Original article:

On the origin of the Norwegian lemming

Vendela K. Lagerholm [1,2*], Edson Sandoval-Castellanos [1,2], Dorothee Ehrich [3], Natalia I. Abramson [4], Adam Nadachowski [5], Daniela C. Kalthoff [6], Mietje Germonpré [7], Anders Angerbjörn [2], John R. Stewart [8], Love Dalén [1]

1. Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, 10405, Sweden
2. Department of Zoology, Stockholm University, Stockholm, 10405, Sweden
3. Department of Arctic and Marine Biology, University of Tromsø, Tromsø, 9037, Norway
4. Zoological Institute, Russian Academy of Sciences, St. Petersburg, 199034, Russia
5. Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Kraków, 31-016, Poland
6. Department of Zoology, Swedish Museum of Natural History, Stockholm, 10405, Sweden
7. Operational Direction "Earth and History of Life", Royal Belgian Institute of Natural Sciences, Brussels, 1000, Belgium
8. Faculty of Science and Technology, Bournemouth University, Dorset, BH12 5BB, United Kingdom

Keywords: Speciation, colonisation, extinction, Pleistocene, ancient DNA, *Lemmus*

* Correspondence: Vendela K. Lagerholm, Fax: +46 (0)8 5195 5181,
E-mail: vendela.k.lagerholm@nrm.se

Running title: Norwegian lemming palaeogenetics

Abstract

The Pleistocene glacial cycles resulted in significant changes in species distributions, and it has been discussed whether this caused increased rates of population divergence and speciation. One species that is likely to have evolved during the Pleistocene is the Norwegian lemming (*Lemmus lemmus*). However, the origin of this species, both in terms of when and from what ancestral taxon it evolved, has been difficult to ascertain. Here, we use ancient DNA recovered from lemming remains from a series of Late Pleistocene and Holocene sites to explore the species' evolutionary history. The results revealed considerable genetic differentiation between glacial and contemporary samples. Moreover, the analyses provided strong support for a divergence time prior to the Last Glacial Maximum, therefore likely ruling out a post-glacial colonisation of Scandinavia. Consequently, it appears that the Norwegian lemming evolved from a small population that survived the Last Glacial Maximum in an ice-free Scandinavian refugium.

Introduction

Changes in the distribution of organisms are thought to be one of the main drivers of speciation (Hewitt 1996; Mayr 1963). This is because habitat fragmentation and long distance colonisation events can lead to isolation of conspecific populations, with subsequent evolutionary divergence due to genetic drift and natural selection. The Pleistocene glacial cycles (*c.* 2.6 million to 12 thousand years ago) are considered to have had a major impact on the distribution and abundance of species, especially in temperate and polar regions (Stewart *et al.* 2010). It has therefore been proposed that glacial cycles have led to increased rates of speciation, both due to prolonged periods of isolation in refugia and due to colonisation of, and adaptation to, previously uninhabitable regions (Hewitt 1996; Johnson & Cicero 2004; Rand 1948). However, this view has been a topic of considerable debate in recent decades, since it has been argued that phylogenetic estimates of speciation rates are not higher during the Pleistocene compared to earlier time periods (Klicka & Zink 1997). Moreover, the lengths of glacials and interglacials have been considered too short to permit evolution of reproductive isolation among conspecific populations (Brochmann *et al.* 2003; Lister 2004). On the other hand, theoretical work (Mayr 1954), as well as some empirical studies (Johnson *et al.* 1996), suggest that speciation can occur rapidly under certain conditions.

One of the main problems in investigating whether the Pleistocene climate fluctuations influenced speciation stems from the difficulty in identifying the precise location and timing of such events. This is partly due to that morphological divergence is expected to be low during early stages of speciation, and thus typically invisible in the fossil record. In addition, inference

using modern DNA data and fossil-based molecular clocks usually lack sufficient resolution to reconstruct past speciation events (Ho *et al.* 2011a; Hofreiter & Barnes 2010; Lister 2004).

One species that likely evolved comparatively recently is the Norwegian lemming (*Lemmus lemmus*). Today, the Norwegian lemming inhabits the mountain tundra of Fennoscandia, a region encompassing the Scandinavian Peninsula, Finland and the Kola Peninsula (Fig. 1), which is thought to have been completely covered by the Scandinavian Ice Sheet between approximately 30 to 16 thousand calendar years before present (kyr BP) (Mangerud *et al.* 2011; Svendsen *et al.* 2004). In this paper, we refer to this time period as the Last Glacial Maximum (LGM), although it should be noted that other more narrow definitions have been used in other studies (Svensson *et al.* 2006).

The Norwegian lemming is the only endemic mammal in Fennoscandia, and its origin is therefore somewhat of a mystery. Previous genetic analyses on modern DNA have shown that it is too different from its sister species, the Siberian lemming (*L. sibiricus*), to have evolved from a post-glacial common ancestor (Fedorov & Stenseth 2001). Consequently, the Norwegian lemming either originates from a non-Siberian source population outside the Scandinavian Ice Sheet (Østbye *et al.* 2006), or it originates from a small population that survived the Last Glacial Maximum in a local northern refugium (Ekman 1922).

Numerous fossil remains have shown that lemmings of the genus *Lemmus* were common inhabitants of the vast steppe-tundra of midlatitude Europe and Asia during the Late Pleistocene glacial period (e.g. Nadachowski 1982). Being members of a cold-adapted genus, these southern populations disappeared during the transition to the current Holocene interglacial, and it has not yet been established whether they became extinct or shifted their distribution to more northern latitudes as the temperature increased. It has been postulated that some of these southern *Lemmus*

populations tracked their tundra habitat to the Scandinavian Peninsula as the ice margin retreated, and subsequently founded the modern Norwegian lemming population (Østbye *et al.* 2006). In contrast to the post-glacial colonisation hypothesis, it has also been proposed that the species actually survived the last glaciation *in situ* in an ice-free area of Scandinavia, possibly on Andøya or on a part of the continental shelf that was exposed during times of low sea level (Ekman 1922; Fedorov & Stenseth 2001). However, there is no fossil evidence of Norwegian lemmings in Scandinavia during the Last Glacial Maximum to support this second hypothesis, although *Lemmus* sp. bones of ~ 36 kyr BP in age have been found in Norway from the Ålesund interstadial, indicating the presence of lemmings in the area *before* the last glacial advance (Larsen *et al.* 1987).

The aim of this study was to use ancient DNA from *Lemmus* spp. remains to further investigate the evolutionary history of the Norwegian lemming. More explicitly, we examined the two contrasting hypotheses (Fig. 2) discussed above to resolve whether the Norwegian lemming is derived either from a post-glacial colonisation from midlatitude Europe (scenario 1), or from a population of lemmings that colonised Scandinavia before the Last Glacial Maximum and then survived locally in an ice-free northern refugium (scenario 2).

Materials and methods

Data collection

A total of 54 Late Pleistocene *Lemmus* spp. mandibles, spanning between ~ 12 kyr and 48 kyr BP in age, were collected from 11 paleontological sites across the genus' glacial range in midlatitude Europe (Fig. 1; Table S1, Supporting Information). Further, we also included 27 mandibles from

early-mid Holocene lemmings (between ~ 3 kyr and 8 kyr BP in age) found in the Sirijorda Cave in northern Norway. A modified version of protocol C in Yang *et al*. (1998) was used to extract DNA from the Late Pleistocene samples, whereas Qiagen's QIAamp Tissue kit was used for the Holocene cave samples as described in Fernández *et al*. (2006). For the modern data set, 17 Norwegian lemming (*L. lemmus*) tissue samples from seven localities along the Swedish mountain range were extracted at the Swedish Museum of Natural History using the QIAamp DNA mini kit (Qiagen), with the protocol DNA Purification from Tissues. In order to get an estimate of the interspecific variation in European *Lemmus* spp. we also extracted DNA from 11 modern Siberian lemming (*L. sibiricus*) bone samples from three localities within the north-western phylogeographic group (Fedorov *et al.* 1999) using the same protocol as for the Late Pleistocene *Lemmus* spp. samples. To avoid confusion, the early-mid Holocene Norwegian cave samples are hereafter called Holocene Scandinavian, while the modern samples of *L. lemmus* (Norwegian lemming) are called modern Scandinavian.

We targeted two mitochondrial regions previously used in modern phylogenetic studies of the *Lemmus* genus, comprising the first hypervariable part of the control region (CR) and parts of the cytochrome *b* (cyt *b*) gene. Further details regarding DNA extraction, PCR amplification and sequencing are presented in the Supporting Information online.

