universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

DNA Analysis of Late Glacial Mammal Remains

Reconstructing temporal demographic changes in the siberian collared lemming (Dicrostonyx torquatus) using Approximate Bayesian Computation (ABC) and Markov Chain Monte Carlo (MCMC) based bayesian inference.

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasserin / Verfasser:	Stefan Prost
Matrikel-Nummer:	0202910
Studienrichtung /Studienzweig (lt. Studienblatt):	Diplomstudium Genetik - Mikrobiologie
Betreuerin / Betreuer:	ao Univ.-Prof. Dr. Doris Nagel

Wien, am 28.08.2009

Contents

Chapter 1

Introduction

All work has been done at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany under the local supervision of Michael Hofreiter and the general supervision of Doris Nagel (Institute for Paleontology, University of Vienna).

Using molecular data to study different aspects of biology (e.g. evolution, population ecology or taxonomy) is a common approach today. Several studies used modern DNA to reveal past population history (e.g. Alter *et al.* (2007); Thalmann *et al.* (2007); Ruzzante *et al.* (2008)). However, studies of demographic events solely based on contemporary DNA have limited strength to reconstruct historical processes (Chan *et al.* , 2006). Thus, the advent of ancient DNA (aDNA) extraction and amplification was of outstanding importance for our understanding of responses of populations to changing environments. One major advantage, when working with ancient DNA, is that temporal demographic changes can be monitored over thousands of generations. Furthermore, temporal sampling improves the accurateness of the results (Chan *et al.* , 2006).
This study focusses on the response of a collared lemming population from the northern Ural mountain to the abrupt climate change at the Pleistocene-Holocene transition using temporal ancient DNA sampling.

In the following I will provide an introduction to different aspects of this work. The first part will give a general introduction to the project and an overview over ancient DNA in general, precautions when working with aDNA and ecological aspects of the study. The last part is devoted to the computational methods used to unravel the demographic history.

1.1 General Introduction

Climate change is a popular and serious issue today both in public perception and scientific literature, due to it presumed impact on the Earth's ecological and biological systems. In general, climate changes such as temperature increase result in loss of biodiversity, polwards expansion of species' ranges, shifts in phenological events (e.g. reproduction, blooming), higher dispersion of diseases and even extinction (Parmesan & Yohe, 2003; Root *et al.* , 2003; Thuiller *et al.* , 2005; Thomas *et al.* , 2004; Patz *et al.* , 2005). Effects of climate changes can be quite drastic. This can be seen in the mass extinction event at the end of the Permian (251 My ago), when nearly 95%

of all species went extinct (Benton & Twitchett, 2003). Many studies treat the impacts of recent global warming (e.g. (Parmesan & Yohe, 2003; Root *et al.* , 2003)) concluding strong effects even when the Earth' climate warmed on by an average of 0.6 °C (Root *et al.* , 2003). According to the IPCC the global average temperature will increase between 1.4 and 5.8 °C over the period 1990 to 2100 (Houghton *et al.* , 2001). This prospect further emphasizes the importance of a proper understanding of climate change.

Taking into account the popularity of this topic surprisingly little is known about its effects on the genetic level (Hadly *et al.* , 2004). A proper understanding of molecular responses is crucial for assessment and prediction of future scenarios. Disciplines such as conservation biology and medicine will be confronted with exacerbated challenges such as differential movements (Overpeck & Webb, 1992), decreasing fitness (Parmesan & Yohe, 2003) and liability to diseases (Patz *et al.* , 2005). In order to minimize the implications of climate change (e.g. loss of biodiversity) an improved knowledge of possible progressions and risks is crucial. An encouraging possibility to improve our understanding of climate change and its implications is to study the past.

The late quaternary period (the past 1 My) was marked by cycles of glacial-interglacial changes (Imbrie *et al.* , 1992; Tzedakis *et al.* , 1997). Thus it can be used to study responses to climatic changes. The most prominent climate change of the late Quaternary was the transition from late Pleistocene to Holocene. This period was characterized by two major climate shifts at $\sim$14.7 and $\sim$11.5 ky cal. BP, respectively, which occurred within a few decades only (Dansgaard *et al.* , 1989; Lehman & Keigwin, 1992; Dansgaard *et al.* , 1993). Ice-core data indicates a warming of $\sim$7 °C within 5 decades at the end of the Younger Dryas stadial (Dansgaard *et al.* , 1989). Furthermore those drastic climate transitions led to vegetation changes (e.g. increased flowering, which is indicated by an increase in pollen concentration in the paleontological record.) (Hoek, 2001).

