



ORIGINAL
ARTICLE



Whole mitochondrial genomes illuminate ancient intercontinental dispersals of grey wolves (*Canis lupus*)

Stephan Koblmüller^{1,2}, Carles Vilà¹, Belen Lorente-Galdos³, Marc Dabad³, Oscar Ramirez³, Tomas Marques-Bonet³, Robert K. Wayne⁴ and Jennifer A. Leonard^{1*}

¹Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), 41092 Sevilla, Spain, ²Institute of Zoology, University of Graz, Universitätsplatz 2, A-8010 Graz, Austria, ³ICREA, Institut de Biologia Evolutiva (CSIC-UPF), Dr. Aiguader 88, 08003 Barcelona, Spain, ⁴Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-1606, USA

ABSTRACT

Aim Grey wolves (*Canis lupus*) are widespread across the Holarctic. Here, we test the previously proposed hypothesis that extant North American wolves originate from multiple waves of colonization from Asia. We also test the hypothesis that land connections have been important in the evolutionary history of other isolated wolf populations in Japan.

Location Holarctic.

Methods We analyse 105 previously published and newly obtained complete mitochondrial genomes from a geographically diverse sample of grey wolves and date critical branches in the phylogenetic tree. Phylogeographical hypotheses are tested in an approximate Bayesian computation approach.

Results We find that the mitogenomes of all living wolves in North America, including Mexican wolves, most likely derive from a single colonization event from Eurasia that expanded the grey wolf range into North America. This colonization occurred while a land bridge connected Eurasia and North America before the Cordillerian and Laurentide ice sheets fused in the Last Glacial Maximum, c. 23 ka, much more recent than predicted based on the fossil record. Pleistocene land bridges also facilitated the separate colonization of Hokkaido and the southern Japanese islands.

Main conclusions Extant wolf lineages in North America derive from wolves that migrated into North America coincident with the formation of the most recent land bridge with Eurasia. The maternal lineages from earlier Pleistocene American wolves are not represented in living American wolves, indicating that they left no descendants. The timing of colonization of North America, Hokkaido and the southern Japanese islands corresponds to the changes in land connectivity as a consequence of changing sea level.

Keywords

Beringia, canid, carnivore, mitogenome, molecular dating, Pleistocene dispersal

*Correspondence: Jennifer Leonard, Conservation and Evolutionary Genetics Group, Estación Biológica de Doñana (EBD-CSIC), 41092 Sevilla, Spain. E-mail: jleonard@ebd.csic.es

INTRODUCTION

The grey wolf (*Canis lupus* L. 1758) is the largest canid and, together with the red fox (*Vulpes vulpes* L. 1758), the most widely distributed wild carnivoran. The species is found throughout the Holarctic from the High Arctic to the deserts of the Middle East and southern North America, but was historically extirpated from parts of its original distribution, such as Japan and much of North America and Europe. It is most closely related to the coyote (*Canis latrans* Say 1819), a

smaller North American endemic canid. The fossil record supports the first appearance of wolves in Eurasia in the early to mid-Pleistocene, 1–2.5 Ma, and then in North America in the mid-Pleistocene (Kurtén, 1968; Tedford *et al.*, 2009).

The discovery of distinct phylogeographical patterns has been elusive in grey wolves. Early large-scale mitochondrial DNA studies found little support for large-scale patterns, and revealed instead a polyphyletic ancestry of North American wolves within the diversity of Eurasian wolves. However,

critical early branching events in the phylogenetic trees were poorly supported (Wayne *et al.*, 1992; Vilà *et al.*, 1999; Leonard *et al.*, 2005; Pilot *et al.*, 2010; Thalmann *et al.*, 2013). The somewhat divergent haplotypes restricted to Mexican wolves (*C. l. baileyi*) were hypothesized to represent an early invasion of wolves into North America (Wayne *et al.*, 1992; Leonard *et al.*, 2005; Thalmann *et al.*, 2013). Later research based on nuclear loci suggested monophyly of North American grey wolves (vonHoldt *et al.*, 2011). Despite considerable genetic research on grey wolves, it is still not clear when they colonized North America, or if there have been subsequent gene flow or colonization events. The mid-Pleistocene record of grey wolves in North America reveals a long, complex history. Conceivably, first colonization occurred during a period of significantly lowered global sea levels when the Bering land bridge was exposed and connected Asia to America, which occurred repeated during the Pleistocene and facilitated colonization of America by Eurasian terrestrial vertebrates (e.g. Brunhoff *et al.*, 2003; Davison *et al.*, 2011; Kutschera *et al.*, 2013). The Bering land bridge was not a narrow strip as the word 'bridge' may suggest, but covered an area of over one and a half million square kilometres and incorporated some present-day islands, such as Wrangel Island. The same fluctuations in sea level that exposed the Bering land bridge between Asia and North America also exposed other land bridges connecting, for example, Hokkaido and the southern Japanese islands to mainland Asia, allowing the expansion of wolves into these islands, where they survived until the beginning of the 20th century (Matsamura *et al.*, 2014).

Here, we analyse new and previously published mitogenomes of ancient and modern grey wolves from throughout their distribution to assess large scale phylogeographical patterns and statistically evaluate alternative scenarios regarding the colonization of North America by means of an approximate Bayesian computation (ABC) approach. The timeframe tested was constrained by the time-calibrated genealogy and reconstructions of Quaternary sea levels and the extent of ice sheets during the Last Glacial Maximum (LGM) (Siddall *et al.*, 2003; Miller *et al.*, 2005; Rabassa & Ponce, 2013). The total time period considered extends to the time of the most recent common ancestor for all of the wolf mitogenomes (80 ka; Thalmann *et al.*, 2013) and is much more recent than the time period suggested by the fossil record (Tedford *et al.*, 2009), but consistent with the recent divergence of Old World wolves suggested by nuclear data (Freedman *et al.*, 2014; Skoglund *et al.*, 2015). Consequently, we hypothesize that because colonization of North America was only possible during periods when the Bering land bridge was present, it must have occurred during the last emergence of the Bering land bridge (Fig. 1). The average water depth of the Bering Strait is 40–50 m. The land bridge was likely open when the sea level dropped more than 50 m below the current level, with a minimum relative sea level (*c.* –140 m) at the LGM, 20 ka (Siddall *et al.*, 2003; Miller *et al.*, 2005). The penultimate emergence of the land bridge was about

c. 130 ka and so is too ancient to be consistent with mitochondrial DNA sequence data. From *c.* 24–13 ka, the fusion of the Laurentide with the Cordilleran ice sheet prevented gene flow between Beringia (and hence Eurasia) and continental North America (e.g. Burns, 2010; Gowan, 2013; Rabassa & Ponce, 2013). After the ice sheets started to retreat (*c.* 13 ka), the Yukon corridor opened and potentially allowed for dispersal between Eurasia and continental North America until the Bering Strait closed in the wake of rising sea levels *c.* 10 ka, finally preventing intercontinental migration of grey wolves. We used this historic and geographical framework to assess if the extant diversity of American wolves is best explained by a single or multiple colonization(s) from Eurasia and to date the arrival of modern wolves to North America.