The pre-PCR work on the Late Pleistocene samples was carried out in the ancient DNA laboratory at the Swedish Museum of Natural History, where no previous work on *Lemmus* spp. had been done. For all Late Pleistocene samples, at least two independent amplifications were done in order to resolve erroneous bases caused by misincorporation during PCR. The Holocene Scandinavian samples were analysed in the ancient DNA laboratory at Laboratoire d'Ecologie Alpine in Grenoble, France, where no rodent samples had been analysed before. Since the

sequences obtained from Sirijorda cave only displayed variation in nucleotide positions that are variable in extant lemming populations, it seemed unlikely that the observed variation could have been caused by PCR misincorporation. The Sirijorda sequences were therefore not considered necessary to replicate through multiple PCRs. All working surfaces and lab equipment were regularly sterilised with UV light, bleach or hydrochloric acid, and extraction and PCR blanks were extensively used to monitor possible contamination. The pre-PCR work on the modern samples was carried out at the Swedish Museum of Natural History, in laboratories physically separated from both the ancient DNA and post-PCR facilities.

Eleven Late Pleistocene lemming mandibles that gave successful DNA sequences were dated at the Oxford Radiocarbon Accelerator Unit. Five of these produced radiocarbon dates, which were calibrated to calendar years before present using OxCal 4.1.7 (Bronk Ramsey 2009) and the IntCal 09 calibration curve (Reimer *et al.* 2009). The remaining six samples failed due to low collagen yields, likely owing to the small size of the lemming mandibles rather than poor biomolecular preservation. All dates, including the inferred ages of the remaining samples in the data set, are listed in Table S1, Supporting Information.

Data analyses

Sequences were aligned and edited using the software SeqMan in the package Lasergene v8.1.5 (DNASTAR). BioEdit v1.7.3 (Hall 1999) was subsequently used to construct a combined data set of 520 bp, consisting of both CR (168 bp) and cyt *b* (352 bp) sequences. Additionally, we used a partial data set of 172 bp (96 bp CR and 76 bp cyt *b*) that also included the Holocene Scandinavian lemming sequences. The Late Pleistocene sample sites were assigned to six

124 geographic regions; England, Belgium, Germany, Poland, Russian plains and Ural Mountains.
125 Genetic diversity within these, Holocene Scandinavia, and the two modern regions (Scandinavia
126 and NW Russia) were calculated with Arlequin v3.5.1.2 (Excoffier & Lischer 2010).

127 Temporal statistical parsimony networks were created with the R-script TempNet v1.4 (Prost &
128 Anderson 2011) to display the haplotypes found in the different time periods (the Holocene and
129 the Late Pleistocene). The phylogenetic relationships among all samples and the divergence times
130 for different lineages were calculated with the software BEAST v1.6.1 (Drummond & Rambaut
131 2007). Initially, the analyses were performed using the nucleotide substitution model GTR+G, as
132 an analysis in MrModeltest v2.3 (Nylander 2004) showed this to be the most appropriate model
133 of nucleotide substitution. However, due to poor mixing of the Markov chain Monte Carlo
134 (MCMC) this was later changed to the simpler model HKY+G. The phylogenetic analyses were
135 performed using a strict molecular clock with fixed mutation rates. Based on the previously
136 published rates of 17 % Myr^{-1} (CR) and 5 % Myr^{-1} (cyt *b*) (Fedorov & Stenseth 2001), and the
137 relative length of each gene region in our combined data sets, we set the mutation rate to 8.9 %
138 Myr^{-1} and 11.7 % Myr^{-1} respectively for the 520 bp and the 172 bp alignments. However, since
139 the mutation rate could have a strong impact on the subsequent analyses, and concerns have been
140 raised about biases in mutation rate estimates (Ho *et al.* 2011a), we also ran the analyses using
141 mutation rates of 30 % and 50 % Myr^{-1} which encompass the range of previously published
142 estimates from ancient DNA data sets on large herbivores, such as saiga (Saiga tatarica; Campos
143 *et al.* 2010a) and bison (Bison bison; Shapiro *et al.* 2004). It should be noted that the issue of
144 elevated mutation rates in ancient DNA data sets is a topic of discussion (Ho *et al.* 2011b;
145 Navascués *et al.* 2010; Navascués & Emerson 2009), wherefore using a rate of 50 % Myr^{-1} in our
146 analyses may seem unreasonably high. Nonetheless, we decided to include it since the mutation

147 rate is inversely proportional to the split time estimated in the coalescent simulations and we
148 wanted to exclude the possibility of selecting the wrong scenario due to using a too low mutation
149 rate. The dates of all ancient sequences, obtained either from direct radiocarbon dating or from
150 inferred ages based on stratigraphy and published dates, were included in the analyses. However,
151 due to uncertainties regarding the age of the Ural deposit, the date of these sequences were
152 instead inferred using the option Tip sampling, with a wide uniform prior encompassing the
153 proposed age (28.6 kyr ± 15 kyr BP). To assess the robustness of the analysis, BEAST was also
154 run using only the sequences that had specific ages, i.e. the modern samples and those of the
155 ancient remains that were successfully radiocarbon dated. However, this did not affect the overall
156 topology of the phylogeny, nor the split times among major clades (data not shown). All analyses
157 were made with a randomly generated starting tree, and the length of the MCMC was set to 30
158 and 50 million generations for the partial and the complete data sets, respectively, with
159 parameters logged to file every 1,000 generations. Two independent runs were made for each
160 analysis, and the results were checked in TRACER v1.5 (Rambaut & Drummond 2007) to ensure
161 that runs were converging on the same distribution. The sample of trees obtained from a BEAST
162 run was summarised with TreeAnnotator v1.6.1 to a maximum clade credibility tree with median
163 node heights, using a burnin of 10 % and a posterior probability limit of 0.5, and the output was
164 graphically edited in FigTree v1.3.1 (Rambaut 2009).

165 We also constructed a phylogenetic tree in MrBayes v3.2.2 (Ronquist *et al.* 2012) from the
166 partial data set, in order to investigate the reliability of our topology also without using sample
167 ages and mutation rate as priors. Using the tundra vole, *Microtus oeconomus*, as an outgroup
168 (GenBank accession no AY305172; Galbreath & Cook 2004) we ran the analyses for 5 million
169 generations with the HKY+G substitution model, with a sample and print frequency set to 100.

Two independent runs were made, with results checked in Tracer v1.5 (Rambaut & Drummond 2007) to ensure convergence before discarding 10 % as burnin. The combined tree file was graphically edited in FigTree v1.3.1 (Rambaut 2009).

Approximate Bayesian Computation coupled with coalescent simulations was carried out using the partial dataset, in order to test the two contrasting hypotheses regarding the Norwegian lemming's (*L. lemmus*) glacial history. The statistical inference relied on one single parameter: the time separating all the Scandinavian lemmings (modern and early-mid Holocene) from their closest non-Scandinavian glacial relatives. This inference was based on hypothesis testing using acceptance ratios (Bayes factors) of the simulations of the two proposed scenarios, and the estimation of the mentioned divergence time. The program Bayesian Serial SimCoal (Anderson *et al.* 2005; Excoffier *et al.* 2000) was used to run coalescent simulations (Fig. S1, Supporting Information) for three different analyses: simulations for performing a model comparison (hypothesis contrast) using Bayes factors, simulations for estimating the parameters of interest, and simulations for a cross-validation test using pseudo-observed datasets. Pilot simulations were carried out to test different prior distributions and their effect on the posteriors, as well as to define proper parameter values. Also, a comprehensive selection of summary statistics was carried out in order to select an appropriate and informative set. Other pilot simulations explored alternative population sizes, mutation rates (fixed or sampled from a prior), as well as alternative scenarios and statistical groups. For the parameters estimation, optimisation simulations were made in order to improve the fit of the simulations to the data, and thereby increasing the accuracy of the estimates. Thus, the prior distributions in the final simulations were tuned according to the obtained posteriors in the optimisation runs (but using wider variances) (Bertorelle *et al.* 2010; Lopes *et al.* 2009).