State-of-the-art ancient DNA techniques (e.g improved extraction methods, high-throughput sequencing technologies, etc.) provide an outstanding tool to study past population dynamics. Ancient DNA allows to survey mammal populations over hundreds to thousands of generations and therefore helps to reconstruct temporal demographic history on a reasonable time scale. This reconstructions can then be linked to prominent climate changes.

The "phylochronological" approach (Hadly *et al.* , 2004) is a powerful tool to study temporal demographic changes using serial ancient DNA sampling. In their study Hadly *et al.* (2004) analyzed temporal changes of genetic diversity in a single study site over 2500y BP. This "phylochronological" approach has been successfully used within the last years to study demographical histories such as bottleneck events (e.g. Chan *et al.* (2006)).

In order to study the effects of such abrupt climate changes as the ones occurred at $\sim$14.7 and $\sim$11.5 ky cal. BP I sampled subfossil and modern bones of the artic collared lemming *Dicrostonyx torquatus* from a single site (Pymva-Shor) in the northern Ural. The layers of this site are dated to 25.244 $\mp$ 359 (L6 up), 15.246. $\mp$ 122 (L6 low), 11.538 $\mp$ 423 (L4) y cal. BP and modern (surface), thus representing sampling before, while and after a possible population decline as expected from its ecological demands. Collared lemmings (Dicrostonyx sp.), the northernmost genus of rodents, evolved in dry landscapes of eastern Siberia and was characteristically associated with dry and cold environment of tundra steppe during Pleistocene (Agadjanian, 1976; Zazhigin, 1976; Kowalski, 1995). Furthermore, the paleontological record from this site shows a decrease in numbers of collected bones after the last glacial maximum (LGM) (see figure 4 in Golovachov & Smirnov (2008)), which decreases further and collapses finally at around 8.000 to 9.000 BP. This temporal sampling indicates a population decline in *Dicrostonyx torquatus*, but has to be treated carefully since it also could reflect a change in the prey-pattern of possible chasers (such as owls).

The aim of this paper is to provide the first study linking temporal demographic changes in a mammal species to climate changes at the Pleistocene- Holocene transition using temporal ancient DNA data. Therefore improving our knowledge of possible responses of mammal population to climate shifts and thus providing helpful information for conservation of species facing future climate changes.

1.2 Ancient DNA

Under usual circumstances DNA gets degraded soon after the death of an organism by endonucleases (Hofreiter *et al.* , 2001a; Lindahl, 1993). In addition chemical damage starts to accumulate in the DNA (Lindahl, 1993). These chemical and enzymatic forces lead to strand breaks and nucleotide modifications which may block elongation during the PCR reaction or lead to incorporation of false nucleotides into the newly synthesized DNA strand (Hofreiter *et al.* , 2001a).

Definition 1.1 (Ancient DNA)

Ancient DNA is DNA from subfossil remains (usually thousands of years old) and museum specimens (tens to hundreds of years).

Thus, sequences derived from ancient DNA samples have to be treated careful. To provide trustful and accurate data 'criteria of authenticity' have been formulated (Cooper & Poinar, 2000; Gilbert *et al.* , 2005). According to the authors, all ancient DNA studies should follow to the nine criteria provided in figure 1.1. However, under certain conditions some citeria may be ignored (Gilbert *et al.* , 2005). In this study I adhered to the first five criteria, which are seen to be sufficient to emphasize reliability of the data.
The next subsection summarizes the sources of ancient DNA damage, kinds of damage and its implications.