MATERIALS AND METHODS

Mitochondrial genomes

We analysed 105 sequences of the 12 protein coding genes located on the H-strand of the mitochondrial genome of grey wolves distributed across Eurasia and North America, but excluding the divergent lineages from India and the Himalayas (Sharma *et al.*, 2004; Aggarwal *et al.*, 2007) (Table 1). Among the studied sequences we included 10 from ancient (> 14,000 years old) Eurasian and North American wolves from Thalmann *et al.* (2013). Thirty-five new mitogenomes from modern wolves were generated for this study, either Sanger sequenced following Björnerfeldt *et al.* (2006) (*n* = 24) and aligned by eye in MEGA 6.05 (Tamura *et al.*, 2013), or reconstructed from whole-genome shotgun (WGS) sequencing (*n* = 11).

For each WGS sample, we captured mitochondrial reads from the whole set of sequenced reads by mapping the raw data against a reference wolf mitochondrial genome (DQ480508, Björnerfeldt *et al.*, 2006). We took advantage of the circularity of mtDNA in order to increase the number of captured reads at the extremes of the assembly. With this goal in mind, we applied a second round of mapping and aligned the reads to a modified sequence assembly, changing the origin of the reference assembly at the middle of the mtDNA (8 kbp from the start). To map, we used BWA 0.6.1 aligner with parameters `-n 6 -q 15` (Li & Durbin, 2009). We kept only high-quality paired-end reads by retaining read pairs that both were mapped and properly paired and had a mapping quality > 50 (with SAMTOOLS 0.1.18 `-f 2 -q 50`; Li *et al.*, 2009). We discarded paired-end pairs if at least one of both paired-ends had a median Phred quality score < 32. We used HAPASSEMBLER 1.1. (`-p Illumina -t 4 -d no -PHRED_OFFSET 33 -MIN_CONTIG_SIZE 1000 -EPSILON 0.05`) (Donmez & Brudno, 2011), a haplotype-specific genome assembly toolkit, to construct longer sequences from multiple overlapping reads (contigs) from the captured reads. The elevated mitochondrial coverage, calculated from the captured high-quality reads, decreases the efficiency of the

assembler. Therefore, we randomly selected reads to obtain a 350× coverage 20 times for each reference assembly (the standard and the one with the modified origin) and used those to construct contigs. This compensates for the random, low frequency representation of reads from nuclear copies of mitochondrial DNA (numts) assembled into the mitochondrial sequence. For each of the 40 times in which we iterated the read reduction and posterior *de novo* assembling, we oriented the contigs using local alignments to the corresponding reference (with BLAST, Altschul *et al.*, 1990) and joined them using MAFFT 6.903 (Katoh & Toh, 2008), incorporating uncalled base pairs (N's) in both existing gaps and sites that remain unresolved due to differences in overlapping contigs. This yields 40 mitochondrial assemblies per sample; the consensus sequence of which is the final mitochondrial sequence per individual.

Phylogenetic analyses and dating

To place our new and other published genomes in the time-calibrated mitogenomic genealogy of Thalmann *et al.* (2013) we used BEAST 1.8.0 (Drummond & Rambaut, 2007) and employed the same settings as used in that study [SRD06 model of molecular evolution (Shapiro *et al.*, 2006); strict molecular clock]. Mean calibrated radiocarbon ages or stratigraphically assigned ages were used as priors for all ancient samples (Table 1). Two independent Markov chain Monte Carlo chains were run for 100 million generations, with model parameters and trees sampled every 1000 generations. The first 10% of generations were discarded from each log and tree file as burn-in before the two chains were combined using LOGCOMBINER (available as part of the BEAST package). Chain convergence to stationarity for all model parameters was assessed in TRACER 1.6 (Rambaut *et al.*, 2013). The pooled post-burn-in effective sample sizes for all parameters exceeded 200, indicating that the pooled log file accurately represented the posterior distribution (Kuhner, 2009). Divergence times were derived from the pooled post-burn-in results and TREEANNOTATOR (available as part of the BEAST package) was used to compute a maximum-clade-credibility tree, which was visualized in FIGTREE 1.4.1 (Rambaut, 2014). Divergence times were calculated as mean node heights of the 95% highest posterior density (HPD) intervals. To ensure that our estimates were not biased because of misdated samples, we repeated the BEAST analysis several times with different sets of ancient samples.

Demographic analyses

Past population size trajectories for Eurasian and North American wolves were inferred assuming the Bayesian skygrid model (Gill *et al.*, 2013; 100 population size intervals, cut-off of 80 kyr – other cut-off values were evaluated but yielded identical results) as implemented in BEAST 1.8.0 and using the evolutionary and coalescent parameters used in the phylogenetic analyses.

Phylogeographical hypothesis testing with ABC

To test alternative hypotheses regarding the colonization of North America by grey wolves we employed an ABC approach as implemented in DIYABC 2.0 (Cornuet *et al.*, 2014). The timeframe tested was constrained by the time-calibrated genealogy inferred in BEAST, reconstructions of Quaternary sea levels and the extent of ice shields during the late Pleistocene (Fig. 2) (Siddall *et al.*, 2003; Miller *et al.*, 2005; Rabassa & Ponce, 2013). Considering these changes in the connectivity between continents four scenarios were evaluated (Fig. 2). Scenario 1 assumed that extant mitochondrial diversity of North American wolves was due to a single colonization wave prior to the fusion of the Laurentide and Cordilleran ice sheets (> 24 ka), which prevented migration between Beringia and continental North America. Scenario 2 assumed a single colonization event along the Yukon Corridor, once the Laurentide and Cordilleran Ice Shields separated after the LGM (< 13 ka). Based on the hypothesis that the Mexican wolves derive from an earlier colonization event than the other North American wolves, Scenarios 3 and 4 assumed two waves of colonization (one prior to, and one after the LGM) but with different admixture rates. Japanese samples were excluded from this analysis because they may represent back migration from North America to Eurasia (samples from Hokkaido, see Results), and this possibility is not contemplated in the model.

