Petrova, Genetic analysis of type material of the Amur lemming resolves nomenclature issues and creates challenges for the taxonomy of true lemmings (*Lemmus*, Rodentia: Cricetidae) in the eastern Palearctic. *Zool. J. Linn. Soc.* **182**, 465–477 (2018).
21. V. K. Lagerholm *et al.*, On the origin of the Norwegian lemming. *Mol. Ecol.* **23**, 2060–2071 (2014).
22. V. Fedorov, N. C. Stenseth, Glacial survival of the Norwegian lemming (*Lemmus lemmus*) in Scandinavia: Inference from mitochondrial DNA variation. *Proc. Biol. Sci.* **268**, 809–814 (2001).
23. M. Andersson, Aposematism and crypsis in a rodent: Antipredator defence of the Norwegian lemming. *Behav. Ecol. Sociobiol.* **69**, 571–581 (2015).
24. N. C. Stenseth, R. A. Ims, "Intra-and interspecific relations: An introduction" in *The Biology of Lemmings*, N. C. Stenseth, R. A. Ims, Eds. (Academic Press, London, UK, 1993), pp. 341–354.
25. N. Alachiotis, P. Pavlidis, RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun. Biol.* **1**, 79 (2018).
26. D. W. Huang *et al.*, The DAVID gene functional classification tool: A novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol.* **8**, R183 (2007).
27. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
28. S. Erlinge, D. Hasselquist, M. Svensson, P. Frodin, P. Nilsson, Reproductive behaviour of female Siberian lemmings during the increase and peak phase of the lemming cycle. *Oecologia* **123**, 200–207 (2000).
29. T. V. Koshkina, A. S. Khalanski, Reproduction of the Norway lemming (*Lemmus lemmus* L.) on the Kola Peninsula. *Zool. Zhur.* 41, 604–615 (1962).
30. D. Hirata *et al.*, Paternal phylogeographic structure of the brown bear (*Ursus arctos*) in northeastern Asia and the effect of male-mediated gene flow to insular populations. *Zool. Lett.* **3**, 21 (2017).
31. L. Yang *et al.*, Evolutionary conservation genomics reveals recent speciation and local adaptation in threatened takins. *Mol. Biol. Evol.* **39**, msac111 (2022).
32. M. Baca *et al.*, Ancient DNA of narrow-headed vole reveal common features of the late pleistocene population dynamics in cold-adapted small mammals. *Proc. Biol. Sci.* **290**, 20222238 (2023).
33. A. Kaikusalo, H. Henttonen, "Lemming movements" in *The Biology of Lemmings*, N. C. Stenseth, R. A. Ims, Eds. (Academic Press, London, UK, 1993), pp. 157–186.
34. L. L. de Kock, A. E. Robinson, Observations on a lemming movement in Jamtland, Sweden, in autumn 1963. *J. Mammal.* **47**, 490–499 (1966).
35. A. V. Pokrovskiy, I. A. Kuznetsova, M. I. Cheprakov, Hybridological studies of reproductive isolation of Palaearctic species of genus *Lemmus* (Rodentia, Cricetidae). *Zool. Zh.* **63**, 904–911 (1984).
36. E. A. Gileva, I. A. Kuznetsova, M. I. Cheprakov, Chromosome sets and taxonomy of the true lemmings (genus *Lemmus*). *Zool. Zh.* **63**, 105–114 (1984).
37. I. J. Jackson, Homologous pigmentation mutations in human, mouse and other model organisms. *Hum. Mol. Genet.* **6**, 1613–1624 (1997).
38. B. McEvoy, S. Beleza, M. D. Shriver, The genetic architecture of normal variation in human pigmentation: An evolutionary perspective and model. *Hum. Mol. Genet.* **2**, R176–R181 (2006).
39. P. Walter, D. Ron, The unfolded protein response: From stress pathway to homeostatic regulation. *Science* **334**, 1081–1086 (2011).
40. H. Kroeger *et al.*, ATF6 is essential for human cone photoreceptor development. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103196118 (2021).
41. R. F. Hillary, U. FitzGerald, A lifetime of stress: ATF6 in development and homeostasis. *J. Biomed. Sci.* **25**, 48 (2018).
42. R. Yu *et al.*, ATF6 deficiency damages the development of spermatogenesis in male Atf6 knockout mice. *Andrologia* **54**, e14350 (2022).
43. J. Y. Lim, C. Liu, K.-Q. Hu, D. E. Smith, X.-D. Wang, Ablation of carotenoid cleavage enzymes (BCO1 and BCO2) induced hepatic steatosis by altering the farnesoid X receptor/miR-34a/sirtuin 1 pathway. *Arch. Biochem. Biophys.* **654**, 1–9 (2018).
44. R. L. Rausch, V. R. Rausch, Taxonomy and Zoogeography of Lemmus spp. (Rodentia: Arvicolinae), with Notes on Laboratory-Reared Lemmings. *Faculty Publications from the Harold W. Manter Laboratory of Parasitology* 859 (1975).
45. A. Semb-Johansson, R. Wiger, C. E. Engh, Dynamics of freely growing, confined populations of the Norwegian Lemming, *Lemmus lemmus*. *Oikos* **33**, 246–260 (1979).
46. P. A. Godfrey, B. Malnic, L. B. Buck, The mouse olfactory receptor gene family. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2156–2161 (2004).
47. X. Zhang, S. Firestein, The olfactory receptor gene superfamily of the mouse. *Nat. Neurosci.* **5**, 124–133 (2002).
48. J. M. Young *et al.*, Different evolutionary processes shaped the mouse and human olfactory receptor gene families. *Hum. Mol. Genet.* **11**, 535–546 (2002).
