

Ancient DNA sequences point to a large loss of mitochondrial genetic diversity in the saiga antelope (*Saiga tatarica*) since the Pleistocene

PAULA F. CAMPOS,* TOMMY KRISTENSEN,* LUDOVIC ORLANDO,* ANDREI SHER,^{†1} MARINA V. KHOLODOVA,+ ANDERS GÖTHERSTRÖM,‡ MICHAEL HOFREITER,§² DOROTHÉE G. DRUCKER,¶ PAVEL KOSINTSEV,** ALEXEI TIKHONOV,†† GENNADY. F. BARYSHNIKOV,†† ESKE WILLERSLEV* and M. THOMAS P. GILBERT*

*Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark, †Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia, ‡Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyv. 18D, S-752 36 Uppsala, Sweden, §Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, D-04103 Leipzig, Germany, ¶Institut für Ur-und Frühgeschichte und Archäologie des Mittelalters, Naturwissenschaftliche Archäologie Universität Tübingen Rümelinstr. 23, D-72070 Tübingen, Germany, **Institute of Plant and Animal Ecology, Urals Branch of the Russian Academy of Sciences, 202 8th of March Street, Ekaterinburg 620144, Russia, ††Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, St. Petersburg 199034, Russia

Abstract

Prior to the Holocene, the range of the saiga antelope (*Saiga tatarica*) spanned from France to the Northwest Territories of Canada. Although its distribution subsequently contracted to the steppes of Central Asia, historical records indicate that it remained extremely abundant until the end of the Soviet Union, after which its populations were reduced by over 95%. We have analysed the mitochondrial control region sequence variation of 27 ancient and 38 modern specimens, to assay how the species' genetic diversity has changed since the Pleistocene. Phylogenetic analyses reveal the existence of two well-supported, and clearly distinct, clades of saiga. The first, spanning a time range from >49 500 ¹⁴C ybp to the present, comprises all the modern specimens and ancient samples from the Northern Urals, Middle Urals and Northeast Yakutia. The second clade is exclusive to the Northern Urals and includes samples dating from between 40 400 to 10 250 ¹⁴C ybp. Current genetic diversity is much lower than that present during the Pleistocene, an observation that data modelling using serial coalescent indicates cannot be explained by genetic drift in a population of constant size. Approximate Bayesian Computation analyses show the observed data is more compatible with a drastic population size reduction (c. 66–77%) following either a demographic bottleneck in the course of the Holocene or late Pleistocene, or a geographic fragmentation (followed by local extinction of one subpopulation) at the Holocene/Pleistocene transition.

Keywords: ancient DNA, LGM, mtDNA, Pleistocene, saiga

Received 5 August 2009; revision received 19 June 2010; accepted 24 June 2010

Correspondence: M Thomas P Gilbert, Fax: +45 35 32 13 00; E-mail: mtpgilbert@gmail.com

¹Deceased.

²Present address: Department of Biology, University of York, PO Box 373, York, YO10 5YW, UK.

Introduction

The major climatic oscillations that characterized the Pleistocene (2 Myr–10 000 years BP) had an important influence on the evolution and distribution of extinct and extant animal and plant taxa (and described in

2002; Hofreiter *et al.* 2004; Leonard *et al.* 2000; Orlando *et al.* 2002; Shapiro *et al.* 2004; Willerslev *et al.* 2003). During cold periods, temperate species distributions became fragmented and often limited to southern refugia, such as the Iberian, Italic and Balkan peninsulas in Europe (Taberlet *et al.* 1998), creating high levels of diversity and endemism in these areas (Hewitt 1996, 2000). For arctic and subarctic species on the other hand, the climatic cooling and associated spread of steppe tundra ecosystems presented ideal conditions for population and range expansions (Dalén *et al.* 2005; Stewart & Lister 2001).

One such subarctic species is the saiga antelope (*Saiga tatarica*, Linnaeus 1776), a fecund, nomadic, nonterritorial herding species that exhibits extremely large annual population movements between northern summer and southern winter ranges. Although presently confined to the dry-steppes and semi-deserts of Central Asia, palaeontological evidence indicates that it experienced its largest geographic distribution during the last glacial age. During this period, its east–west distribution stretched from England to the Northwest Territories of Canada and from the New Siberian Islands in the north to France and the Caucasus in the south (Sher 1974). While the Iberian and Apennine Peninsulas were never colonized, as the Pyrenees and the Alps seemed to constitute a barrier to migration, saiga was able to disperse into other mountainous areas, for example the Urals as far as 62°N (Kuzmina 1971). With the warming of the climate around the Pleistocene–Holocene transition, and

the subsequent replacement of the steppe tundra ecosystems by taiga forests, the range of the saiga contracted to their previous extent, the central Asiatic plains (Vereshchagin & Baryshnikov 1984). While the climate was probably the main cause of range contraction, in modern times, further contractions have been caused by the ever-growing human impact on the continent's steppe regions. Poaching and illegal trade in horns, uncontrolled hunting for meat, destruction of habitat and construction of irrigation channels, roads and other obstacles preventing natural dispersion and migration have all contributed to recent saiga population declines (Lushchekina & Struchkov 2001), putting the species on the brink of extinction. Their numbers have dropped more than 95%, from >1 000 000 to <50 000 in under two decades (Milner-Gulland *et al.* 2001) making this the most sudden and dramatic population crash of a large mammal in recent times. Moreover, most of the 50 000 animals remaining are females, which is leading to a further reproductive collapse in saiga antelope harems (Milner-Gulland *et al.* 2003).

The genus *Saiga* is nowadays represented by two recognized subspecies, *Saiga t. tatarica*, present in Kalmykia and Kazakhstan, and *Saiga t. mongolica*, that inhabits the basin of the Great Western Lakes in Outer Mongolia (Fig. 1). The latter differs slightly from the former in size and cranial bone structure (Bekenov *et al.* 1998). Mitochondrial DNA studies of modern populations reveal low levels of genetic diversity, especially in the Mongolian subspecies, and limited genetic structure