Priors were given uniform distributions, informed when possible by the results of the Bayesian skygrid analyses, the closing and opening of the Bering Strait, and the presence of ice sheets during the LGM (priors for the different scenarios evaluated are given in Table S1 in Appendix S1 in Supporting Information). As prior for the substitution rate, we employed the 95% HPD interval inferred from a Bayesian skygrid analysis in BEAST. As the SRD06 model of molecular evolution is not implemented in DIYABC and we wanted to refrain from unnecessary over-parameterization of the analy-

Figure 1 Dated phylogeny of wolves as inferred in BEAST. Global sea level changes are indicated on the top of the figure, coloured lines denote the drop in sea level required to expose a land connection and thus allow for colonization of North America (Bering Strait) or the Japanese islands from Asia (Soya and Korean Straits). The colours in the phylogeny indicate geographical origin of the haplotype: red: Europe; yellow: Near East; pink: India; violet: Asian Far East; green: Hokkaido (northern Japan); light blue: southern Japan; dark blue: North America. The ancient American haplotypes are represented by the truncated branches, which end at the radiocarbon date associated with the sequence. Nodal support in form of posterior probabilities of ≥ 0.99 , ≥ 0.95 , ≥ 0.90 and ≥ 0.70 is indicated by black, dark grey, light grey and white circles respectively, 95% confidence intervals for the age of the nodes are represented by gray bars. Time is indicated along the bottom of the figure in thousands of years before present (ka).

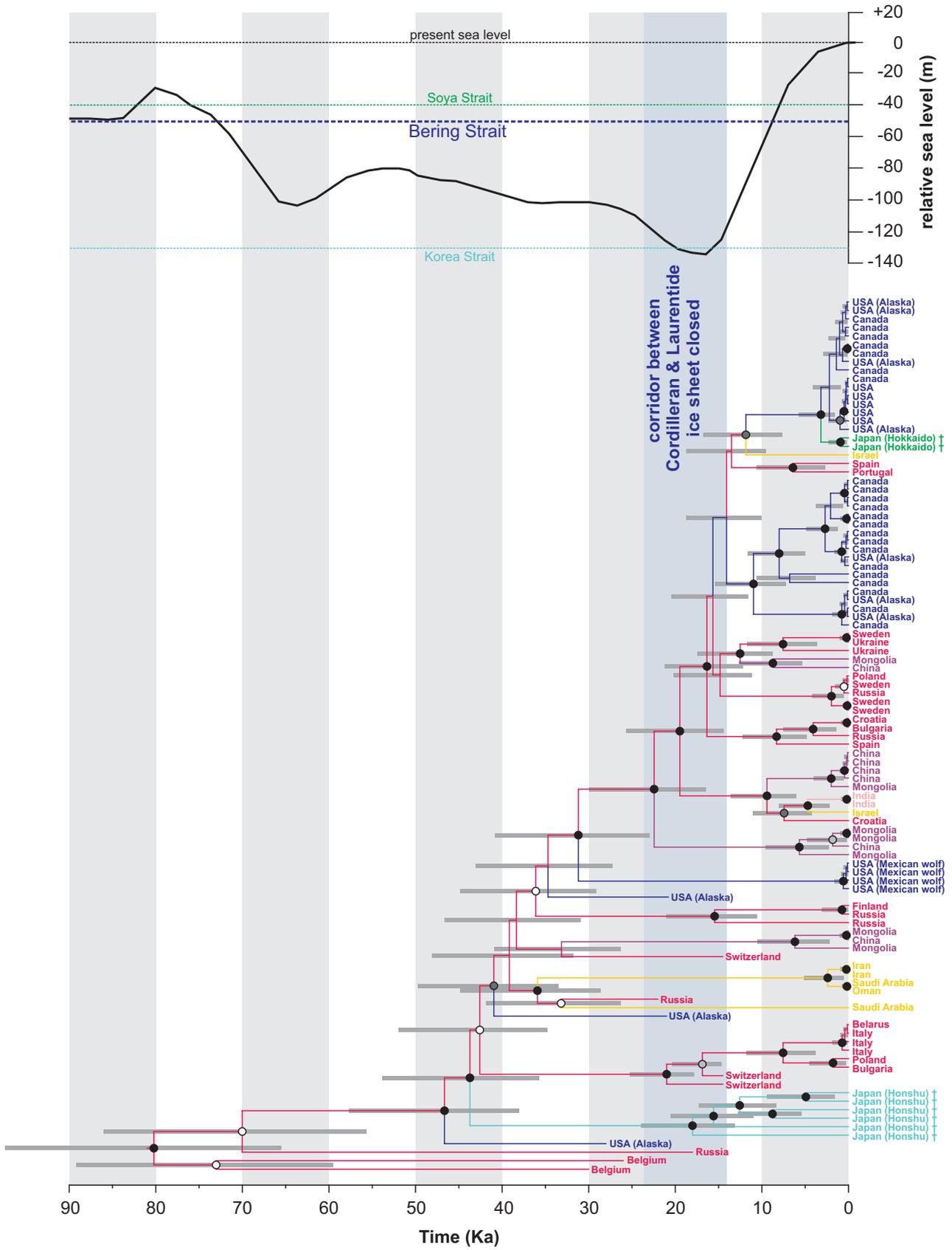


Table 1 List of samples used in this study with approximate age of ancient samples that were used to calibrate the molecular clock, or location for the sequences from modern animals.

GenBank	Sample ID ⁺	References	Approximate age (bp)
<i>Ancient samples</i>			
KF661078	Belgium_26000	Thalmann <i>et al.</i> (2013)	26,000
KF661080	Belgium_30000	Thalmann <i>et al.</i> (2013)	30,000
KF661081	Russia_18000	Thalmann <i>et al.</i> (2013)	18,000
KF661085	Russia_22000	Thalmann <i>et al.</i> (2013)	22,000
KF661087	Switzerland1_14500	Thalmann <i>et al.</i> (2013)	14,500
KF661088	Alaska_28000	Thalmann <i>et al.</i> (2013)	28,000
KF661089	Alaska_21000	Thalmann <i>et al.</i> , 2013;	21,000
KF661190	Alaska_20800	Thalmann <i>et al.</i> (2013)	20,800
KF661091	Switzerland2_14500	Thalmann <i>et al.</i> (2013)	14,500
KF661095	Switzerland3_14500	Thalmann <i>et al.</i> (2013)	14,500
<i>Modern samples</i>			
			Location
KU644667	YellowstoneA*	This study	Yellowstone N.P., USA
KU644665	YellowstoneB*	This study	Yellowstone N.P., USA
KU644663	YellowstoneC*	This study	Yellowstone N.P., USA
KU644664	MexicanWolfA*	This study	Captive Mexican wolf
KU644661	MexicanWolfB*	This study	Captive Mexican wolf
KU644669	China5*	This study	China
KU644666	India2*	This study	India
KU644671	Iran2*	This study	Iran
KU644662	Italy2*	This study	Italy
KU644668	Portugal*	This study	Portugal
KU644670	Spain2*	This study	Spain
KU696388	Bul729	This study	Bulgaria
KU696389	IT9	This study	Italy
KU696390	Bel13	This study	Belarus
KU696391	China282135	This study	China
KU696392	Mongolia1	This study	Mongolia
KU696393	Mongolia2	This study	Mongolia
KU696394	Mongolia3	This study	Mongolia
KU696395	Mongolia4	This study	Mongolia
KU696396	Mongolia5	This study	Mongolia
KU696397	CLUUkraine	This study	Ukraine
KU696398	WCRO7	This study	Croatia
KU696399	CVA354_Bulg	This study	Bulgaria
KU696400	Can1	This study	Alberta, Canada
KU696401	Can2	This study	Alberta, Canada
KU696402	Can3	This study	Alberta, Canada
KU696403	Can4	This study	Alberta, Canada
KU696404	Can5	This study	Alberta, Canada
KU696405	Can6	This study	Alberta, Canada
KU696406	Can7	This study	Alberta, Canada
KU696408	Can8	This study	Alberta, Canada
KU696407	Can9	This study	Alberta, Canada
KU696409	Can10	This study	NWT, Canada
KU696410	Can11	This study	Canada
KU696411	Ala1	This study	Alaska
EU789787	EU789787_China	Pang <i>et al.</i> (2009)	China
GQ374438	Mongolia	Unpublished	Mongolia
AM711902	AM711902_Sweden	Arnason <i>et al.</i> (2007)	Sweden
DQ480507	DQ480507_SaudiArab	Björnerfeldt <i>et al.</i> (2006)	Saudi Arabia
DQ480506	DQ480506_SaudiArab	Björnerfeldt <i>et al.</i> (2006)	Saudi Arabia
DQ480505	DQ480505_Spain	Björnerfeldt <i>et al.</i> (2006)	Spain
KF661038	Finland	Thalmann <i>et al.</i> (2013)	Finland
KF661039	Russia1	Thalmann <i>et al.</i> (2013)	Western Russia
KF661040	Sweden2	Thalmann <i>et al.</i> (2013)	Sweden
KF661041	China3	Thalmann <i>et al.</i> (2013)	China
KF661042	Israel1	Thalmann <i>et al.</i> (2013)	Israel