193 The model that was simulated (Fig. S1, Supporting Information) consisted of four populations
194 (Scandinavia, Siberia, glacial England and glacial continental Europe) whose lineages coalesced
195 backwards in time. Population sizes were simulated with initial exponential priors (λ=250,000).
196 This was used because exponential priors sample uniformly in a logarithmic scale, which is
197 advantageous when parameters have ranges covering several orders of magnitude, as in the case
198 of lemming populations sizes which potentially can reach millions of individuals. The parameter
199 value was set to 250,000 for an optimal acceptance rate of the simulations. The Scandinavian
200 population was also set to have an exponential growth starting 11.5 kyr BP, corresponding to a
201 post-glacial population expansion into previously ice-covered Scandinavian areas, since that was
202 expected under both of the hypothesised scenarios. The ages of all Late Pleistocene samples were
203 assigned from normal prior distributions (around the dates listed in Table S1, Supporting
204 Information) to account for the uncertainty in the age estimates, both when these were derived
205 from radiocarbon dating and when inferred from stratigraphic contexts. Generation time was set
206 to 1 per year. As in the BEAST analyses, the simulations were made with three fixed mutation
207 rates; 11.7 %, 30 % and 50 % Myr^{-1}. Post simulation analyses were made in a custom software
208 (available upon request) written in the programming language Fortran 95. In order to deal with
209 the large number of summary statistics employed, the rejection was performed by using a vector
210 containing the threshold distances for every summary statistic (Table S2, Supporting
211 Information). In addition, summary statistics were normalised with the distance between the
212 median of the simulated values and the observed value, which empirically yielded better results
213 than using the variance. Further details regarding the simulation procedures are given in the
214 Supporting Information online.

215

216 Results

217 *Data set*

218 The complete 520 bp sequence targeted in this study was obtained from 23 Late Pleistocene and
219 27 modern samples, while a partial 172 bp fragment was obtained from eight of the early-mid
220 Holocene *Lemmus* spp. samples (Table S1, Supporting Information). We therefore had two data
221 sets containing 50 and 58 sequences, respectively (GenBank accession numbers: JX483882-
222 JX483939).

223

224 *Genetic diversity and phylogenetic relationships*

225 There was a high genetic variation in the glacial data set with a total of 19 and 17 unique
226 haplotypes found in the complete and partial alignments, respectively (Fig. 3; Fig. S2 and Table
227 S4, Supporting Information). This was also reflected in the estimates of nucleotide and haplotype
228 diversities, which generally were higher in the overall glacial data set, although regional levels of
229 diversity in the glacial populations were comparable to those in the modern-day populations.
230 Bayesian phylogenetic analyses showed that the diversity is distributed into three clades (Fig.
231 4). The first two (clades A and B) include the representatives of each of the two modern species,
232 whereas the third (clade C) is basal and only includes Late Pleistocene lemmings. The modern
233 Scandinavian samples form a well supported monophyletic group together with all but one of the
234 Holocene Scandinavian cave samples (Figs 3 and 4; Fig. S3, Supporting Information). For all
235 mutation rates used (11.7, 30 and 50 % Myr^{-1}), the estimated time to the most recent common
236 ancestor (tMRCA) for this Scandinavian group and the most closely related Late Pleistocene

sequences pre-date the final retreat of the Scandinavian Ice Sheet (100 kyr, 43 kyr and 32 kyr BP respectively; Fig. 4; Table 1). Very similar results were obtained from the network and phylogeny constructed from the 520 bp data set (Fig. S2, Supporting Information), and the tMRCA estimates pre-dated the last glacial retreat also when the 520 bp sequences were used (Table S5, Supporting Information). Further, the robustness of the overall tree topology estimated in BEAST was supported by Bayesian phylogenetic analyses that did not incorporate sequence dates or pre-defined mutation rates (Fig. S3, Supporting Information), although some deeper internal nodes found in the BEAST analyses could not be resolved.

Bayesian coalescent simulations of population divergence times

Consistent with the phylogenetic results, the Bayesian coalescent simulations of the two hypothesised scenarios strongly supported a population divergence that pre-dated the last glacial retreat (Fig. 5; Table 2). The acceptance ratio yielded a higher support for this scenario, with Bayes factors of 7.4, 48.3 and 37.2 (for mutation rates of 11.7, 30 and 50 % Myr^{-1}, respectively). In the pseudo-observed datasets (PODs) analysis, the probabilities of selecting the right scenario were 0.67 and 0.81 for scenarios 1 and 2, respectively, when a mutation rate of 11.7% Myr^{-1} was assumed. Additionally, the analysis that took into account the observed Bayes factor, in which the only PODs that were considered were those with a Bayes factor equal to or larger than the observed one, yielded values of 0.98 and 0.89 for scenarios 1 and 2 respectively. For the mutation rate of 30% Myr^{-1}, the corresponding values were 0.90 and 0.90 in the first run, and 0.97 and 0.97 when considering the observed Bayes factor. The mutation rate of 50% Myr^{-1} resulted in values

of 0.94 and 0.90 in the first run, which rose to 0.98 and 0.99 respectively after the observed Bayes factor was taken into account.

Two of the summary statistics allowed a good differentiation between our hypotheses; the mean number of pairwise differences and the F_{ST} between Scandinavian lemmings (including both modern and early-mid Holocene cave samples) and their closest glacial relatives (Fig. S4, Supporting Information). The observed values for both these statistics were too high to correspond to scenario 1 (*p*-value 0.06-0.0098), but were not significantly differentiated from scenario 2 (*p*-value 0.36-0.09).

When using coalescent simulations coupled with Approximate Bayesian Computation analysis, there is always a concern that the true scenario is not incorporated among the models tested (Templeton 2009). However, one way to address this problem is to assess how well the models fit to the empirical data (Csillery *et al.* 2010). To address this, we compared the posterior distributions with the summary statistics of the observed data set. The results showed that the observed data had a close fit to the non-rejected summary statistics in the simulated data sets, which indicates that the simulated models provide a good fit to the empirical data (Fig. S5, Supporting Information).

Discussion

Our results indicate a large genetic variation in the lemming populations that inhabited the steppe-tundra region of midlatitude Europe during the Late Pleistocene period. In particular, the glacial populations in Eastern Europe appear to have had a very high nucleotide diversity, which could reflect long-term occupation in the region (Table S4, Supporting Information). It should be

noted, however, that these diversity estimates may to some extent be inflated due to the heterochronous nature of the data (Depaulis *et al.* 2009). Nonetheless, the seemingly high genetic variation in the glacial populations, as well as the large effective population sizes estimated in the Bayesian coalescent simulations (Fig. S6, Supporting Information), support the view that *Lemmus* spp. were common in the European Late Pleistocene steppe-tundra ecosystem.

The modern Scandinavian population displays low levels of nucleotide diversity and a star-like pattern in the haplotype network (Fig. 3; Table S4, Supporting Information), which indicate a previous reduction in population size followed by a demographic expansion, as also previously demonstrated in a mismatch distribution test done by Fedorov & Stenseth (2001). This could either correspond to a bottleneck during the Last Glacial Maximum in line with the hypothesis of local glacial survival (i.e. scenario 2), or a post-glacial founder event (i.e. scenario 1). Both these hypotheses are supported by the observation that all but one of the Scandinavian cave samples from the early-mid Holocene fall within the diversity of the modern samples (Figs. 3 and 4), thus making a more recent genetic bottleneck unlikely.

The central, and most common, haplotype in Scandinavia (Fig. 3) is likely to represent either the haplotype that survived the hypothesised LGM bottleneck, or alternatively, the founding haplotype during a post-glacial colonisation. However, this haplotype was not observed in any of the glacial populations that surrounded the Scandinavian Ice Sheet, which could have been expected if the Norwegian lemming (*Lemmus lemmus*) originated from a post-glacial colonisation from these southern populations. Instead, the most recent common ancestor (MRCA) to the Norwegian lemming and the most closely related glacial lemmings was estimated to have lived between 100 kyr to 32 kyr BP. Even for the extreme mutation rate of 50 % Myr^{-1}, the lower bound of the 95% highest posterior density (HPD) interval does not include the time

after the final retreat of the Scandinavian Ice Sheet (Table 1). Although the tMRCA may predate the time of actual population divergence, this difference is reduced when the populations are small, which likely was the case for the founder population of the Norwegian lemming. Furthermore, the Bayesian coalescent simulations provided a markedly higher support for a pre-LGM divergence between Scandinavian and glacial European populations, with an estimated population divergence time of more than 78 kyr BP (lower 95 % HPD for 50 % mutation rate = 32 kyr BP; Table 2). It should be noted that the coalescent framework takes the temporal dimension and co-ancestry relationships into account at once. Therefore, the possibility that the dominant haplotype in extant *L. lemmus* existed outside the ice sheet during the Last Glacial Maximum, but was not sampled, would not only be the likelihood that it was absent in our glacial European sample, but the likelihood that it was not in the sample and that it did not coalesce with any of the sampled lineages. The estimation performed in the Approximate Bayesian Computation analyses targeted the time of the divergence of the Scandinavian lemmings from their ancestral population, and not the time to the origin of the lineage. The time window between those two events (the origin of the lineage and the origin of the population) is therefore where potentially unsampled lineages could have coalesced, thus producing a more recent origin of the *L. lemmus* lineage. However, such a scenario was not supported, and instead the results suggest that none of the populations that lived south of the Scandinavian Ice Sheet during the end of the last glaciation were the direct ancestors of the Norwegian lemming. Consequently, the most parsimonious explanation is that the species originates from a population that survived the Last Glacial Maximum in a northern refugium.