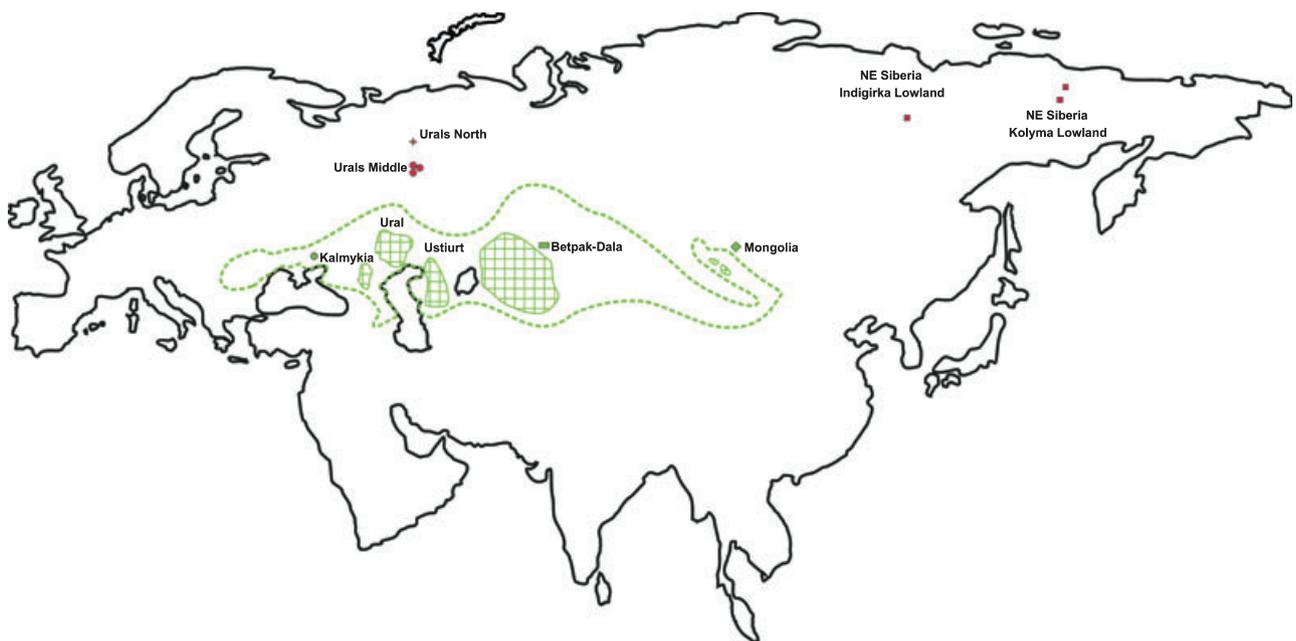


Fig. 1 Geographic origin of modern and ancient samples used in this study (ancient samples in red and modern samples in green). Dashed line represents the area where the species is known to have existed during historical times (17th–18th century).

between the other populations (Kholodova *et al.* 2006). Previous genetic analyses have also demonstrated a slight, but clear distinction between *S. t. mongolica* and *S. t. tatarica*, an observation used to support the current designation of *S. t. mongolica* as a subspecies rather than a separate species (Kholodova *et al.* 2006). The overall low mtDNA genetic diversity of the species is likely explained by either a long period of very small population size or a severe bottleneck in very recent times, although which explanation applies to the saiga has been unclear. As fossil samples are available, however, one way to solve this question is through the use of ancient DNA (aDNA).

Ancient DNA has been extensively used in the past few years to address evolutionary, ecological and palaeogeographical questions (e.g. Campos *et al.* 2010; Gilbert *et al.* 2008; Hofreiter *et al.* 2002; Leonard *et al.* 2005; Shepherd & Lambert 2008; Valdiosera *et al.* 2008). Its capacity to record evolutionary processes within populations over geologically significant timescales provides the means to examine past population variation. Lost lineages or haplotypes, that would never be found when only modern DNA analyses are conducted, can be revealed by using aDNA (e.g. Calvignac *et al.* 2008). Ancient DNA is also a powerful tool for analysing how organisms have responded to past climate and environmental changes (Shapiro *et al.* 2004). Geographic and temporal divisions apparent in the genetic data can be correlated to specific ecological events, such as advancing ice sheets, volcanic eruptions or human colonization. In light of the above, in this study, we have used ancient DNA to investigate the evolution and demographic history of Pleistocene saiga and to assess the level of genetic variation that has been lost during the last 50 000 years.

Material and methods

DNA extraction and sequencing

Ancient material. A total of 122 saiga bones, teeth and horns were obtained from across its past geographical range (Table S2). Samples included both *S. tatarica* and several individuals identified as *Saiga borealis*, a term used by some to describe Pleistocene saiga from northern Eurasia (Baryshnikov & Tikhonov 1994). DNA was extracted from 0.02 to 0.05 g of bone powder using a silica-column-based method (Yang *et al.* 1998) in a specialized ancient DNA facility that is physically isolated from the location where post-PCR-amplified products are manipulated. To guard against contamination, all reagents used were molecular biology grade, and blank PCR and extraction controls were used to monitor for contamination. A 280 bp subsection of the mitochon-

drial control region was PCR amplified using a combination of the overlapping primers listed in Table S1. The overall poor condition of the samples limited the length of the amplification products to under 150 bp, including primers. PCR amplification was performed in 25 μ L volumes, using 1 \times PCR buffer, 2 mM of MgSO₄, 1.6 mg/mL Bovine Serum Albumin (BSA), 0.4 μ M of each primer, 0.1 mM of dNTPs and 1 μ L of High Fidelity Platinum Taq (Invitrogen). Cycling conditions were 94 °C for 2 min followed by 50 cycles of 94 °C for 30 s, 50–60 °C for 30 s and 72 °C for 45 s with a final extension step at 72 °C for 7 min. All PCR products were cloned using the TOPO TA cloning kit (Invitrogen). A minimum of six clones was sequenced for each fragment. The multiple overlapping primer pairs resulted in a substantial degree of sequence replication for each sample. Furthermore, the 27 samples that successfully yielded the full 280-bp fragment (in overlapping PCRs) were subsequently re-extracted, re-amplified and directly sequenced to ensure against erroneous sequences that might be caused by DNA damage or contamination. DNA sequences were edited and aligned using SEQUENCHER 4.7 (Gene Codes Corporation).

Modern samples. DNA was extracted from a total of 38 modern saiga hair samples (using multiple hair shafts and roots per individual), sampled from three populations (Kalmykia, Betpak Dala-Kazakhstan and Mongolia) representing members of both modern subspecies. DNA was extracted using organic solvents following Sambrook *et al.* (1989). The complete mtDNA control region was amplified from these samples using primers tRNA^F and tRNA^R (see Supporting Information). PCR amplification was performed in 20 μ L volumes, using 1 \times PCR buffer, 2 mM of MgCl₂, 0.4 μ M of each primer, 0.2 mM of dNTPs and 0.4 U of Smart Taq (Dialat). Cycling conditions were 94 °C for 3 min; 35 cycles of 94 °C for 30 s, 54 °C for 30 s and 72 °C for 1 min followed by 72 °C for 7 min. PCR products were sequenced in both directions on an ABI-3730 (Applied Biosystems) using the BigDye Terminator Cycle Sequencing kit version 3.1. The sequences were edited and the forward and reverse strands aligned using SEQUENCHER 4.7.

Phylogenetic analyses

Phylogenetic relationships were estimated using Bayesian and Maximum Likelihood (ML) algorithms. The best substitution model was selected using the Aikake Information Criterion using MODELTEST v. 3.7 (Posada & Crandall 1998), and ML analyses were performed with PHYML (Guindon & Gascuel 2003) using the online interface (<http://www.atgc-montpellier.fr/phyml/>). The

transition/transversion ratio, the proportion of invariable sites and the gamma distribution were estimated, and the starting tree was determined by a BioNJ analysis of the data set (default settings). Using optimization options (for branch length and tree topology), 10 000 bootstrap replicates were performed. Markov Chain Monte Carlo sampling was performed as implemented in the phylogenetic analysis software BEAST version 1.4.6 (Drummond & Rambaut 2007) under the HKY + G nucleotide substitution model, a strict molecular clock model, using radiocarbon dates as tip calibration points, along with the Bayesian skyline plot demographic model (Drummond *et al.* 2005). Two MCMC chains were run for 20 million iterations, with parameters written to file every 1000 iterations. Convergence and mixing were evaluated using Tracer v1.4.1 (Rambaut & Drummond 2007). The first quarter (25%) of the trees was discarded as burn-in, and the remaining trees were summarized with the majority-rule consensus approach, using posterior probability (pp) as a measure of clade support. A Bayesian Skyline reconstruction was subsequently performed on the data using Tracer v1.4.1 (Drummond *et al.* 2005; Rambaut & Drummond 2007) to investigate whether any signal of population changes through time could be observed.