Table 1 Continued

GenBank	Sample ID ⁺	References	Approximate age (bp)
KF661043	India	Thalmann <i>et al.</i> (2013)	India
KF661044	Russia2	Thalmann <i>et al.</i> (2013)	Western Russia
KF661045	Poland1	Thalmann <i>et al.</i> (2013)	Poland
KF661046	Russia3	Thalmann <i>et al.</i> (2013)	Western Russia
KF661047	Ukraine	Thalmann <i>et al.</i> (2013)	Ukraine
KF661048	Italy	Thalmann <i>et al.</i> (2013)	Italy
KF661049	Poland2	Thalmann <i>et al.</i> (2013)	Poland
KF661050	Oman	Thalmann <i>et al.</i> (2013)	Oman
KF661051	Iran	Thalmann <i>et al.</i> (2013)	Iran
KF661052	Sweden3	Thalmann <i>et al.</i> (2013)	Sweden
KF661053	China4	Thalmann <i>et al.</i> (2013)	China
KF661054	Croatia	Thalmann <i>et al.</i> (2013)	Croatia
KF661055	Israel2	Thalmann <i>et al.</i> (2013)	Israel
KF661056	Canada1	Thalmann <i>et al.</i> (2013)	Canada
KF661057	Canada2	Thalmann <i>et al.</i> (2013)	Canada
KF661058	Alaska1	Thalmann <i>et al.</i> (2013)	Alaska
KF661059	Canada4	Thalmann <i>et al.</i> (2013)	Canada
KF661060	Mexico1	Thalmann <i>et al.</i> (2013)	Captive Mexican wolf
KF661061	Canada5	Thalmann <i>et al.</i> (2013)	Canada
KF661062	Canada6	Thalmann <i>et al.</i> (2013)	Canada
KF661063	Canada7	Thalmann <i>et al.</i> , 2013;	Canada
KF661064	USA1	Thalmann <i>et al.</i> (2013)	Contiguous U.S.A.
KF661065	Mexico2	Thalmann <i>et al.</i> (2013)	Captive Mexican wolf
KF661066	Alaska2	Thalmann <i>et al.</i> (2013)	Alaska
KF661067	Alaska3	Thalmann <i>et al.</i> (2013)	Alaska
KF661068	USA2	Thalmann <i>et al.</i> (2013)	Contiguous U.S.A.
KF661069	USA3	Thalmann <i>et al.</i> (2013)	Contiguous U.S.A.
KF661070	USA4	Thalmann <i>et al.</i> (2013)	Contiguous U.S.A.
KF661071	Alaska4	Thalmann <i>et al.</i> (2013)	Alaska
KF661072	Alaska5	Thalmann <i>et al.</i> (2013)	Alaska
KF661073	Alaska6	Thalmann <i>et al.</i> (2013)	Alaska
KF661074	Canada8	Thalmann <i>et al.</i> (2013)	Canada
KF661075	Canada9	Thalmann <i>et al.</i> (2013)	Canada
KF661076	Canada10	Thalmann <i>et al.</i> (2013)	Canada
KF661077	Canada11	Thalmann <i>et al.</i> (2013)	Canada
AB499818	AB499818_Jw229	Matsamura <i>et al.</i> (2014)	Japan
AB499819	AB499819_Jw233	Matsamura <i>et al.</i> (2014)	Japan
AB499820	AB499820_Jw235	Matsamura <i>et al.</i> (2014)	Japan
AB499821	AB499821_Jw237	Matsamura <i>et al.</i> (2014)	Japan
AB499822	AB499822_Jw239	Matsamura <i>et al.</i> (2014)	Japan
AB499823	AB499823_Jw240	Matsamura <i>et al.</i> (2014)	Japan
AB499824	AB499824_Jw257	Matsamura <i>et al.</i> (2014)	Japan
AB499825	AB499825_Jw258	Matsamura <i>et al.</i> (2014)	Japan
DQ480504	DQ480504_Sweden	Björnerfeldt <i>et al.</i> (2006)	Sweden
DQ480503	DQ480503_Russia	Björnerfeldt <i>et al.</i> (2006)	Russia
EU789788	EU789788_China	Pang <i>et al.</i> (2009)	China
KC461238	desertorum_China	Zhang <i>et al.</i> (2014)	China
KC896375	campestris_Mongoli	Zhang <i>et al.</i> (2014)	Mongolia

*Mitogenomes reconstructed from whole-genome shotgun (WGS) sequencing.

sis, we used a standard model of evolution (TN93+G, Tamura & Nei, 1993; as inferred by employing the Bayesian information criterion in jMODELTEST 0.1, Posada, 2008) on unpartitioned data. Topology and divergence times of the BEAST tree based on the TN93+G model (not shown) were almost identical to the original BEAST tree with the SRD06 model, such that we were confident that the use of the simpler model would not introduce a bias in the inferences

drawn from the ABC simulations. We assumed a generation time of 3 years (Mech & Seal, 1987) and employed wide and flat priors of migration rate (r) and N_e for all populations (see Table S1 in Appendix S1). The mutation rate prior μ spanned the 95% HPD interval inferred in BEAST [Uniform (1.85×10^7 – 3.00×10^7 per generation)]. We simulated one million generations for each scenario and calculated the following summary statistics for the observed and simulated