The hypothesis of small ice-free refugia in Scandinavia during the Last Glacial Maximum recently gained support in a study by Parducci *et al*. (2012), which reported paleoecological and

326 genetic data suggesting a local glacial survival of pine (*Pinus sylvestris*) and spruce (*Picea abies*)
327 in Scandinavia (but see Birks *et al.* 2012). The data presented in this study thus lends further
328 support to the local northern refugium hypothesis, and suggests that this putative ice-free area
329 was diverse or large enough to harbour both Arctic taxa like lemmings, as well as Boreal trees.
330 Alternatively, there might have existed multiple refugia that were inhabited by differently
331 adapted plant and animal communities, for example at higher altitudes or on the part of the
332 continental shelf that was flooded by rising sea levels during the Holocene (Nesje *et al.* 2007).
333 As indicated by the divergence time estimates in the phylogeny and the Bayesian coalescent
334 simulations, it appears likely that Scandinavia was colonised by European lemmings (*Lemmus*
335 sp.) during an interstadial period sometime between the Karmøy glaciation, which ended ~ 60 kyr
336 BP (Mangerud *et al.* 2011), and the last glacial advance ~ 30 kyr BP (see Fig. 1a). The
337 occurrence of *Lemmus* sp. fossil remains in Scandinavia dating to the Ålesund interstadial ~ 36
338 kyr BP (Larsen *et al.* 1987) also confirms that the region was populated at this time period,
339 although unfortunately we have no genetic information on these. As the Scandinavian Ice Sheet
340 started to grow during late Marine Isotope Stage (MIS) 3, culminating in the full glacial
341 conditions during MIS 2 (Svensson *et al.* 2006), the lemming population in Scandinavia must
342 have become increasingly small and isolated from the surrounding southern populations. As the
343 ice sheet melted during the early Holocene, Norwegian lemmings originating from the ice-free
344 northern refugium likely expanded into the previously glaciated regions of Scandinavia.
345 Interestingly, the observation in this study of a ~ 8 kyr BP old specimen from Sirijorda Cave in
346 Norway carrying a haplotype today only found in *L. sibiricus* (Figs. 3 and 4) indicates that
347 Siberian lemmings may have expanded into Scandinavia as the ice sheet melted. Alternatively,
348 introgression between the two species may have led to inclusion of *L. sibiricus* haplotypes in the

L. lemmus gene pool. Haplotypes belonging to *L. sibiricus* have, however, not been observed in any modern Norwegian lemmings (this study; Fedorov & Stenseth 2001) and the species does not inhabit the region today. Moreover, this observation is based on one single sample, making further evaluation of the existence and extent of past gene flow from *L. sibiricus* into Scandinavia difficult at present.

The results presented here indicate that the end-Pleistocene midlatitude European *Lemmus* populations did not contribute to the gene pool of the contemporary lemming populations in Scandinavia and northwest Russia. Instead, it appears that the midlatitude populations became extinct at the Pleistocene-Holocene transition, and that this led to a marked decrease in genetic diversity that included the loss of a major mitochondrial (mt) DNA clade (Figs. 3 and 4). Whether this extinct clade represents a divergent population or a separate species is difficult to ascertain at present, since we have only analysed mtDNA. In any case, the observed loss of an entire clade adds to a growing body of evidence suggesting that many glacial populations and species were unable to track the shifts and contractions in habitat that took place at the end of the last Ice Age (Campos *et al.* 2010a; Campos *et al.* 2010b; Dalén *et al.* 2007).

An inability of populations to track reductions in habitat availability implies that a succession of expansions and contractions in species ranges, such as the ones that likely took place at the end of the Pleistocene, would have been characterised by a series of population extinctions (Brace *et al.* 2012). This could provide an explanation for the observation that many extant Holarctic species appear to have lost significant amounts of genetic diversity since the Late Pleistocene (Hofreiter & Barnes 2010). With the ongoing increases in global temperatures, this in turn raises concerns about the fate of extant cold-adapted populations that inhabit the southern margins of the Arctic biome.

From an evolutionary perspective, it appears likely that the northern survival of a small and isolated *Lemmus* sp. population during the Last Glacial Maximum may have contributed to the evolution of the Norwegian lemming, or possibly even represents the speciation event itself. This is consistent with the hypothesis that adopting a new refugium provides a mechanism of speciation (Stewart & Stringer 2012). Both the isolation and the small population size could have led to rapid evolutionary changes, consistent with the model of peripatric speciation (Mayr 1963). Such evolutionary changes may also have been reinforced by shifts in local ecological conditions (Orr & Smith 1998) due to the changes in temperature and precipitation associated with the onset of the Last Glacial Maximum, as well alterations in the lemmings' realised niche if their key predators and competitors were unable to persist in the local refugium (Dalén *et al.* 2007; Hewitt 1996; Stewart 2008). Further analyses of autosomal genes, including those under natural selection, from serially sampled Norwegian lemming specimens could thus constitute a unique opportunity to study the speciation process in real time.

Acknowledgements

We are grateful to field workers within the Scandinavian arctic fox project (SEFALO+), staff at the Tarfala research station, and Maria Noring (Royal Institute of Technology) for contributing with modern samples. We thank Eivind Østbye (University in Oslo) for providing the Holocene Norwegian samples, Ralf W. Schmitz (University of Bonn), Nick Barton (Oxford University), Christelle Draily (Director of Walou Cave Excavations), Annelise Folie (Royal Belgian Institute

394 of Natural Sciences) and Anne Taylor (Cambridge University Museum of Archaeoology and
395 Anthropology) for giving us access to samples from the Neandertal site excavation, Merlin's
396 Cave, Walou Cave, Caverne Marie-Jeanne and Bridged Pot Cave, respectively. We also thank
397 Pierre Taberlet and Christian Miquel (Université de Grenoble) respectively for providing access
398 to the DNA facilities in Grenoble and for technical support in the same lab. Additional field
399 collection of modern samples was supported through grants from the Royal Swedish Academy of
400 Sciences and Riksmusei Vänner to V.K.L. The genetic analyses were funded through grants to
401 L.D. from the Swedish Research Council (VR) and FORMAS through the FP6 BiodivERsA
402 ERA-NET program. E.S-C. also acknowledges support from the Strategic Research Programme
403 EkoKlim at Stockholm University.

Tables

Table 1 Median ages, and the 95% highest posterior density (HPD) interval, in thousands (k) of years before present for the phylogenetic tree nodes shown in Fig. 4. The estimated times to the most recent common ancestor of all Scandinavian lemmings and the most closely related Late Pleistocene European lemmings are shown in bold. Based on BEAST analyses of the partial data set, using mutation rates of 11.7 %, 30 % and 50 % Myr^{-1}.

	11.7 % Myr^{-1}		30 % Myr^{-1}		50 % Myr^{-1}	
	Node age	95 % HPD	Node age	95 % HPD	Node age	95 % HPD
A	467 k	725 k - 285 k	190 k	284 k - 123 k	130 k	184 k - 87 k
B	260 k	414 k - 144 k	128 k	183 k - 84 k	97 k	133 k - 71 k
C	300 k	476 k - 173 k	112 k	173 k - 69 k	75 k	109 k - 52 k
D	152 k	263 k - 72 k	57 k	90 k - 35 k	41 k	59 k - 28 k
E	**100 k**	**169 k - 49 k**	**43 k**	**64 k - 26 k**	**32 k**	**46 k - 21 k**
F	86 k	164 k - 36 k	38 k	59 k - 22 k	30 k	42 k - 20 k
G	78 k	148 k - 33 k	32 k	53 k - 16 k	22 k	36 k - 12 k
H	65 k	113 k - 33 k	29 k	45 k - 17 k	22 k	33 k - 13 k

Table 2 Descriptive statistics for the estimated population divergence times between Scandinavian and Late Pleistocene European lemmings, as shown in Fig. 5. The estimates are based on the posterior probability distributions obtained in the Bayesian coalescent simulations of the partial data set, using three different mutation rates.