Population genetic analyses

The number of segregating sites (S) and haplotypes (h), and haplotype (Hd) and nucleotide (π) diversities (Tajima 1983) were calculated using DNASP version 4.9 (Table 1) (Rozas *et al.* 2003). Different data set partitions were considered respective to the age of the samples (ancient, modern or both combined) and/or the

number of sites showing missing data (Table 1). To test for neutrality and demographic expansions Fu's F_s (Fu 1997), Fu and Li's F^* and D^* statistics (Fu & Li 1993) and Tajima's D were calculated from the combined ancient and modern samples in DNASP version 4.9 (Rozas *et al.* 2003). The significance of Fu's F_s estimate was assessed by performing coalescent simulations (1000 replicates) in DNASP (Rozas *et al.* 2003). For mitochondrial data (i.e. haploid), a rough estimate of the population size can be provided by θ_π using the formula $2N_e\mu$ (where N_e is the population size and μ is the mutation rate) if no time difference among sequences (heterochrony) is present in the data set. Using a mutation rate of 52.042% mutations per site per million year as calculated in the BEAST analyses [95% highest posterior density (HPD) 79.201–24.173%], we calculated the effective population size (N_e) of the observed modern samples as c. 13 500–15 000 individuals. In situations where a substantial time structure is present in a data set, the population size can still be estimated from θ_π but using an unbiased estimator correcting for the average time difference between pairs of sequences (and described in Depaulis *et al.* 2009). This leads to the estimates of c. 22 000–52 000 individuals, for the ancient, and total sequence data sets, respectively (Table 1). As sample SA-044 had an infinite radiocarbon date, it was excluded from the data set for the latter analyses and further described serial coalescent simulations.

Serial coalescent modelling

The interpretation of patterns observed in ancient genetic data using only summary statistics and phylogenetic patterns can be misleading if the heterochronous

Table 1 Genetic diversity measures for different data set partition of the mtDNA control region

	#Seq.	#Sites	S	h	Hd	π	π^{**}	$N_e(\pi^{**})$	%Bias	F_s	F^*	D^*	D
Modern	38	250	25	12	0.785 ± 0.063	0.01385 ± 0.00247	0.01385	13307	0	-1.755	-2.318	-2.228	-1.428
	38	201	22	11	0.781 ± 0.062	0.01574 ± 0.00266	0.01574	15122	0	-1.401	-2.163	-2.078	-1.335
	38	246	25	12	0.785 ± 0.063	0.01407 ± 0.00251	0.01407	13518	0	-1.755	-2.318	-2.228	-1.428
Ancient	27	224	39	20	0.954 ± 0.028	0.05196 ± 0.00501	–	–	–	-4.227	0.011	-0.107	0.256
£	26	201	35	18	0.948 ± 0.029	0.05422 ± 0.00498	0.05016	48345	7.49%	-2.953	0.045	-0.108	0.356
£	26	224	39	19	0.951 ± 0.030	0.05251 ± 0.00494	0.04887	52330	6.93%	-3.532	0.029	-0.085	0.261
\$	22	272	37	19	0.987 ± 0.018	0.04036 ± 0.00527	–	–	–	-6.399	0.196	0.102	0.317
£\$	21	246	34	17	0.981 ± 0.020	0.04286 ± 0.00557	0.03882	37474	9.43%	-4.372	0.425	0.326	0.452
£\$	21	272	37	18	0.986 ± 0.019	0.04104 ± 0.00523	0.03738	39714	8.92%	-5.505	0.220	0.125	0.335
Combined	65	201	46	30	0.918 ± 0.026	0.03716 ± 0.00478	–	–	–	-8.641	-1.623	-1.593	-0.972
£	64	201	46	29	0.915 ± 0.026	0.03731 ± 0.00483	0.03187	30625	14.58%	-7.714	-1.611	-1.577	-0.972
\$	60	246	47	30	0.912 ± 0.030	0.02770 ± 0.00415	–	–	–	-10.888	-1.955	-1.882	-1.229
£\$	59	246	47	29	0.909 ± 0.031	0.02780 ± 0.00420	0.02272	21830	18.27%	-9.825	-1.942	-1.862	-1.231

The summary statistics used for serial coalescent simulations and Approximate Bayesian Computation are outlined in grey. N , number of samples; S , number of polymorphic sites; h , number of haplotypes; Hd , haplotype diversity; π , nucleotide diversity. **, unbiased estimates considering the time structure of the data set; %Bias, extent of the bias in the estimate of π , relative to the value uncorrected for heterochrony; \$, removing sequences with missing data; £, Removing Sample SA-044 with infinite date.

nature of the data is not taken into account (Depaulis *et al.* 2009). To help account for this, we used Bayesian Serial SimCoal (SSC, Anderson *et al.* 2005; as implemented at <http://www.stanford.edu/group/hadlylab/ssc/BayeSSC.htm>) to model several scenarios against which the observed data could be compared. We initially used BSSC to model whether the observed number of segregating sites and haplotypes, and haplotype and nucleotide diversities (see grey rows in Table 1) could be explained solely as a result of our sampling scheme in a single, constant sized, population (Null model; Table 2). In doing this, we treated all modern saiga as a single population, as both their currently fragmented distribution is a very recent phenomenon, and as there is no evidence that the modern populations and subspecies are distinct at the mtDNA level. Radiocarbon dates of the ancient samples were converted to calendar years and generation times set to 1 year in accordance with saiga life history (Sokolov & Zhirnov 1998).

We then used BSSC to explore further population histories, for example whether the observed data could be modelled by an initially constant sized population undergoing a single sudden (Fig. 2, model B₁), or gradual, bottleneck (exponential decrease, Fig. 2 model B₂). The time of the bottleneck was sampled from a uniform prior ranging from 100 to 40 000 years BP. This time period includes the first large-scale contact between humans and saiga, major climatic changes at the end of

the Pleistocene and the age range of the ancient samples (c. 10 000–50 000 years BP). Additional simulations based on a larger prior 100 to 75 000 years BP were found to have no effect on posterior distributions (data not shown). After 200 000 simulations, the posterior distributions of the parameters were estimated according to a Approximate Bayesian Computation procedure (ABC, Beaumont *et al.* 2002) using the observed values of S , h , H_d and π , the locfit, akima, lattice R packages (R Development Core Team 2008) and the reject function available at <http://www.stanford.edu/group/hadlylab/ssc/eval.r>. Briefly, a Euclidean distance between the simulated and observed summary statistics is computed for each simulation, and only simulations showing Euclidean distances within a 0.2 tolerance interval (delta value) were conserved, resulting in a number of c. 1500–8500 simulations per model (Table 2). Two hundred thousand further simulations were performed per model using the previous posterior distributions of the parameters as priors, and a two-sided P -value for each model was computed using the C statistics described in (Fabre *et al.* 2009; Ghirotto *et al.* 2009; Voight *et al.* 2005). The posterior probability of each model (i.e. marginal likelihood) was further estimated using a rejection method based on 200 000 simulations and counting up the proportion of simulated points within a tolerance region of 0.2 (Pritchard *et al.* 2000). The two demographic scenarios were then contrasted to the null model using Bayes Factors. In addition, we relied on