49. Y. Niimura, A. Matsui, K. Touhara, Extreme expansion of the olfactory receptor gene repertoire in African elephants and evolutionary dynamics of orthologous gene groups in 13 placental mammals. *Genome Res.* **24**, 1485–1496 (2014).
50. U. W. Huck, E. M. Banks, Social olfaction in male brown lemmings (*Lemmus sibiricus* = trimucronatus) and collared lemmings (*Dicrostonyx groenlandicus*): I. Discrimination of species, sex, and estrous condition. *J. Comp. Psychol.* **98**, 54–59 (1984).
51. P. Pavlidis, D. Živkovic, A. Stamatakis, N. Alachiotis, SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
52. M. Fagny *et al.*, Exploring the occurrence of classic selective sweeps in humans using whole-genome sequencing data sets. *Mol. Biol. Evol.* **31**, 1850–1868 (2014).
53. J. K. Pickrell *et al.*, Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
54. G. M. Allen, Mammals., Notes on the birds and mammals of the Arctic coast of east Siberia. *Proc. N. Engl. Zoöl. Club* **5**, 49–66 (1914).
55. C. B. Ramsey, Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337–360 (2009).
56. P. J. Reimer *et al.*, The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**, 725–757 (2020).
57. D. Y. Yang, B. Eng, J. S. Waye, J. C. Dudar, S. R. Saunders, Technical Note: Improved DNA extraction from ancient bones using silica-based spin columns. *Am. J. Phys. Anthropol.* **543**, 539–543 (1998).
58. E. Ersmark *et al.*, Population demography and genetic diversity in the Pleistocene Cave Lion. *Open Q.* **1**, 4 (2015).
59. A. W. Briggs *et al.*, Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res.* **38**, e87 (2010).
60. E. Palkopoulou *et al.*, Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).
61. M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 2010, pdb.prot5448 (2010).
62. C. Hahn, L. Bachmann, B. Chevreux, Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–A baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).
63. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint] (2013). https://doi.org/10.48550/arXiv.1303.3997 (Accessed 1 January 2025).
64. M. Kearse *et al.*, Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
65. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
66. P. Pečnerová *et al.*, Genome-based sexing provides clues about behavior and social structure in the woolly mammoth. *Curr. Biol.* **27**, 3505–3510.e3 (2017).
67. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
68. J. Butler *et al.*, ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Res.* **18**, 810–820 (2008).
69. A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
70. M. Seppey, M. Manni, E. M. Zdobnov, "BUSCO: Assessing genome assembly and annotation completeness" in *Gene Prediction: Methods and Protocols*, M. Kollmar, Ed. (Springer, New York, NY, 2019), pp. 227–245.
71. V. E. Kutschera *et al.*, GenErode: A bioinformatics pipeline to investigate genome erosion in endangered and extinct species. *BMC Bioinf.* **23**, 228 (2022).
72. H. Li *et al.*, The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
73. A. McKenna *et al.*, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
74. E. Lord *et al.*, Population dynamics and demographic history of Eurasian collared lemmings. *BMC Ecol. Evol.* **22**, 126 (2022).
75. K. Malde, R. Skern, K. A. Glover, Using sequencing coverage statistics to identify sex chromosomes in minke whales. arXiv [Preprint] (2019). https://doi.org/10.48550/arXiv.1902.06654 (Accessed 1 January 2025).
76. D. Nishimura, RepeatMasker. *Biotech. Softw. Internet Rep.* **1**, 36–39 (2000).
77. J. M. Flynn *et al.*, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
78. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
79. D. Darriba, G. L. Taboada, R. Doallo, D. Posada, jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
80. M. A. Suchard *et al.*, Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
81. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
82. A. Rambaut, FigTree, a graphical viewer of phylogenetic trees (2007). https://tree.bio.ed.ac.uk/software/figtree (Accessed 20 April 2022).
83. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of next generation sequencing data. *BMC Bioinf.* **15**, 356 (2014).
84. V. Lefort, R. Desper, O. Gascuel, FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).
85. J. L. A. Paijmans *et al.*, African and Asian leopards are highly differentiated at the genomic level. *Curr. Biol.* **31**, 1872–1882.e5 (2021).
86. A. R. Quinlan, I. M. Hall, BEDtools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

87. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
88. R. R. Bouckaert, Densitree: Making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372–1373 (2010).
89. S. H. Martin, S. M. Van Belleghem, Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
90. N. Dussex *et al.*, Population genomics of the critically endangered kākāpō. *Cell Genom.* **1**, 100002 (2021).
91. E. E. Deinum *et al.*, Recent evolution in *Rattus norvegicus* is shaped by declining effective population size. *Mol. Biol. Evol.* **32**, 2547–2558 (2015).
92. A. Uchimura *et al.*, 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 (2015).
93. X. Wang, S. Peischl, G. Heckel, Demographic history and genomic consequences of 10,000 generations of isolation in a wild mammal. *Curr. Biol.* **33**, 2051–2062.e4 (2023).
94. R. E. Green *et al.*, A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
95. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
96. P. Skoglund *et al.*, Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
97. P. Cingolani *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
98. E. Eden, R. Navon, I. Steinfeld, D. Lipson, Z. Yakhini, GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinf.* **10**, 48 (2009).
99. B. T. Sherman *et al.*, DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
100. E. Lord *et al.*, Genome analyses suggest recent speciation and postglacial isolation in the Norwegian Lemming. European Nucleotide Archive. https://www.ebi.ac.uk/ena/browser/view/PRJEB87511. Deposited 28 May 2025.