	11.7 % Myr^{-1}	30 % Myr^{-1}	50 % Myr^{-1}
Mode	101 k	117 k	82 k
Median	93 k	90 k	79 k
Mean	86 k	87 k	78 k
95% HPD Lower	23 k	37 k	32 k

Figures

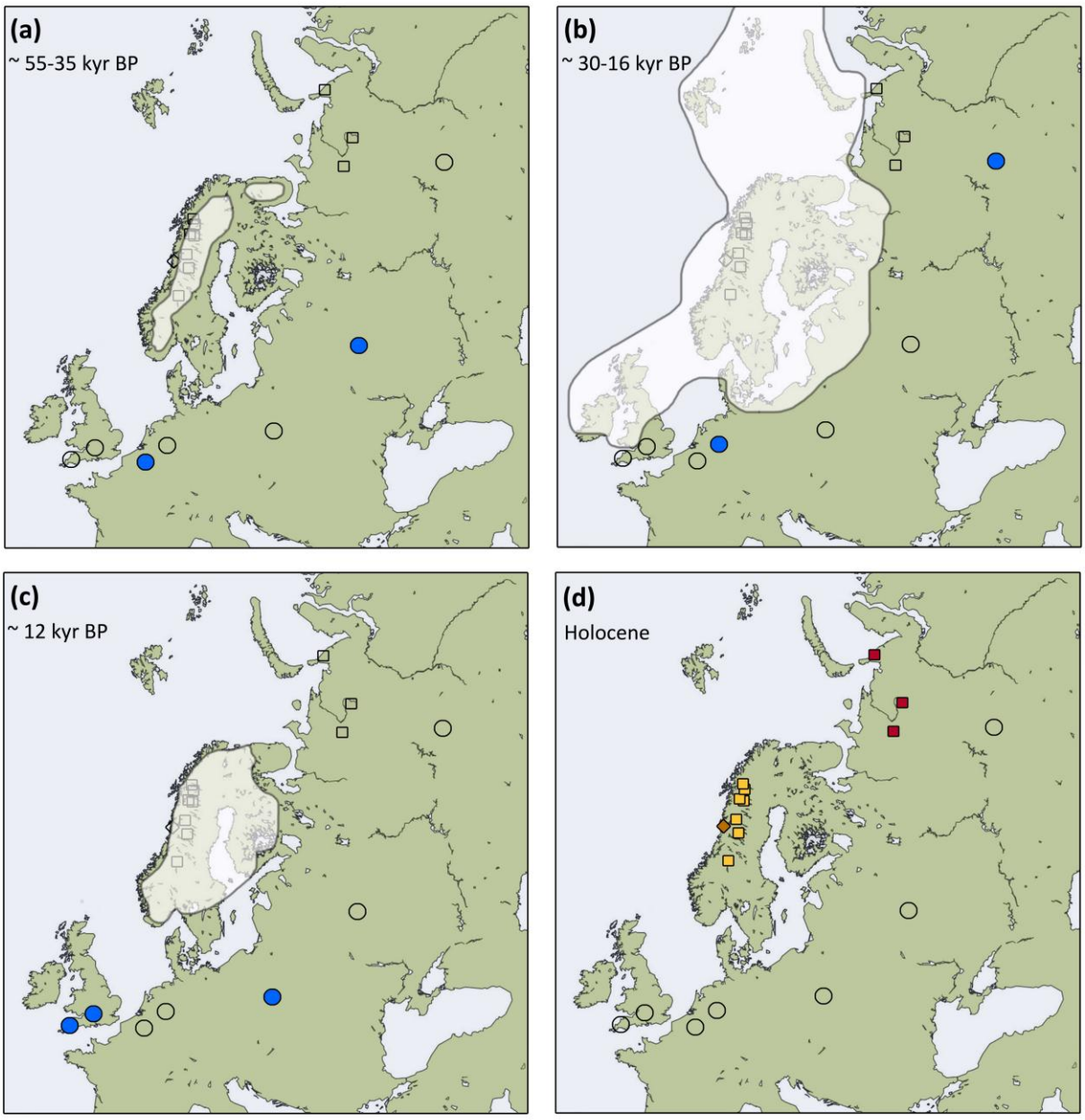

Fig. 1

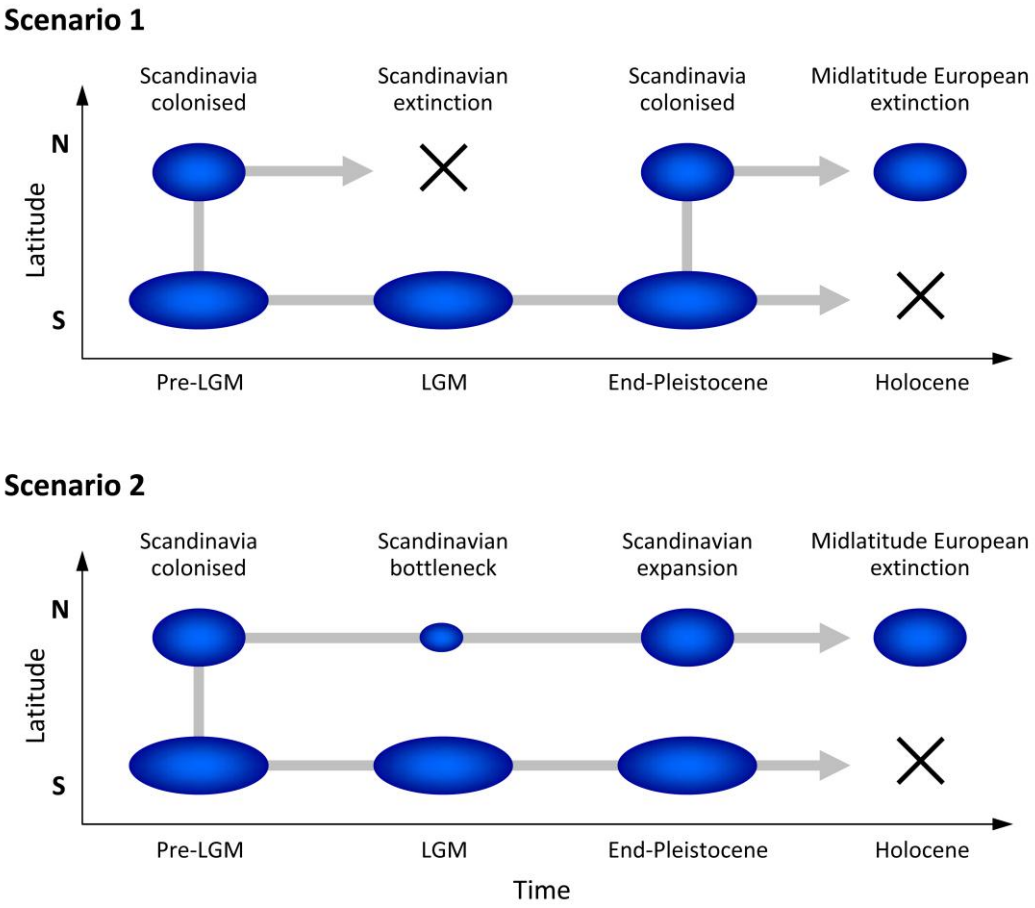
Scenario 1
Scandinavia colonised
Scandinavian extinction
Scandinavia colonised
Midlatitude European extinction
N
S
Latitude
Pre-LGM
LGM
End-Pleistocene
Holocene
Scenario 2
Scandinavia colonised
Scandinavian bottleneck
Scandinavian expansion
Midlatitude European extinction
N
S
Latitude
Pre-LGM
LGM
End-Pleistocene
Holocene
Time