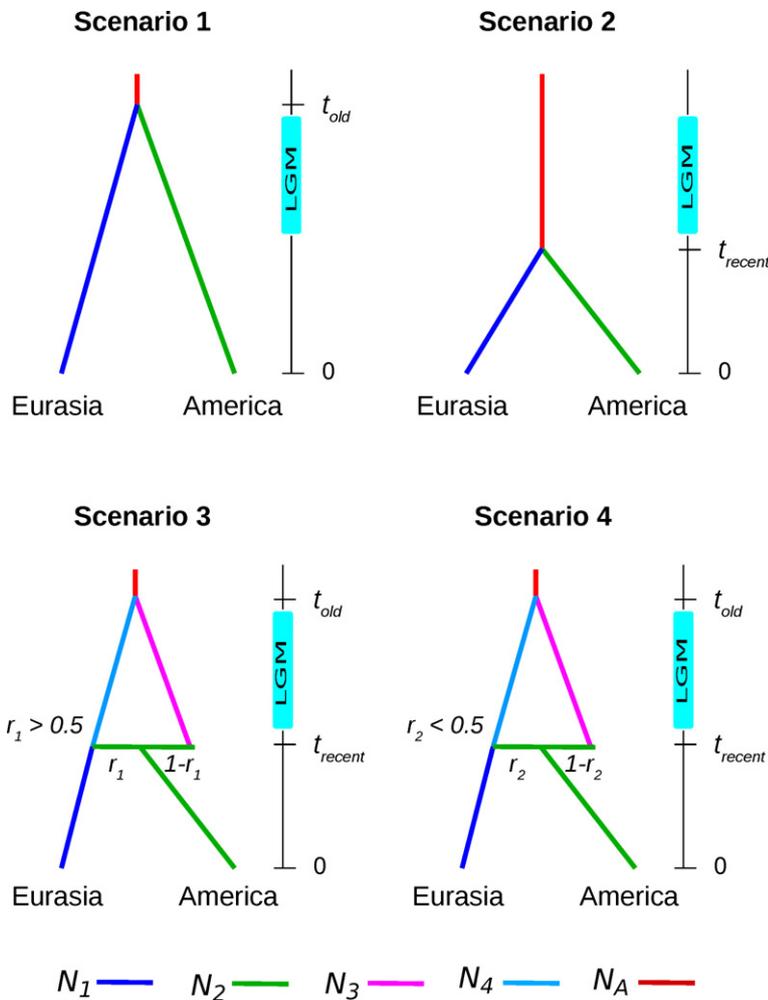


Figure 2 Graphical representation of the four alternative scenarios simulated in DIYABC. Scenario 1 assumes that North American wolves split from Eurasian wolves due to the closure of an ice-free corridor between the Cordilleran and Laurentide ice sheets during the Last Glacial Maximum (LGM) (*c.* 24 ka). Scenario 2 assumes that extant mitogenomic diversity in North American grey wolves traces back to colonization of continental North America after the re-opening of the corridor between the Cordilleran and Laurentide ice sheets at the end of the LGM (*c.* 13 ka). Scenarios 3 and 4 assume two colonization waves into continental North America (one prior to the LGM, one after the formation of the ice-free corridor at the end of the LGM), but with different admixture rates. t_{old} and t_{recent} indicate population splitting times prior to, or after the LGM respectively. N_1 , N_2 , N_3 , N_4 , N_A refer to the different effective population sizes through time; r_1 and r_2 indicate different admixture rates. For details, see Table S1 in Appendix S1.

data sets: number of haplotypes, number of segregating sites, mean of pairwise differences, variance of pairwise differences, Tajima's D , private segregating sites, mean number of the rarest nucleotide at segregating sites and variance of numbers of the rarest nucleotide at segregating sites. The posterior probability of each scenario was estimated using a logistic regression approach implemented in DIYABC.

For the best supported model, posterior distributions of this model's parameters were estimated using a local linear regression of the 1% simulated data sets closest to the observed data following logistic transformation of the parameter values (see Table S2 in Appendix S1; Cornuet *et al.*, 2008). Confidence in scenario choice was assessed by comparing the match between observed and simulated summary statistics and by calculating type I and type II error rates (Cornuet *et al.*, 2010). Therefore, 500 test data sets were simulated under each of the four scenarios, and posterior probabilities of the scenarios were assessed for each of the simulated data sets. Type I error rates were estimated by counting the proportion of times the 'true' (i.e. the best fitting scenario) scenario did not have the highest posterior probability among the alternative scenarios. Type II error rates were calculated as the proportion of times the 'true'

scenario had the highest posterior probability among the competing scenarios when the simulation was based on an alternative (sub-optimal) scenario. To evaluate whether the selected scenario fits with the observed data, we used the model checking option implemented in DIYABC (Cornuet *et al.*, 2010). Specifically, we simulated 10^3 data sets with parameters drawn from the posterior sample (10^4 values of the posterior distribution of parameters obtained from the 10^6 data sets simulated under the best supported scenario). The observed test statistics were compared with the 10^3 simulated test statistics, and their P -values were calculated and corrected for multiple comparisons following Benjamini & Hochberg (1995).

RESULTS

After exclusion of the control region which was only partially sequenced in previous studies that used Sanger sequencing methods (e.g. Leonard *et al.*, 2005, 2007), all sequences generated for this study were complete and of high-quality and consisted of the 12 protein coding genes of the H-strand only (see Appendix S2 for more information; GenBank accession numbers in Table 1).

Consistent with the mitogenomic analyses by *Thalmann et al. (2013)* and *Matsamura et al. (2014)* our time-calibrated phylogeny of grey wolf haplotypes (Fig. 1) revealed polyphyly of the North American wolves, and a basal position of the extinct Japanese wolves from the southern Japanese islands (*C. l. hodophilax*) among extant ‘modern’ wolves. Mexican wolves were divergent from other North American wolves, which formed two closely related clades. The extinct Ezo wolves from Hokkaido (*C. l. hattai*; $n = 2$) had sequences that were very different from southern Japanese wolves but very closely related to one of these American clades, and may indicate migration from America into Eurasia. Our estimate for the most recent common ancestor for all wolf sequences (*c.* 80 ka, Fig. 1) agrees with *Thalmann et al. (2013)*, but is considerably younger than in *Matsamura et al. (2014)*. Similarly, we find a much more recent common ancestor for the Japanese wolves than in *Matsamura et al. (2014)*. These differences result from a different calibration of the molecular clock as *Matsamura et al. (2014)* used the wolf-coyote split as a single calibration point, whereas we used mean calibrated radiocarbon ages or stratigraphically assigned ages of several ancient samples to calibrate the tree (following *Thalmann et al., 2013*). The use of internal points to calibrate the clock allowed us to largely circumvent the problems associated with a potential time dependency of the molecular clock (*Ho et al., 2005*) and obtain more robust and reliable divergence time estimates (*Ho et al., 2008*). Replicate BEAST analyses with different sets of ancient samples revealed consistent results across runs, indicating that inferred divergence time estimates were not driven by single misdated samples (not shown).