Table 2 Prior and Posterior Distributions, Empirical and Marginal likelihood for population models B₁, B₂, S₁ and S₂. The Null model corresponds to a single panmictic population with constant size. Population models B₁, B₂, S₁ and S₂ are described in Fig. 2

		201 sites					246 sites				
		Post. Distrib.	#Acc.	EL (%)	PP model (%)	BF	Post. Distrib.	#Acc.	EL (%)	PP model (%)	BF
Null	N0	N:42 900, 11 700	1678	16.393	1.095	–	N:36 200, 10 300	3175	21.202	1.433	–
B1	N0	N:20 200, 5840	6605	95.340	24.569	22.44	N:19 300, 5980	5697	93.051	5.268	4.81
	TB	N:18 300, 10 600					N:17 200, 11 100				
	IB	G:4.45, 0.977					G:3.58, 0.954				
B2	N0	N:22 000, 5240	2894	22.950	4.360	3.98	N:22 100, 5200	3846	21.554	5.107	4.66
	TB	G:1.7, 0.000113					G:1.5, 0.0000931				
	N1	N:67 700, 17 800					N:59 700, 18 400				
S1	N0	N:28 000, 7130	5634	36.771	3.895	3.56	N:31 400, 7380	2178	26.977	2.785	2.54
S2	N0	N:19 100, 7190	6702	84.615	13.187	12.04	N:22 400, 8110	8585	66.066	9.687	8.85
	X	G:3.23, 0.94					G:3.18, 0.923				

The posterior distributions of model parameters have been estimated after Approximate Bayesian Computation (ABC) following 200 000 simulations using 0.2 as tolerance. Model posterior distributions (marginal likelihood) have been estimated after a second round of ABC following 200 000 simulations, using 0.2 as tolerance and the posterior distributions of model parameters as priors. Bayes Factors are reported to estimate model support (compared to the Null model). Empirical likelihood (EL) is defined as the two-sided P -value of the C statistics described in Voight *et al.* (2005) following 200 000 simulations.

G: X , Y , gamma distribution with X and Y as scale and shape parameters; N: X , Y , normal distribution with X and Y as mean and variance parameters; U: X , Y , uniform distribution within a minimum (X), maximum (Y) range.

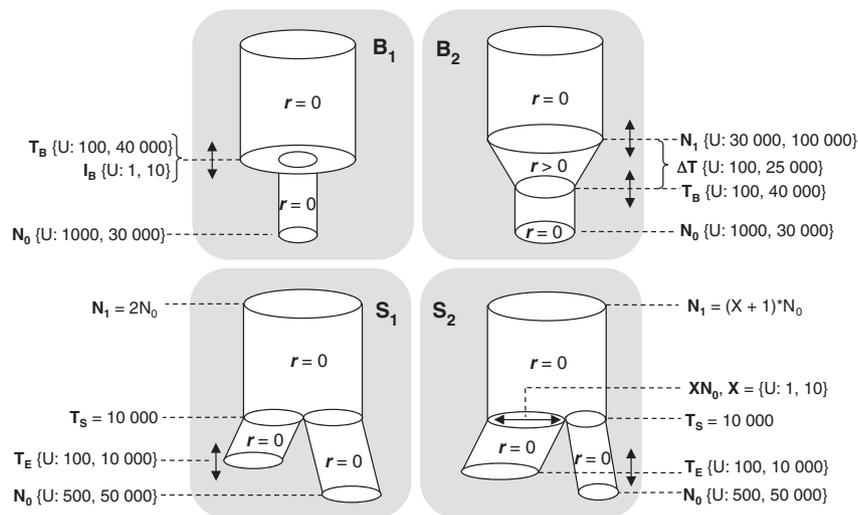


Fig. 2 Population models used in BSSC simulations followed by Approximate Bayesian Computation. Model B₁ corresponds to a single population that experienced a sudden demographic bottleneck T_B generations ago. The overall effective size was reduced by a factor I_B . Model B₂ is a modification from model B₁ where the demographic decline (from N_0 to N_1) is assumed as an exponential decrease that lasted for ΔT generations. In model S₁, the saiga population splitted into two subpopulations of equal sizes 10 000 generations ago. Then, one of the subpopulations became extinct T_E generations ago while the second gave rise to the modern saiga population. Model S₂ is a modification from model S₁, allowing for uneven population sizes among subpopulations (the ratio of the population size between the subpopulation that will become extinct to the modern one is noted X ; i.e. $X/(X + 1)$ of the past population will become extinct sometime after fragmentation).

the same simulation framework to investigate whether a population split that would have occurred at the end of the Pleistocene (combined with a further extinction of one subpopulation; Fig. 2, models S₁ and S₂) may have caused the observed loss of genetic diversity.

Results

Although 51 of the 122 ancient samples studied yielded DNA, complete or almost complete fragments (247–277 bp) could only be recovered from 27 samples. After repeated attempts at obtaining the missing sections using a range of alternate primers failed to fill the gaps in the remaining 24 samples, the samples with incomplete sequences were not used in the analysis. The completed 27 sequences, as well as the sequences obtained for the modern samples are deposited in GenBank with accession numbers HM625915–HM625978. We are confident that the control region sequences are of mitochondrial origin and are not nuclear-encoded copies of mitochondrial sequences (numts), as they were consistent between fragments generated with different primer pairs and replicable between amplifications when the same primer pair was used. Furthermore, no alternative sequence was observed among the clones. The final data set of 65 samples consists of 38 modern and 27 ancient specimens. To provide additional insights into the data, we obtained radiocarbon dates for the 27 ancient specimens, through the commercial accelerator

mass spectrometry (AMS) ¹⁴C dating facility at the Department of Physics and Astronomy, University of Århus. The samples range in age between >49 500 and 10 250 ¹⁴C years before present (ybp), thus are all from the Pleistocene period (Table S2, Fig. S1).

The combined modern and ancient genetic data set consists of 201 nucleotides in all samples (246 excluding sample SA-044) and 46 (47) segregating sites, which define 30 different haplotypes, none of which are shared between the ancient and the modern samples (Table 1). Once sites showing missing data are removed from the analyses, the 27 ancient sequences constitute a 224 nucleotide-long alignment, and define 20 different haplotypes, 75% of which are unique (Table 1). The different geographic locations also are distinct genetically, with only two haplotypes shared between them (one in both the north and middle Urals, and the other in both middle Urals and Northeast Yakutia). In contrast, the 38 modern samples contain only 12 different haplotypes (Table 1). The most frequent haplotype was found in two of the modern populations: 15 times in Kalmykia and two times in Kazakhstan. Consistent with the above observations, haplotype (Hd) and nucleotide (π) diversities were significantly higher for the ancient samples than for the modern samples, even after having corrected the latter for heterochrony (Table 1). Tajima's D -values and Fu and Li F^* and D^* statistics for the combined ancient plus modern data set were not significantly different from zero, which confirms that all

mutations are selectively neutral. Note that when a moderate time structure is present within a data set, Tajima's *D* statistics tend to be shifted toward negative values under a null neutral model, resulting in higher rates of false positive rejections of neutrality (Depaulis *et al.* 2009). Because the null model was accepted here, our conclusion is conservative, regardless of heterochrony-related biases. Similarly, Fu's *F_s* index was negative but not significant (Table 1).