Fig. 2

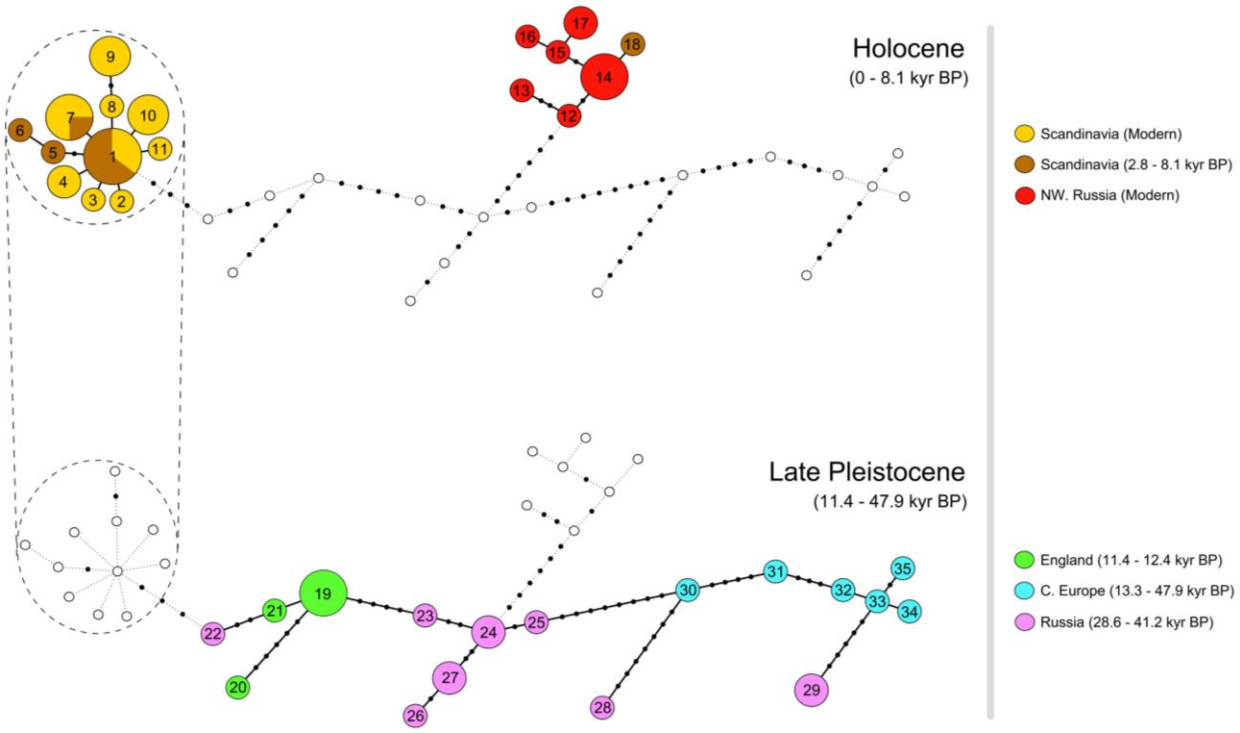
Holocene
(0 - 8.1 kyr BP)
Late Pleistocene
(11.4 - 47.9 kyr BP)
Scandinavia (Modern)
Scandinavia (2.8 - 8.1 kyr BP)
NW. Russia (Modern)
England (11.4 - 12.4 kyr BP)
C. Europe (13.3 - 47.9 kyr BP)
Russia (28.6 - 41.2 kyr BP)

Fig. 3

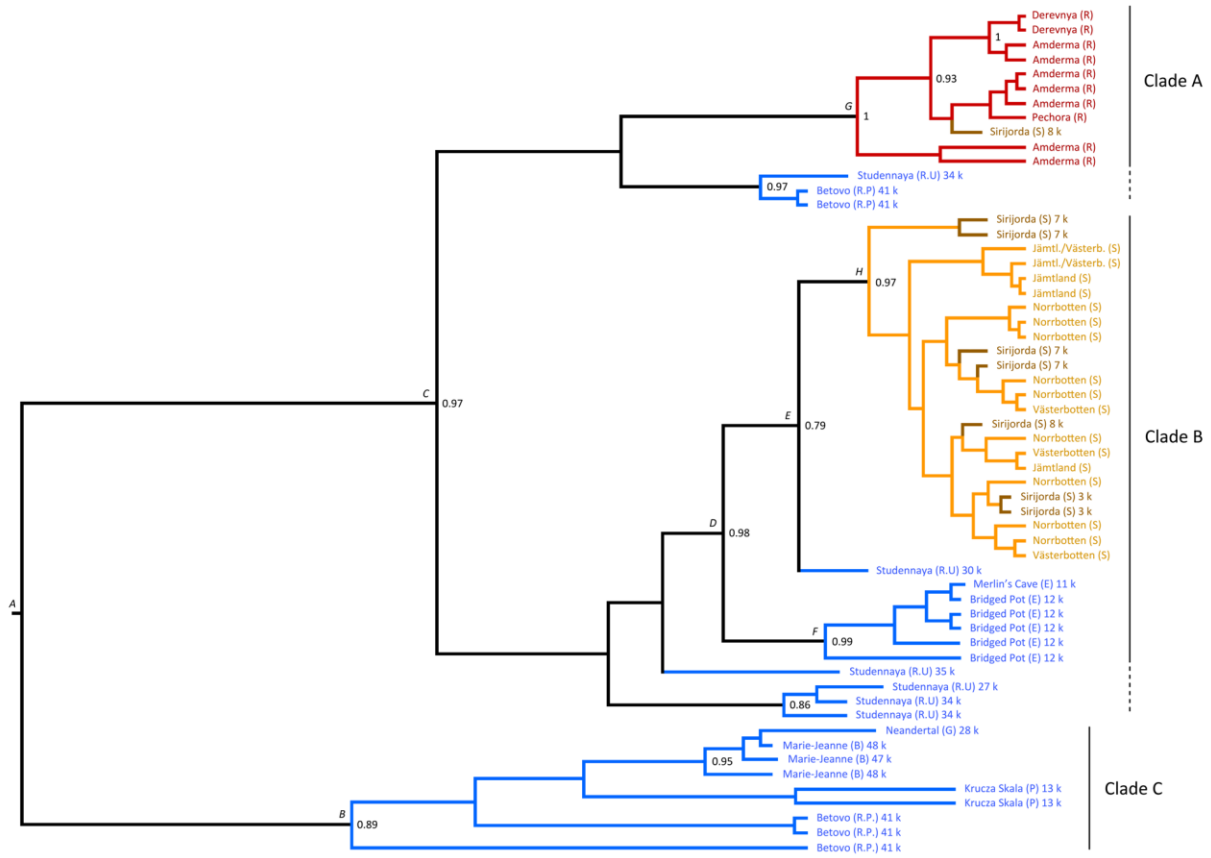
Derevnya (R)
Derevnya (R)
Amderma (R)
Amderma (R)
Amderma (R)
Amderma (R)
Amderma (R)
Pechora (R)
Sirijorda (S) 8 k
Amderma (R)
Amderma (R)
Studennaya (R.U) 34 k
Betovo (R.P) 41 k
Betovo (R.P) 41 k
Clade A
Sirijorda (S) 7 k
Sirijorda (S) 7 k
Jämtl./Västerb. (S)
Jämtl./Västerb. (S)
Jämtland (S)
Jämtland (S)
Norrbotten (S)
Norrbotten (S)
Norrbotten (S)
Sirijorda (S) 7 k
Sirijorda (S) 7 k
Norrbotten (S)
Norrbotten (S)
Västerbotten (S)
Sirijorda (S) 8 k
Norrbotten (S)
Västerbotten (S)
Jämtland (S)
Norrbotten (S)
Sirijorda (S) 3 k
Sirijorda (S) 3 k
Norrbotten (S)
Norrbotten (S)
Västerbotten (S)
Studennaya (R.U) 30 k
Merlin's Cave (E) 11 k
Bridged Pot (E) 12 k
Bridged Pot (E) 12 k
Bridged Pot (E) 12 k
Bridged Pot (E) 12 k
Bridged Pot (E) 12 k
Studennaya (R.U) 35 k
Studennaya (R.U) 27 k
Studennaya (R.U) 34 k
Studennaya (R.U) 34 k
Clade B
Neandertal (G) 28 k
Marie-Jeanne (B) 48 k
Marie-Jeanne (B) 47 k
Marie-Jeanne (B) 48 k
Krucza Skala (P) 13 k
Krucza Skala (P) 13 k
Betovo (R.P.) 41 k
Betovo (R.P.) 41 k
Betovo (R.P.) 41 k
Clade C
A
B
C
D
E
F
G
H
1
0.93
0.97
1
0.97
0.97
0.79
0.98
0.99
0.86
0.95
0.89