Reconstructions of past population size trajectories by means of Bayesian skygrid analyses suggest fairly constant population size through time for both the Eurasian and the American wolves until a few thousand years ago (Fig. 3), which is in stark contrast to the findings from genome wide SNP analyses, which suggest declining population sizes during the last *c.* 50 kyr for Eurasian and American wolf populations (*vonHoldt et al., 2011; Freedman et al., 2014*).

The ABC simulations provided best support for Scenario 1, a single colonization of North America during the last opening of the Bering Strait prior to the closure of the Yukon corridor in between the Laurentide and Cordilleran ice shields during the LGM (see Table S3; Fig. S1 in Appendix S1). The posterior probability of this scenario was 43.57% (95% CI: 42.93–44.21) and did not overlap with any other scenarios. Scenarios 2 and 4 received equal, but considerably lower support (20.41%, 95% CI: 19.88–20.94%; and 20.99, 95% CI: 20.50–21.47%). Scenario 3 was the least supported scenario (15.03%, 95% CI: 14.61–15.46%).

Power analyses based on test data sets simulated under the four scenarios indicated that the method has fairly high power to distinguish between the competing scenarios, considering that the scenarios tested are very similar and cover a very short time-scale (see Table S4 in Appendix S1). Type II errors were relatively large only in simulations of Scenario 4, which is to be expected given that this scenario is quite similar to the best scenario, as it assumes only low levels of immigration from Eurasia into continental North America after the emergence of an ice-free corridor at the end of the LGM (Fig. 2).

The implications of the ABC results are twofold. First, haplotypes unique to Mexican wolves do not trace back to

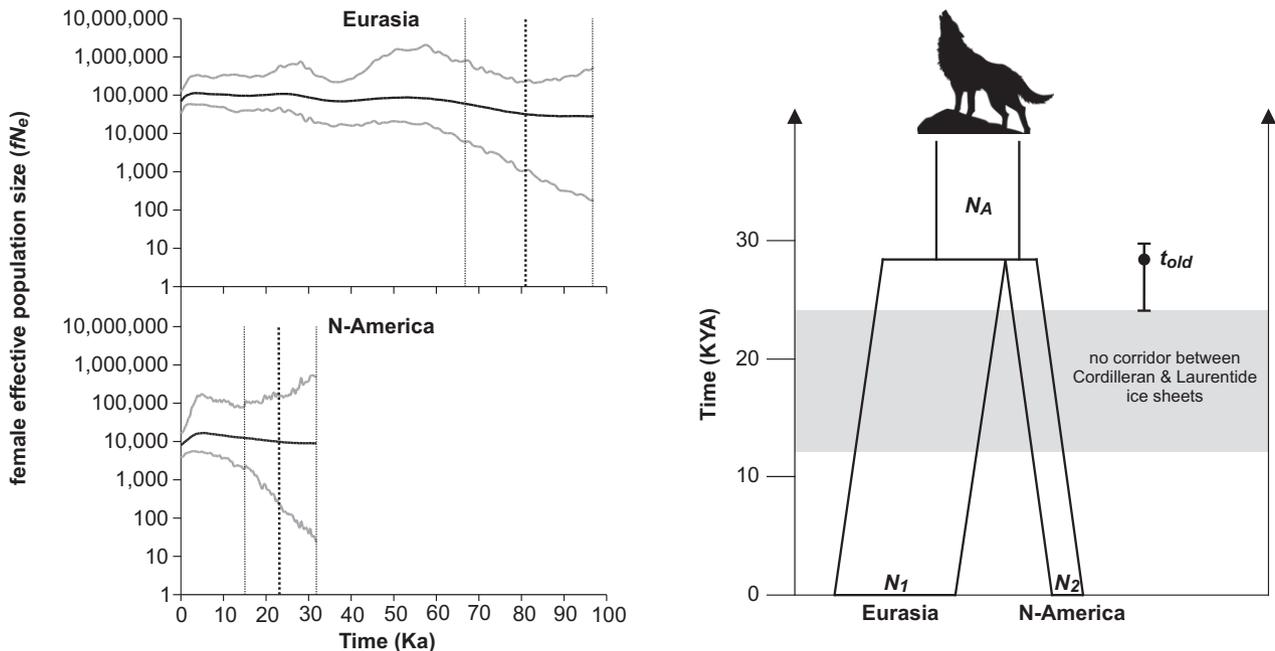


Figure 3 (a) Bayesian skygrid plots showing estimates of past effective population sizes of female wolves in Eurasia (top panel) and North America (bottom panel); (b) best scenario as inferred by DIVERABC (width of branches proportional to inferred modal values of population size; see Table S2 in Appendix S1).

an earlier colonization wave than other North American wolves. Second, mitochondrial diversity present in North American wolves was entirely replaced by haplotypes from wolves invading North America prior to the fusion of the Cordilleran and Laurentide ice sheets at the LGM (*c.* 24 ka). Our analysis suggests a coalescence of these invading haplotypes *c.* 80 ka, which is much more recent than the appearance of wolf-like forms in the fossil record.

DISCUSSION

The fossil record considerably pre-dates the inferred mitochondrial most recent common ancestor of the extant North American grey wolves (Tedford *et al.*, 2009), implying replacement of original haplotypes by novel, potentially more successful grey wolves or following continent-wide extinction. As no apparent gap is identified in the fossil record, replacement of the original haplotypes by novel ones suggests competitive displacement, especially when considering the possibility of admixture of ancient and invading wolves. Indeed, based on short fragments of the mitochondrial control region, complete and nearly complete replacement of a mitochondrial haplogroup was previously reported for Eastern Beringian (Alaskan) and Eurasian grey wolves respectively (Leonard *et al.*, 2007; Pilot *et al.*, 2010). Our mitogenome phylogeny also supports this hypothesis, as the ancient wolf haplotypes mostly occupy basal positions with respect to sequences from modern wolves (Fig. 1). Alternatively, ancient wolves may have suffered extinction, and been so rapidly replaced that the extinction was not detected in the fossil record.