The Hasegawa–Kishino–Yano + gamma + invariant sites (HKY + G + I) model, which incorporates different rates for transitions and transversions, rate variation across sites and a proportion of invariable sites, was used to generate Bayesian posterior probabilities. The

same model was used to generate the ML tree and bootstraps. The phylogenetic analyses revealed two distinct clades. Although representing both subspecies and three geographic regions, there is no evidence that the modern samples form distinct groups at the mtDNA level. Furthermore, they cluster into one of the clades, along with ancient samples from the Urals (north and middle) and Northeast Yakutia (Fig. 3). This clade ranges in age from >49 500 ¹⁴C ybp to the present. Two *Saiga borealis* specimens successfully yielded DNA sequences that place them in clade 1 along with all the other modern and some of the ancient samples. As such, the data suggest that *S. borealis* does not constitute a distinct subspecies or species. The second clade

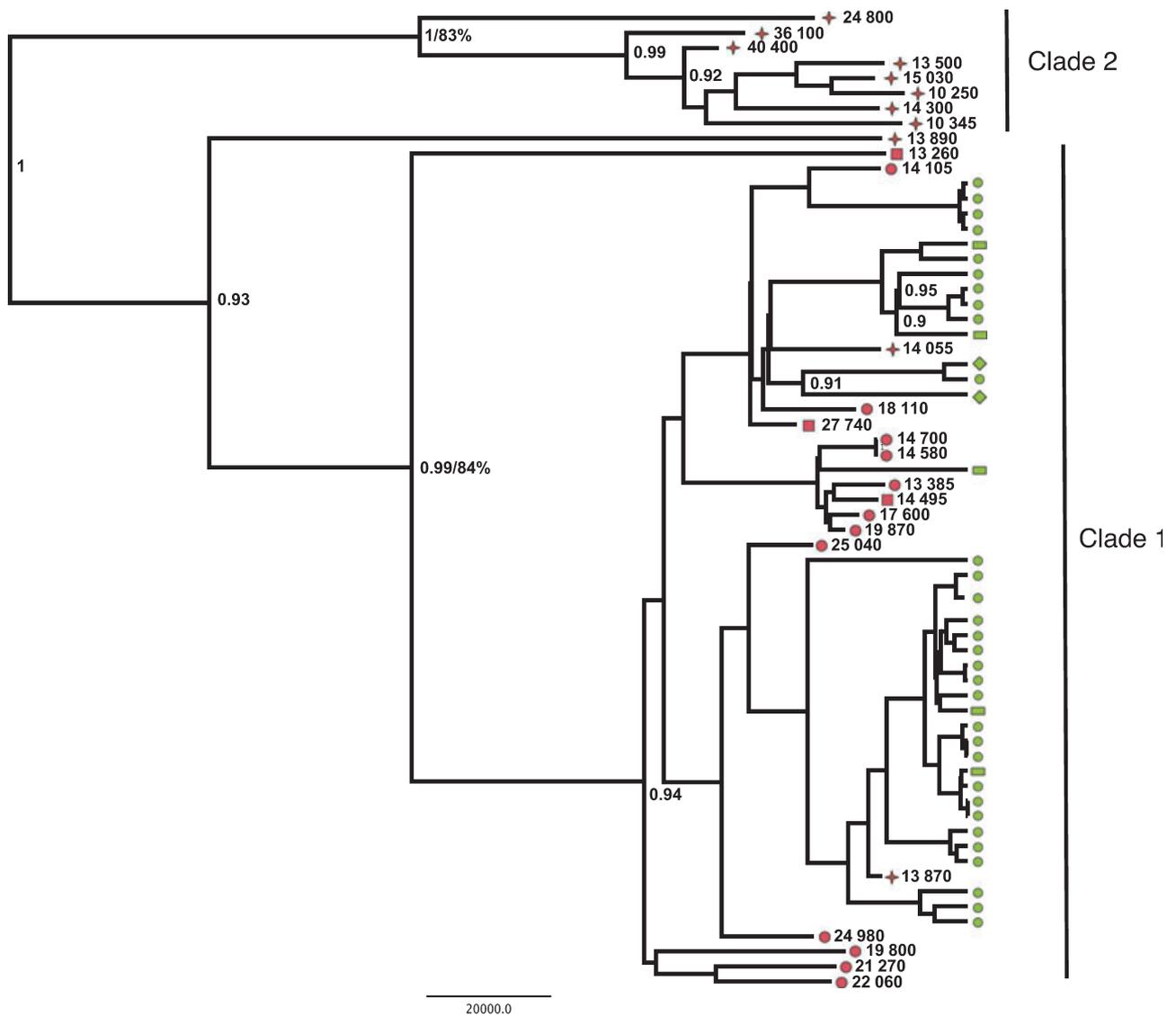


Fig. 3 Bayesian phylogeny of ancient and modern saiga antelope. Posterior probabilities above 0.9 and bootstrap values above 50% are shown for major nodes. Symbols on the tip of the branches correspond to population and numbers to the sample age. Scale bar is given in years.

is restricted to samples from the North Urals, with sample ages spanning from 40 000 to 10 250 ^{14}C ybp. One of the samples from North Urals falls outside of these two clades. Both clades are well supported, presenting very high posterior probabilities and moderately high bootstraps values (0.99/84% and 1/83%, respectively; Fig. 3).

Although there is no clear evidence of a large-scale change in N_e from the results of the skyline analysis (Fig. 4), this is not surprising given the limited sequence length and number of samples analysed. In contrast, BSSC simulations of the bottleneck as a single, sudden event (Fig. 2, model B₁) followed by ABC provided a mean estimate of 3.4–4.3 as for the population size reduction (the mean of a gamma distribution is equal to the product of shape and scale parameters; Table 2, Fig. 5). Note that this value is in agreement with the ratio of N_e before and after demographic decline as calculated from an unbiased estimator for θ_π (c. 2.8–3.2; see column $N_e(\text{Pi}^*)$, Table 1). We note that both empirical likelihood and Bayes Factor analysis provides strong support for population decline (Model B1). However, the genetic data does not contain enough information to estimate the bottleneck time with great accuracy; this event could have occurred at any time between approximately 2000 and 25 000 BP (Fig. 5), but most probably did not occur during recent historical times. Interestingly, a more biologically plausible model of the bottleneck as a gradual exponential decrease (Fig. 2, model B₂) receives less support (Table 2), suggesting that a short and severe, rather than a long and moderate, bottleneck is compatible with the genetic data recovered (Table 2).

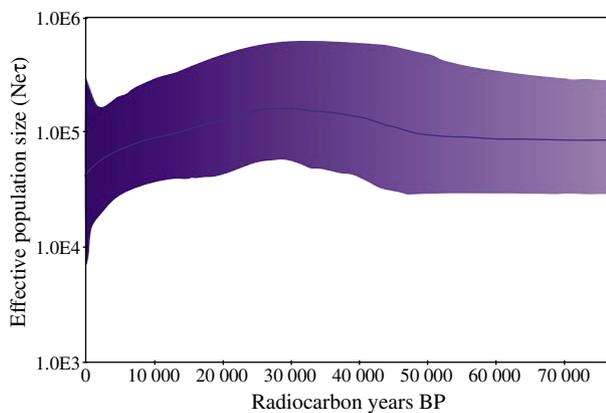


Fig. 4 Bayesian Skyline Plot (BSP) derived from the analysis of the ancient and modern saiga data sets. The x axis is in units of radiocarbon years in the past, and the y axis is equal to $N_e\tau$ (the product of the effective population size and the generation length in radiocarbon years). The thick solid line is the median estimate, and the blue shaded area the 95% highest posterior density limits.