Fig. 4

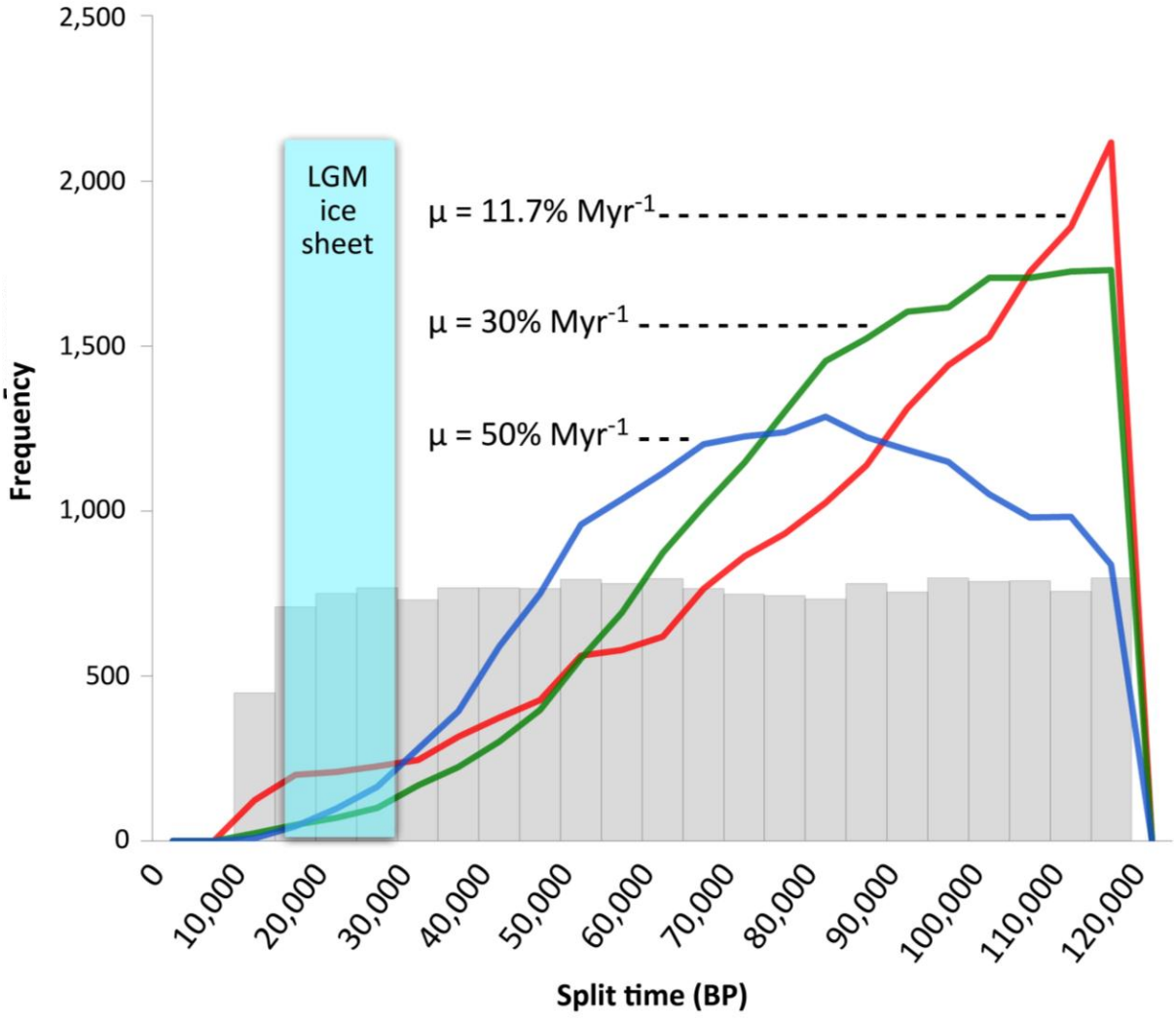
2,500
2,000
1,500
1,000
500
0
Frequency
LGM ice sheet
μ = 11.7% Myr-1
μ = 30% Myr-1
μ = 50% Myr-1
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
90,000
100,000
110,000
120,000
Split time (BP)

Fig. 5

References

Anderson CNK, Ramakrishnan U, Chan YL, Hadly EA (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics*, **21**, 1733-1734.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609-2625.

Birks HH, Giesecke T, Hewitt GM, *et al.* (2012) Comment on "Glacial survival of boreal trees in Northern Scandinavia". *Science*, **338**, 742.

Brace S, Palkopoulou E, Dalén L, *et al.* (2012) Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proceedings of the National Academy of Sciences*.

Brochmann C, Gabrielsen TM, Nordal I, Landvik JY, Elven R (2003) Glacial survival or tabula rasa? The history of North Atlantic biota revisited. *Taxon*, **52**, 417-450.

Bronk Ramsey C (2009) Bayesian analysis of radiocarbon dates. *Radiocarbon* **51**, 337-360.

Campos PF, Kristensen T, Orlando L, *et al.* (2010a) Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene. *Molecular Ecology*, **19**, 4863-4875.

Campos PF, Willerslev E, Sher A, *et al.* (2010b) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos moschatus*) population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5675-5680.

Csillery K, Blum MGB, Gaggiotti O, Francois O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends Ecol Evol*, **25**, 410 - 418.

Dalén L, Nyström V, Valdiosera C, *et al.* (2007) Ancient DNA reveals lack of postglacial habitat tracking in the arctic fox. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 6726-6729.

Depaulis F, Orlando L, Hänni C (2009) Using Classical Population Genetics Tools with Heterochroneous Data: Time Matters! *PLoS ONE*, **4**, e5541.

Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Ekman S (1922) *Djurvärldens utbredningshistoria på den Skandinaviska halvön* Albert Bonniers Förlag, Stockholm.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567.

Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: A general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *Journal of Heredity*, **91**, 506-509.

Fedorov V, Goropashnaya A, Jarrell GH, Fredga K (1999) Phylogeographic structure and mitochondrial DNA variation in true lemmings (*Lemmus*) from the Eurasian Arctic. *Biological Journal of the Linnean Society*, **66**, 357-371.

Fedorov VB, Stenseth NC (2001) Glacial survival of the Norwegian lemming (*Lemmus lemmus*) in Scandinavia: inference from mitochondrial DNA variation. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 809-814.

Fernández H, Hughes S, Vigne J-D, *et al.* (2006) Divergent mtDNA lineages of goats in an Early Neolithic site, far from the initial domestication areas. *Proceedings of the National Academy of Sciences*, **103**, 15375-15379.

Galbreath KE, Cook JA (2004) Genetic consequences of Pleistocene glaciations for the tundra vole (Microtus oeconomus) in Beringia. *Molecular Ecology*, **13**, 135-148.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* **58**, 247-276.

Ho SYW, Lanfear R, Bromham L, *et al.* (2011a) Time-dependent rates of molecular evolution. *Molecular Ecology*, **20**, 3087-3101.

Ho SYW, Lanfear R, Phillips MJ, *et al.* (2011b) Bayesian estimation of substitution rates from ancient DNA sequences with low information content. *Systematic Biology*.

Hofreiter M, Barnes I (2010) Diversity lost: are all Holarctic large mammal species just relict populations? *BMC Biology*, **8**, 46.

Johnson NK, Cicero C (2004) New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution*, **58**, 1122-1130.

Johnson TC, Scholz CA, Talbot MR, *et al.* (1996) Late Pleistocene desiccation of Lake Victoria and rapid volution of Cichlid fishes. *Science*, **273**, 1091-1093.

Klicka J, Zink RM (1997) The importance of recent ice ages in speciation: a failed paradigm. *Science*, **277**, 1666-1669.

Lagerholm VK, Sandoval-Castellanos E, Ehrich D, *et al.* Data from: On the Origin of the Norwegian Lemming. Dryad Data Repository. doi:10.5061/dryad.jp8r1.

Larsen E, Gulliksen S, Lauritzen SE, *et al.* (1987) Cave stratigraphy in western Norway; multiple Weichselian glaciations and interstadial vertebrate fauna. *Boreas*, **16**, 267-292.

Lister AM (2004) The impact of Quaternary Ice Ages on mammalian evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 221-241.

Lopes JS, Balding D, Beaumont MA (2009) PopABC: a program to infer historical demographic parameters. *Bioinformatics*, **25**, 2747 - 2749.