A sea level drop of *c.* 50 m is required to expose the Bering land bridge, which occurred during the Wisconsin glacial period *c.* 70–10 ka. For most of this period, the land bridge was wider than during previous periods, and potentially sustained a panmictic wolf population ranging from Eurasia to North America, or at least allowed for substantial gene flow between the continents. Around 24 ka, gene flow between southern North America and Beringia (and hence Eurasia) ceased with the fusion of the Cordilleran and Laurentide ice sheets. Wolves survived both north and south of the ice sheets until the end of the Pleistocene. However, north of the ice sheets only a specialized hypercarnivorous wolf ecotype was found which subsequently went extinct (Leonard *et al.*, 2007). Therefore, as the ice sheets receded from 12–6 ka, southern wolves progressively re-colonized the formerly glaciated areas in North America. Consequently, all extant North American grey wolves descend from the wolves that were south of the ice sheets during the LGM. The differentiation between Mexican wolves and other wolf lineages in North America could be the result of divergence prior to invasion and/or population structure within North America driven by environmental factors (vonHoldt *et al.*, 2011; Leonard, 2014). Moreover, a large portion of the historic distribution of wolves in North America along with their unique genetic diversity was lost due to predator control efforts and

habitat change during the 20th century (Leonard *et al.*, 2005).

The dates of colonization of Hokkaido and the southern Japanese islands correspond with low sea level stands that imply the presence of land bridges, which was necessary for the colonization of these islands. With an average depth of 20–40 m, the Soya Strait separating Hokkaido from Sakhalin (mainland Russia) was mostly dry during the previous 75 kyr. Nonetheless, judging from the two mitogenomic sequences available (Matsamura *et al.*, 2014), Hokkaido appears to have been colonized only very recently (Fig. 1). Recent differential lineage sorting in Hokkaido and North America due to genetic drift might have resulted in more genetic similarity today than existed before the closure of the land bridge when Hokkaido was part of a much larger Holarctic population and contained a greater diversity of haplotypes. Unfortunately, the small sample size ($n = 2$) for the extinct Ezo wolf from Hokkaido does not allow for further evaluation of alternative scenarios. Although it appears that the southern Japanese islands were colonized much earlier during the last glacial cycle, the bathymetry of the Korea Strait and reconstructed sea level changes imply that southern Japan was colonized < 20 ka (Fig. 1). Indeed, the MRCA of the Japanese wolf dates to this time. That no close relatives of the southern Japanese wolf were found might be related to the apparent rapid turnover of mitochondrial haplogroups generally observed throughout the Old World distribution (Fig. 1; Pilot *et al.*, 2010). These findings suggest that a similar pattern would be found in the historic wolves of the British Isles, which were also connected to mainland Eurasia in the Pleistocene. Islands closer to mainland sources appear not to require land bridges for colonization (Muñoz-Fuentes *et al.*, 2010) as at least some populations of wolves will cross narrow stretches of open ocean by swimming (Darimont & Paquet, 2002; Muñoz-Fuentes *et al.*, 2009).

In conclusion, our results support the intercontinental dispersal of wolves during low sea level stands and suggest a far more recent coalescence of wolf mitogenomes than predicted by the fossil record. Thus, extant wolves represent a genetic veneer capping a complex history of turnover and evolution. We find that the grey wolves that now occupy North America represent a recent migration with the genetic record of previous migration events erased from the landscape. This surprisingly recent origin of modern North American grey wolf mitochondrial diversity contrasts with the greater antiquity of coyote haplotypes in North America (Vilà *et al.*, 1999; Koblmüller *et al.*, 2012).

ACKNOWLEDGEMENTS

We thank the individuals and agencies that assisted in sample collection. This project was supported by a starting grant from CSIC (J.A.L.), a grant from Fundació Barcelona Zoo de Barcelona and Ajuntament de Barcelona (O.R.) and a grant from the Programa de Captació del Conocimiento para Andalucía (C.V.).