The current saiga population range in the central Asiatic plains may alternatively be a result of range contraction and fragmentation, coupled with local population extinction following the warming of the climate around the Pleistocene–Holocene transition, and subsequent replacement of the steppe tundra ecosystems by taiga forests. Therefore, we investigated if the detected loss of genetic diversity could have resulted from a population fragmentation around the Pleistocene–Holocene boundary (here assumed at 10 000 years BP, i.e. the age of the youngest fossil analysed at the DNA level) followed by later extinction of one of the subpopulations (Fig. 2, models S₁ and S₂). Allowing for subpopulation size differences after splitting (model S₂), the posterior distribution of the ratio between the sizes of the extinct to the current subpopulation peaks at around approximately 2.0 (Fig. 6). In addition, model S₂ receives substantial, to strong, support (Bayes Factor) and best fits the observed genetic data as long as the longer data set is considered (246 nucleotides; Table 2). On the contrary, model S₁ (assuming even fragmentation of the original population into two subpopulations) appears less likely, as attested by lower empirical and marginal likelihood (Table 2). Therefore, the genetic data are both compatible with a sudden demographic decline (model B1) and a fragmentation of saiga populations at the Holocene–Pleistocene boundary followed by later extinction of one (the largest) subpopulation (model S₂).

Discussion

We believe that the relatively low success rate of the extractions and the difficulties in obtaining the full region from many of the samples reflects the state of preservation of the material. The degradation of endogenous DNA commences immediately following cell death, and several factors such as high temperature, proximity to free water, environmental salt content and exposure to radiation increase the rate of DNA decay. Although most of the samples originate from high latitude sites, where low temperatures and rapid desiccation should prolong DNA survival (Lindahl 1993), none of the samples were recently excavated. Thus, their storage at room temperature since excavation might have further added to the DNA decay that they would have undergone since death (Pruvost *et al.* 2007).

The mitochondrial phylogeny indicates that the modern subspecies and populations are not reciprocally monophyletic. Although both additional modern samples (in particular of the *S. t. mongolica* subspecies) and nuclear DNA analyses would be required to further expand on this relationship, given what is known about

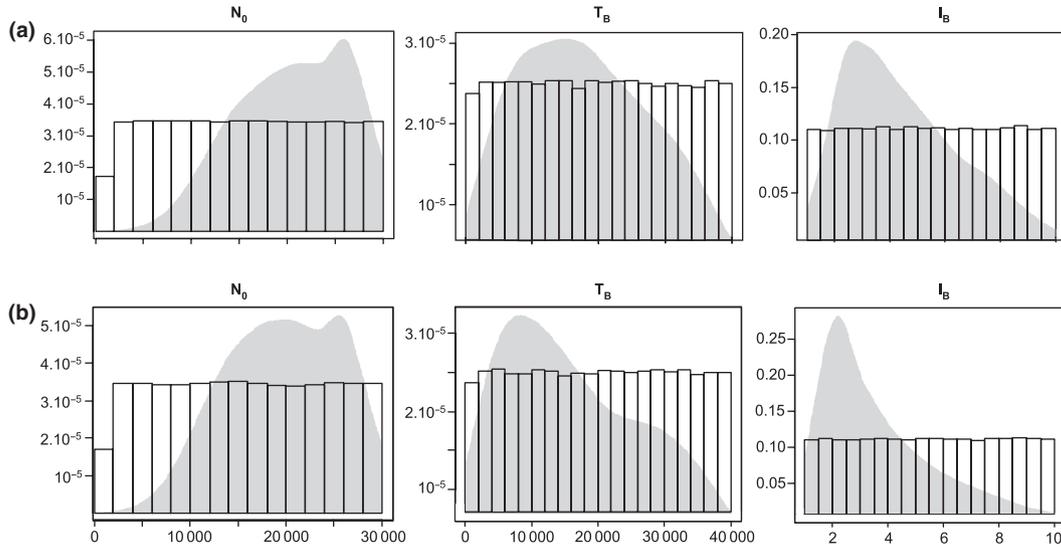


Fig. 5 Posterior distributions for selected parameters of models B_1 (Panel a: 201 sites; Panel b: 246 sites). The histograms represent prior sampling among accepted simulations. The density curve of posterior distributions is coloured in grey.

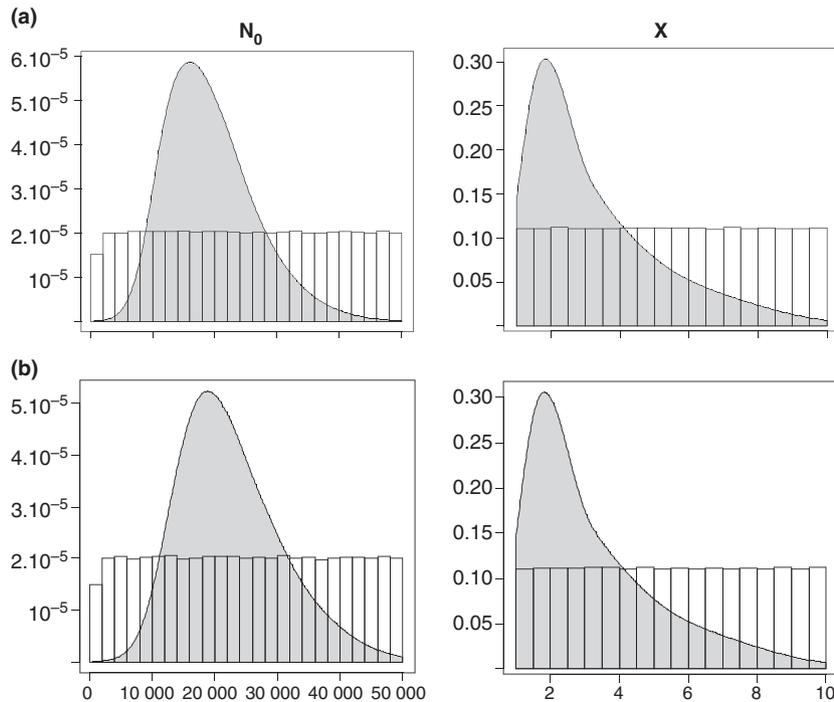


Fig. 6 Posterior distributions for selected parameters of models S_2 (Panel a: 201 sites; Panel b: 246 sites). The histograms represent prior sampling among accepted simulations. The density curve of posterior distributions is coloured in grey.

the history of modern saiga, in particular the very recent fragmentation of its range, it is plausible that the different groups represent a single taxonomic unit. As such, it is unlikely that the observed differences between the levels of modern and ancient genetic diversity would stem from differential contributions of the various subtypes in the two data sets. In comparison

with other antelopes, the total saiga mtDNA nucleotide diversity (i.e. that for both modern and ancient samples) is similar to that observed in modern African impala (*Aepyceros melampus*) and the greater kudu (*Tragelaphus strepsiceros*) (Nersting & Arctander 2001), both species with populations that were, until recently, extremely large and wide ranging. Furthermore, the total sai-

ga diversity is twofold higher than that observed for the roan antelope (*Hippotragus equinus*), a widely distributed sub-Saharan savannah dwelling species that is also threatened by habitat loss and human settlement (Alpers *et al.* 2004), although if only modern populations are considered, the two species present similar levels of diversity. When compared to the Mongolian gazelle (*Procapra gutturosa*), a species native to the semi-arid Central Asian steppes of Mongolia, Siberia and China, and with similar ecological specifications and reproductive biology, saiga diversity (both total and modern), is considerably lower (Sorokin *et al.* 2005).