Mangerud J, Gyllencreutz R, Lohne Ø, Svendsen JI (2011) Glacial history of Norway. In: *Quaternary glaciations - Extent and chronology. A closer look* (eds. Ehlers J, Gibbard PL, Hughes PD), pp. 279-298. Elsevier, Amsterdam.

Mayr E (1954) Change of genetic environment and evolution. In: *Evolution as a Process* (eds. Huxley J, Hardy AC, Ford EB), pp. 157-180. Allen & Unwin, London.

Mayr E (1963) *Animal species and evolution* Harvard University Press, Cambridge.

Nadachowski A (1982) *Late Quaternary rodents of Poland with special reference to morphotype dentition analysis of voles* Panstwowe Wydawnictwo Naukowe, Warszawa.

Navascués M, Depaulis F, Emerson BC (2010) Combining contemporary and ancient DNA in population genetic and phylogeographical studies. *Molecular Ecology Resources*, **10**, 760-772.

Navascués M, Emerson BC (2009) Elevated substitution rate estimates from ancient DNA: model violation and bias of Bayesian methods. *Molecular Ecology*, **18**, 4390-4397.

Nesje A, Dahl SO, Linge H, *et al.* (2007) The surface geometry of the Last Glacial Maximum ice sheet in the Andøya-Skånland region, northern Norway, constrained by surface exposure dating and clay mineralogy. *Boreas*, **36**, 227-239.

Nylander JAA (2004) MrModeltest v2, p. http://www.abc.se/~nylander/mrmodeltest2/mrmodeltest2.html

Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology & Evolution*, **13**, 502-506.

Parducci L, Jørgensen T, Tollefsrud MM, *et al.* (2012) Glacial survival of boreal trees in northern Scandinavia. *Science*, **335**, 1083-1086.

Prost S, Anderson CNK (2011) TempNet: a method to display statistical parsimony networks for heterochronous DNA sequence data. *Methods in Ecology and Evolution*, **2**, 663-667.

Rambaut A (2009) FigTree v1.3.1: Tree figure drawing tool, p. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummond AJ (2007) Tracer v1.4: MCMC trace analyses tool, p. http://tree.bio.ed.ac.uk/software/tracer/.

Rand AL (1948) Glaciation, an isolating factor in speciation. *Evolution*, **2**, 314-321.

Reimer P, Baillie M, Bard E, *et al.* (2009) IntCal09 and Marine09 radiocarbon age calibration curves, 0 - 50,000 years cal BP. *Radiocarbon* **51**, 1111-1150.

Ronquist F, Teslenko M, van der Mark P, *et al.* (2012) MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**, 539-542.

Shapiro B, Drummond AJ, Rambaut A, *et al.* (2004) Rise and fall of the Beringian steppe bison. *Science*, **306**, 1561-1565.

Stewart JR (2008) The progressive effect of the individualistic response of species to Quaternary climate change: an analysis of British mammalian faunas. *Quaternary Science Reviews*, **27**, 2499-2508.

Stewart JR, Lister AM, Barnes I, Dalén L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 661-671.

Stewart JR, Stringer CB (2012) Human evolution out of Africa: The role of refugia and climate change. *Science*, **335**, 1317-1321.

Svendsen JI, Alexanderson H, Astakhov VI, *et al.* (2004) Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*, **23**, 1229-1271.

Svensson A, Andersen KK, Bigler M, *et al.* (2006) The Greenland ice core chronology 2005, 15-42 ka. Part 2: comparison to other records. *Quaternary Science Reviews*, **25**, 3258.

Templeton AR (2009) Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. *Molecular Ecology*, **18**, 319-331.

Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR (1998) Improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*, **105**, 539-543.

Østbye E, Lauritzen S-E, Moe D, Østbye K (2006) Vertebrate remains in Holocene limestone cave sediments: faunal succession in the Sirijorda Cave, northern Norway. *Boreas*, **35**, 142-158.

Figure legends

Fig. 1 Temporal and geographic sample distribution. The Scandinavian Ice Sheet's approximate extent is shown for the four time periods, corresponding to (a) the Bø and Ålesund interstadials ~55 kyr to 35 kyr BP (without the short Skjonghelleren stadial ~ 39 kyr BP); (b) the Last Glacial Maximum ~30 kyr to 16 kyr BP; (c) the Younger Dryas stadial ~12 kyr BP (Mangerud *et al.* 2011; Svendsen *et al.* 2004); and (d) the Holocene. The Late Pleistocene sites that yielded successful ancient DNA sequences are illustrated with circles, which are filled blue at their respective time period. Sampling locations for modern specimens are shown as yellow (*L. lemmus*) and red (*L. sibiricus*) squares. The brown diamond represents the cave site from where early-mid Holocene samples were obtained.

Fig. 2 Schematic illustration of the two hypotheses regarding the evolutionary history of the Norwegian lemming (*Lemmus lemmus*). In scenario 1, the modern population is derived from a post-glacial colonisation from midlatitude Europe following the retreat of the Scandinavian Ice Sheet, whereas the population in scenario 2 has survived in Scandinavia since before the Last Glacial Maximum (LGM). In both scenarios, the lemming populations that inhabited midlatitude Europe ultimately went extinct during Holocene climate warming.

Fig. 3 Temporal statistical parsimony network. Haplotypes are temporally divided into the Holocene (including both modern and Holocene cave samples) and the Late Pleistocene, with empty circles indicating a haplotype that is missing in one temporal layer but is present in the other. Black dots represent missing haplotypes in the total data set. The number of individuals sharing a haplotype is reflected by its size. The dashed circles and connecting lines between the two temporal layers illustrate the absence of the Scandinavian haplogroup in the Late Pleistocene data set. The analysis is based on the partial data set. The

haplotypes are coloured according to their sample region, with numbers referring to the specific haplotype identifiers listed in Table S1 (Supporting Information).

Fig. 4 Bayesian phylogeny. Modern NW Russian *L. sibiricus* are shown in red, modern Scandinavian *L. lemmus* in yellow, early-mid Holocene Scandinavian samples in brown and Late Pleistocene European samples in blue. The ages of all ancient samples are shown in thousands (k) of years before present, with those from the Studennaya site referring to the calculated median ages obtained from BEAST. S = Scandinavia; R = Russia; R.P = Russian plains; R.U = Russian Urals; P = Poland; G = Germany; B = Belgium; E = England. Posterior probabilities of internal nodes above 0.8 are shown, with letters *A* to *H* referring to the estimated divergence times listed in Table 1. The analysis was performed in BEAST, using the partial data set and a mutation rate of 30 % Myr^{-1}.

Fig. 5 Posterior probability distributions for the population divergence time between Scandinavian lemmings and their closest glacial relatives, based on the Bayesian coalescent simulations of the partial data set, using three different mutation rates. The posterior distributions are truncated at 120 kyr BP, which represents the start of the Late Pleistocene glaciation. The uniform prior distribution is shown with grey bars.

Author contributions

V.K.L and D.E. designed and performed the DNA analyses, and V.K.L also computed population-genetic statistics, carried out Bayesian phylogenetic analyses and co-wrote the paper; E.S-C. performed and wrote the text on Bayesian coalescent simulations; N.A., A.N., D.C.K., M.G. and A.A. contributed with material and data; J.R.S. contributed with material and information, and helped interpret the data; L.D. conceived and designed the project and co-wrote the paper. All authors discussed the results and contributed to the preparation of the manuscript.

Data Accessibility

All sequences have been deposited in GenBank under the accession numbers JX483882-JX483939. The sequence alignments, as well as the input files and resulting tree files from BEAST and MrBayes have been deposited in the Dryad Data Repository, doi:10.5061/dryad.jp8r1.

Supporting Information

Additional supporting information can be found in the online version of this article.

Text S1 Materials and methods.

Table S1 All samples included in the study.

Table S2 Summary statistics employed in the Bayesian coalescent simulations.

Table S3 Genus-specific primers developed for the study.

Table S4 Genetic diversity within modern, Holocene and Late Pleistocene sample regions.

Table S5 Phylogenetic tree node ages.

Fig. S1 Bayesian coalescent simulation methodology.

Fig. S2 Temporal statistical parsimony network and Bayesian phylogeny constructed in BEAST.

Fig. S3 Bayesian phylogeny constructed in MrBayes.

Fig. S4 Posterior distributions of the employed summary statistics for each scenario separately.

Fig. S5 Posterior distributions of the employed summary statistics for both scenarios combined.

Fig. S6 Obtained posterior probability distributions for different effective population sizes.