REFERENCES

- Aggarwal, R.K., Kivisild, T., Ramadevi, J. & Singh, L. (2007) Mitochondrial DNA coding region sequences support the phylogenetic distinction of two Indian wolf species. *Journal of Zoological Systematics and Evolutionary Research*, **45**, 163–172.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Arnason, U., Gullberg, A., Janke, A. & Kullberg, M. (2007) Mitogenomic analyses of caniform relationships. *Molecular Phylogenetics and Evolution*, **45**, 863–874.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.
- Björnerfeldt, S., Webster, M.T. & Vilà, C. (2006) Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Research*, **16**, 990–994.
- Brunhoff, C., Galbreath, K.E., Fedorov, V.B., Cook, J.A. & Jaarola, M. (2003) Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. *Molecular Ecology*, **12**, 957–968.
- Burns, J.A. (2010) Mammalian faunal dynamics in Late Pleistocene, Alberta, Canada. *Quaternary International*, **217**, 37–42.
- Cornuet, J.-M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.M., Balding, D.J., Guillemaud, T. & Estoup, A. (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.
- Cornuet, J.-M., Ravigne, V. & Estoup, A. (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.
- Cornuet, J.-M., Pudlo, P., Veysier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.-M. & Estoup, A. (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism. DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.
- Darimont, C.T. & Paquet, P.C. (2002) The gray wolves, *Canis lupus*, of British Columbia's central and north coast: distribution and conservation assessment. *Canadian Field Naturalist*, **116**, 416–422.
- Davison, J., Ho, S.Y.W., Bray, S.C., Korsten, M., Tammeleht, E., Østbye, K., Østbye, E., Lauritzen, S.-E., Austin, J., Cooper, A. & Saarma, U. (2011) Late Quaternary biogeographic scenarios for the brown bear (*Ursus arctos*), a wild mammal model species. *Quaternary Science Reviews*, **30**, 418–430.
- Donmez, N. & Brudno, M. (2011) Hapsembler: an assembler for highly polymorphic genomes. *Research in computational molecular biology*. Vol. 6577 of Lecture Notes in Computer Science (ed. by V. Bafna and S. Sahinalp), pp. 38–52. Springer, Berlin/Heidelberg, Germany.
- Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Freedman, A.H., Gronau, I., Schweizer, R.N. *et al.* (2014) Genome sequencing highlights the dynamic early history of dogs. *PLoS Genetics*, **10**, e1004016.
- Gill, M.S., Lemey, N., Faria, R., Rambaut, A., Shapiro, B. & Suchard, M.A. (2013) Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution*, **30**, 713–724.
- Gowan, E.J. (2013) An assessment of the minimum timing of ice free conditions of the western Laurentide ice sheet. *Quaternary Science Review*, **75**, 100–113.
- Ho, S.Y., Phillips, M.J., Cooper, A. & Drummond, A.J. (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561–1568.
- Ho, S.Y., Saarma, U., Barnett, R., Halle, J. & Shapiro, B. (2008) The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS ONE*, **3**, e1615.
- vonHoldt, B.M., Earl, D.A., Knowles, J.C., Boyko, A.R., Parker, H., Geffen, E., Pilot, M., Jędrzejewski, W., Jędrzejewska, B., Sidorovich, V., Greco, C., Randi, E., Musiani, M., Kays, R., Bustamante, C.D., Ostrander, E., Novembre, J. & Wayne, R.K. (2011) A genome-wide perspective on the evolutionary history of enigmatic wolf-like canids. *Genome Research*, **21**, 1294–1305.
- Katoh, K. & Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.
- Koblmüller, S., Wayne, R.K. & Leonard, J.A. (2012) Impact of Quaternary climatic changes and interspecific competition on the demographic history of a highly mobile generalist carnivore, the coyote. *Biology Letters*, **8**, 644–647.
- Kuhner, M.K. (2009) Coalescent genealogy samples: windows into population history. *Trends in Ecology and Evolution*, **24**, 86–93.
- Kurtén, B. (1968) *Pleistocene mammals of Europe*. William Clowes and Sons, London.
- Kutschera, V.E., Lecomte, N., Janke, A., Selva, N., Sokolov, A.A., Haun, T., Steyer, K., Nowak, C. & Hailer, F. (2013) A range-wide synthesis and timeline for phylogeographic events in the red fox (*Vulpes vulpes*). *BMC Evolutionary Biology*, **13**, 114.
- Leonard, J.A. (2014) Ecology drives evolution in gray wolves. *Evolutionary Ecology Research*, **16**, 461–473.
- Leonard, J.A., Vilà, C. & Wayne, R.K. (2005) Legacy lost: genetic variability and population size of extirpated US gray wolves (*Canis lupus*). *Molecular Ecology*, **14**, 9–17.
- Leonard, J.A., Vilà, C., Fox-Dobbs, K., Koch, P.L., Wayne, R.K. & Van Valkenburgh, B. (2007) Megafaunal extinctions and the disappearance of a specialized wolf ecomorph. *Current Biology*, **17**, 1146–1150.
- Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & 1000 Genome Project Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
- Matsamura, S., Inoshima, Y. & Oshiguro, N. (2014) Reconstructing the colonization history of lost wolf lineages by the analysis of the mitochondrial genome. *Molecular Phylogenetics and Evolution*, **80**, 105–112.
- Mech, L.D. & Seal, U.S. (1987) Premature reproductive activity in wild wolves. *Journal of Mammalogy*, **68**, 871–873.
- Miller, K.G., Kominz, M.A., Browning, J.V., Wright, J.D., Mountain, G.S., Katz, M.E., Sugarman, P.J., Cramer, B.S., Christie-Blick, N. & Pekar, S.F. (2005) The Phanerozoic record of global sea-level change. *Science*, **310**, 1293–1298.
- Muñoz-Fuentes, V., Darimont, C.T., Wayne, R.K., Paquet, P. & Leonard, J.A. (2009) Ecological factors drive differentiation in wolves from British Columbia. *Journal of Biogeography*, **36**, 1516–1531.
- Muñoz-Fuentes, V., Darimont, C.T., Paquet, P. & Leonard, J.A. (2010) The genetic legacy of extirpation and re-colonization in Vancouver Island wolves. *Conservation Genetics*, **11**, 547–556.
- Pang, J.F., Kluetsch, C., Zou, X.-J., Zhang, A.-B., Luo, L.-Y., Angleby, H., Ardalán, A., Ekström, C., Sköllermo, A., Lundberg, J., Matsumara, S., Leitner, T., Zhang, Y.-P. & Savolainen, P. (2009) MtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 year ago, from numerous wolves. *Molecular Biology and Evolution*, **26**, 2849–2864.
- Pilot, M., Branicki, W., Jędrzejewski, W., Goszczyński, J., Jędrzejewska, B., Dykyy, I., Shkvyrya, M. & Tsingarska, E. (2010) Phylogeographic history of gray wolves in Europe. *BMC Evolutionary Biology*, **10**, 104.
- Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Rabassa, J. & Ponce, J.F. (2013) The Heinrich and Dansgaard-Oeschger climatic events during Marine Isotopic Stage 3: searching for appropriate times for human colonization of the America. *Quaternary International*, **299**, 94–105.
- Rambaut, A. (2014) *FigTree, version 1.4.1*. Computer program and documentation distributed by the author. Available at: <http://tree.bio.ed.ac.uk/software/figtree>.
- Rambaut, A., Suchard, M. & Drummond, A.J. (2013) *Tracer, version 1.6*. Computer program and documentation distributed by the author. Available at: <http://tree.bio.ed.ac.uk/software/tracer>.
- Shapiro, B., Rambaut, A. & Drummond, A.J. (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*, **23**, 7–9.
- Sharma, D.K., Maldonado, J.E., Jhala, Y.V. & Fleischer, R.C. (2004) Ancient wolf lineage in India. *Proceedings of the Royal Society B: Biological Sciences*, **272**(Suppl.), S1–S4.
- Siddall, M., Rohling, E.J., Almogi-Labin, A., Hemleben, C., Meischner, D., Schmelzer, I & Smeed, D.A. (2003) Sea-level fluctuations during the last glacial cycle. *Nature*, **423**, 853–858.
- Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. (2015) Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology*, **25**, 1515–1519.
- Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, **30**, 2725–2729.
- Tedford, R.H., Wang, X. & Taylor, B.E. (2009) Phylogenetic systematics of the North American fossil Caninae (Carnivora: Canidae). *Bulletin of the American Museum of Natural History*, **325**, 1–218.
- Thalmann, O., Shapiro, B., Cui, P. *et al.* (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, **342**, 871–874.
- Vilà, C., Amorim, I.R., Leonard, J.A., Posada, D., Castroviejo, J., Petrucci-Fonseca, F., Crandall, K.A., Ellegren, H. & Wayne, R.K. (1999) Mitochondrial DNA phylogeography and population history of the gray wolf *Canis lupus*. *Molecular Ecology*, **8**, 2089–21013.
- Wayne, R.K., Lehman, N., Allard, N.W. & Honeycutt, R.L. (1992) Mitochondrial DNA variability of the gray wolf: genetic consequences of population decline and habitat fragmentation. *Conservation Biology*, **6**, 559–569.
- Zhang, H., Zhang, J., Chen, L. & Liu, G. (2014) The complete mitochondrial genome of Chinese Shinjiang wolf. *Mitochondrial DNA*, **24**, 106–108.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Details of the approximate Bayesian computation (ABC) model testing and power analyses.

Appendix S2 Data for mitogenomes reconstructed from whole-genome shotgun sequencing data.

BIOSKETCH

This research was led by the Conservation and Evolutionary Genetics Group at the Estación Biológica de Doñana (www.consevol.org) where we have a strong line in the molecular ecology of large canids. This project was conceived of by J.A.L., the data were collected and analysed by S.K., the NGS sequences were constructed by B.L.G., M.D., O.R. and T.M.-B. The paper was written by S.K. and J.A.L. with support from all authors. Logistical support and materials were provided by J.A.L., C.V. and R.K.W.

Editor: Jim Provan