When the data set is subdivided to reflect the ancient and modern data separately, the observations support the hypothesis that saiga, like a large number of other Pleistocene relict species (Barnes *et al.* 2002; Campos *et al.* 2010; MacPhee *et al.* 2005; Rohland *et al.* 2005; Shapiro *et al.* 2004), had higher genetic diversity in the past. A striking feature of this observation is the loss of a complete, and very distinct, clade. Under the assumption that the observation is not simply a sampling artefact, this lost clade appears geographically restricted to the north of the Ural Mountains (64°N to 59°N). The fossil record indicates that the saiga antelope was not present in the North Urals after the Bølling/Allerød stage (Late Weichselian II, 12 400–10 900 years; Bachura & Kosintsev 2007). As such, it would seem that the disappearance of this clade occurred around the Pleistocene/Holocene boundary, in close correlation with the deglaciation of the Eurasian ice sheet in the northern Urals (Svendsen *et al.* 2004). The serial coalescent modelling results (Fig. 2, model B₂) suggest that the loss of genetic diversity could have resulted from geographic fragmentation of saiga populations at the Pleistocene/Holocene boundary followed by a later local extinction of the North Urals subpopulation. Modern saiga are a nomadic species with a very high dispersal capability (thus under normal circumstances easily capable of crossing the 500 km separating North and Middle Ural sampling sites). Therefore, such population fragmentation would have required significant barriers to the species. The warming of the climate around the Pleistocene–Holocene transition led to replacement of the steppe tundra ecosystems required by saiga by taiga forests, which could have acted as such a barrier. Alternatively, the presence of a now-absent geographical barrier could have isolated the north Urals saiga from the southern populations and confined saiga movements to narrow migratory corridors. As seen in the modern Mongolian population (Berger *et al.* 2008), the impediment may not have needed to be large, and over time the cessation of contact between the north Urals and the larger population elsewhere may have lead to the extinction

of the northern Urals clade. Interestingly, such a situation has been observed in other species, for example the brown bear (*Ursus arctos*), in which the disappearance of the populations in the Atlas mountains ultimately led to the extinction of a highly divergent clade during historical times (Calvignac *et al.* 2008).

An alternate explanation supported by the modelling is that the saiga population was not fragmented, but simply reduced by *c.* 70% in response to both contraction of its home range and possibly human pressure. It is well documented that the saiga's natural range both expanded and then contracted heavily in the Pleistocene (Vereshchagin & Baryshnikov 1984), as both the extent of Russian/Siberian glaciations (Svendsen *et al.* 2004) and the distribution of the steppe and tundra biomes changed (Hubberten *et al.* 2004; Tarasov *et al.* 2000). Toward the present, the home range of the saiga was forced South and South-Eastwards to the southern and central parts of Siberia and to the steppe regions of Kazakhstan where habitats became more preferable.

In addition to the natural causes, within more recent time periods, human occupation of the land inhabited by the saiga may also have been important in the reduction of saiga diversity and range (albeit ABC suggests that the bottleneck probably occurred >2000 BP; Fig. 5). Until the 17th and 18th centuries, the saiga's natural range reached as far as the Carpathian foothills in the West and the Kiev region in the North (Sokolov & Zhirnov 1998) occupying a far more extensive area than in the present. However, over the last few decades, saiga have been heavily exploited, in particular following the recent collapse of the Soviet regime, after which conservation measures and hunting protection rules have essentially become nonexistent. Even though during the late Pleistocene Neolithic humans intruded upon the range of the saiga, their impact in reduction of populations was probably extremely limited, and no evidence of overhunting is known. Similarly, there was probably limited effect from saiga's top predators, wolves (*Canis lupus*) and bears (*Ursus sp.*). Saiga's high fecundity, short generation time and migratory behaviour most likely enabled ancient populations to avoid drastic reductions because of human impact.

Our data clearly demonstrate that the different modern saiga populations and subspecies are not reciprocally monophyletic at the mtDNA level. Furthermore, as with a number of other megafauna that have been recently studied using aDNA (Hofreiter 2007), mitochondrial genetic diversity of the saiga was previously much greater than at present and that there has been a significant reduction in the saiga's genetic diversity over the past 40 000 years. Based on our modelling results, we present two possible hypotheses for the cause, either a large bottleneck in a single population, or population

fragmentation into two isolated groups, followed by extinction of one population. This latter explanation is particularly interesting, in the light of the results of a similar recent study on another highly mobile mammal, the Arctic fox (*Alopex lagopus*) (Dalén *et al.* 2007). In that study, the authors found that Arctic foxes from mid-latitude Europe did not expand into Scandinavia as climate change opened up new environments to them, and thus their response to environmental change did not conform to the model of 'habitat tracking' (Eldredge & Eldridge 1989) the hypothesis that animals should be able to track changes in environmental availability. The authors furthermore hypothesize that this inability to habitat track may be a general pattern among mammalian species (Dalén *et al.* 2007) something which our latter explanation for the loss of saiga genetic diversity would support. As future ancient DNA studies appear that similarly address this question, we look forward to seeing how general a rule this may be.

Acknowledgements

The authors thank the following for funding the research. Forsknings- og Innovationsstyrelsen 272-07-0279 Skou grant (MTPG), Marie Curie Actions 'GeneTime' grant (PFC, EW).

References

- Alpers DL, Van Vuuren BJ, Arctander P, Robinson TJ (2004) Population genetics of the roan antelope (*Hippotragus equinus*) with suggestions for conservation. *Molecular Ecology*, **13**, 1771–1784.
- 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.
- Bachura O, Kosintsev P (2007) Late Pleistocene and Holocene small- and large-mammal faunas from the Northern Urals. *Quaternary International*, **160**, 121–128.
- Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene Population Extinctions in Beringian Brown Bears. *Science*, **295**, 2267–2270.
- Baryshnikov G, Tikhonov A (1994) Notes on skulls of Pleistocene saiga of northern Eurasia. *Historical Biology*, **8**, 209–234.
- Beaumont M, Zhang W, Balding D (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025.
- Bekenov AB, Grachev IA, Milner-Gulland EJ (1998) The ecology and management of the saiga antelope in Kazakhstan. *Mammal Review*, **28**, 1–52.
- Berger J, Young JK, Berger KM (2008) Protecting migration corridors: challenges and optimism for Mongolian Saiga. *PLoS Biology*, **6**, e165.
- Calvignac S, Hughes S, Tougaard C *et al.* (2008) Ancient DNA evidence for the loss of a highly divergent brown bear clade during historical times. *Molecular Ecology*, **17**, 1962–1970.
- Campos PF, Willerslev E, Sher A *et al.* (2010) 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*, **107**, 5675–5680.
- Dalén L, Fuglei EVA, Hersteinsson P *et al.* (2005) Population history and genetic structure of a circumpolar species: the arctic fox. *Biological Journal of the Linnean Society*, **84**, 79–89.
- 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*, **104**, 6726.
- Depaulis F, Orlando L, Hänni C (2009) Using classical population genetics tools with heterochronous data: time matters! *PLoS ONE*, **4**, 5, e5541.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185–1192.
- Eldredge N, Eldridge N (1989) *Macroevolutionary Dynamics: Species, Niches, and Adaptive Peaks*. McGraw-Hill. New York.
- Fabre V, Condeani S, Degioanni A (2009) Genetic evidence of geographical groups among Neanderthals. *PLoS ONE*, **4**, 4, e5151.
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693–709.
- Ghirotto S, Mona S, Benazzo A *et al.* (2009) Inferring genealogical processes from patterns of bronze-age and modern DNA variation in Sardinia. *Molecular Biology and Evolution*, **27**, 875–886.
- Gilbert MTP, Drautz DI, Lesk AM *et al.* (2008) Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 8327–8332.
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by Maximum Likelihood. *Systematic Biology*, **52**, 696–704.
- 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.
- Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.
- Hofreiter M (2007) Pleistocene extinctions: haunting the survivors. *Current Biology*, **17**, 609–611.
- Hofreiter M, Capelli C, Krings M *et al.* (2002) Ancient DNA analyses reveal high mitochondrial DNA sequence diversity and parallel morphological evolution of late pleistocene cave bears. *Molecular Biology and Evolution*, **19**, 1244–1250.
- Hofreiter M, Serre D, Rohland N *et al.* (2004) Lack of phylogeography in European mammals before the last glaciation. *Proceedings of the National Academy of Sciences*, **101**, 12963–12968.
- Hubberten HW, Andreev A, Astakhov VI *et al.* (2004) The periglacial climate and environment in northern Eurasia during the Last Glaciation. *Quaternary Science Reviews*, **23**, 1333–1357.

- Kholodova MV, Milner-Gulland EJ, Easton AJ *et al.* (2006) Mitochondrial DNA variation and population structure of the Critically Endangered saiga antelope *Saiga tatarica*. *Oryx*, **40**, 103–107.
- Kuzmina IE (1971) Formation of the Theriofauna of the Northern Ural Mountains in the Late Anthropogene. *Tr. Zool. Inst. Akad. Nauk SSSR*, **49**, 44–122.
- Leonard JA, Wayne RK, Cooper A (2000) Population genetics of Ice Age brown bears. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1651–1654.
- Leonard JA, Rohland N, Glaberman S *et al.* (2005) A rapid loss of stripes: the evolutionary history of the extinct quagga. *Biology Letters*, **1**, 291–295.
- Lindahl T (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
- Lushchekina A, Struchkov A (2001) The saiga antelope in Europe: once again on the brink? *The Open Country*, **3**, 11–24.
- MacPhee R, Tikhonov A, Mol D, Greenwood A (2005) Late Quaternary loss of genetic diversity in muskox (Ovibos). *BMC Evolutionary Biology*, **5**, 49.
- Milner-Gulland EJ, Kholodova MV, Bekenov A *et al.* (2001) Dramatic declines in saiga antelope populations. *Oryx*, **35**, 340–345.
- Milner-Gulland EJ, Bukreeva OM, Coulson T *et al.* (2003) Reproductive collapse in saiga antelope harems. *Nature*, **422**, 135.
- Nersting LG, Arctander P (2001) Phylogeography and conservation of impala and greater kudu. *Molecular Ecology*, **10**, 711–719.
- Orlando L, Bonjean D, Bocherens H *et al.* (2002) Ancient DNA and the Population Genetics of Cave Bears (*Ursus spelaeus*) Through Space and Time. *Molecular Biology and Evolution*, **19**, 1920–1933.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Pruvost M, Schwarz R, Correia VB *et al.* (2007) Freshly excavated fossil bones are best for amplification of ancient DNA. *Proceedings of the National Academy of Sciences*, **104**, 739.
- R Development Core Team (2008) R: A language and environment for statistical computing. R foundation for statistical computing.org. Retrieved from <http://www.r-project.org>.
- Rambaut A, Drummond AJ (2007) Tracer v1.4. Available from: <http://beast.bio.ed.ac.uk/Tracer>. Accessed 5/5/2009.
- Rohland N, Pollack JL, Nagel D *et al.* (2005) The population history of extant and extinct hyenas. *Molecular Biology and Evolution*, **22**, 2435–2443.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) *DnaSP, DNA Polymorphism Analyses by the Coalescent and Other Methods*. pp. 2496–2497. Oxford University Press, Oxford.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual* (ed.). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Shapiro B, Drummond AJ, Rambaut A *et al.* (2004) Rise and fall of the Beringian Steppe Bison. *Science*, **306**, 1561–1565.
- Shepherd LD, Lambert DM (2008) Ancient DNA and conservation: lessons from the endangered kiwi of New Zealand. *Molecular Ecology*, **17**, 2174–2184.
- Sher AV (1974) Pleistocene mammals and stratigraphy of the Far Northeast USSR and North America. *International Geology Review*, **16**, 1–89.
- Sokolov VE, Zhirnov LV (1998) *Saigak: filogeniya, sistematika, ekologiya, okhrana i ispol'zovanie* [The Saiga Antelope: Phylogeny, Systematics, Ecology, Conservation, and Use]. Russian Academy of Sciences, Moscow.
- Sorokin PA, Kiriliuk VE, Lushchekina AA, Kholodova MV (2005) Genetic Diversity of the Mongolian Gazelle *Procapra guttorosa* Pallas, 1777. *Russian Journal of Genetics*, **41**, 1101–1105.
- Stewart JR, Lister AM (2001) Cryptic northern refugia and the origins of the modern biota. *Trends in Ecology & Evolution*, **16**, 608–613.
- Svendsen JI, Alexanderson H, Astakhov VI *et al.* (2004) Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews*, **23**, 1229–1271.
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, **7**, 453–464.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
- Tarasov PE, Volkova VS, Webb T *et al.* (2000) Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *Journal of Biogeography*, **27**, 609–620.
- Valdiosera CE, Garcia-Garitaigoitia JL, Garcia N *et al.* (2008) Surprising migration and population size dynamics in ancient Iberian brown bears (*Ursus arctos*). *Proceedings of the National Academy of Sciences*, **105**, 5123.
- Vereshchagin NK, Baryshnikov GF (1984) Quaternary mammalian extinctions in northern Eurasia. In: *Quaternary Extinctions: A Prehistoric Revolution* (eds Martin PS, Klein RG). pp. 416–483, University of Arizona Press, Tucson.
- Voight B, Adams A, Frisse L *et al.* (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proceedings of the National Academy of Sciences*, **102**, 18508.
- Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.
- Yang DY, Eng B, Waye JS, Dudar JC, Saunders SR (1998) Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American Journal of Physical Anthropology*, **105**, 539–543.

Paula Campos is interested in megafauna population dynamics in the late Pleistocene and Holocene and in the early human migrations into the New World. Tom Gilbert's research interests focus on the challenges offered by the study of nucleic acids in degraded materials, solutions to these problems, and the evolutionary biology, archaeological and anthropological questions that can be answered through using such samples. Eske Willerslev is interested in evolutionary biology, past environmental reconstructions, population genetics and phylogenetics using both ancient and contemporary DNA sequence data.

Supporting information

Additional supporting information may be found in the online version of this article:

Table S1 Primers used in this study.

Table S2 Detailed description of specimens used in this study.

Fig. S1 Distribution of the 40 fossil saiga dates used in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.