Box 1. The nine criteria for authenticity*

(i) Isolation of work areas: to separate samples and extracted DNA from PCR amplified products.
(ii) Negative control extractions and amplifications: to screen for contaminants entering the process at any stage.
(iii) Appropriate molecular behaviour: owing to DNA degradation, the successful amplification of large DNA fragments in ancient DNA studies should be treated with caution.
(iv) Reproducibility: multiple PCR and extractions should yield consistent results.
(v) Cloning of products: to assess for damage, contamination and jumping PCR.
(vi) Independent replication: the generation of consistent results by independent research groups.
(vii) Biochemical preservation: preservation of other biomolecules that correlate with DNA survival (e.g. collagen or amino-acid racemization) should indicate good sample preservation.
(viii) Quantification: by competitive PCR or Real-Time PCR to give an indication of the number of starting templates in the reaction.
(ix) Associated remains: are associated remains equally well preserved, and do they show evidence of contamination?

Figure 1.1: **Nine criteria of authenticity after Cooper & Poinar (2000).**

The next subsection summarizes the sources of ancient DNA damage, kinds of damage and its implications.

1.2.1 Ancient DNA damage and contamination

As mentioned above chemical and enzymatic damage start to accumulate in the DNA soon after death of an organism. Under fortunate circumstances DNA can survive in the bones and teeth for thousands of years. Such circumstances comprise low temperatures, fast evaporation and high salt concentrations (Hofreiter *et al.* , 2001a). Given that these conditions are relatively rare, it seems quite stunning that ancient DNA can be obtained from many caves in Eurasia, America and even New Zealand. However, aDNA preservation can be quite different within even a single site.
Most common forms of DNA damage in aDNA are strand breaks and deamination at the cytosine base (Hofreiter *et al.* , 2001a,b; Stiller *et al.* , 2006). This deamination leads to C to T and G to A transitions in the DNA sequences. Stiller *et al.* (2006) showed that deamination at the cytosine and to a much lesser extent guanine lead to misincorporation in the DNA. They further argue that C to T changes are the most prominent damage in ancient DNA. In addition Briggs *et al.* (2007) and Brotherton *et al.* (2007) argue that C to T transitions are the only source for misincorporation and G to A changes are merely a result from the C to T changes at the other DNA strand. Principles sites and kinds of DNA damage can be seen in figure 1.2.
To avoid sequence errors due to damage each sequence position has to be determined from at least two independent amplifications. If consistent differences between two independent amplifications are found, a third PCR has to be performed to clarify which nucleotide represents the correct sequence.

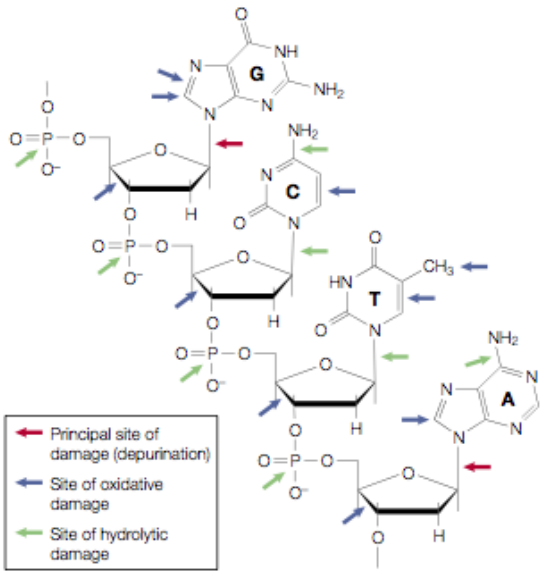

Figure 1.2: **Principle sites and kinds of DNA damage after Hofreiter *et al.* (2001a) modified after Lindahl (1993).**

Another prominent source of error, when dealing with ancient DNA is contamination with modern sequences (such as PCR products in the air or chemicals). Leonard *et al.* (2007) showed that many different animal sequences can be found in chemicals used within the laboratory work. Special precaution has to be taken especially when working on ancient human remains (Abbott, 2003). Sources for human contamination are diverse and frequent, such as researchers own DNA. However, human contamination is not a big issue when working with animal remains. In order to minimize the possibility to contaminate aDNA extracts and PCRs all pre-PCR work has to be done in a so called clean-room, which has to be separated spatially from the post-PCR-laboratories. One should not be allowed to enter the clean-room after working in the post-PCR laboratory. PCR products can be found nearly everywhere (including the air) within post-PCR laboratories. Thus, just entering such a laboratory can be sufficient to contaminate the researchers clothes. Work within a clean-room is shown in figure 1.3.

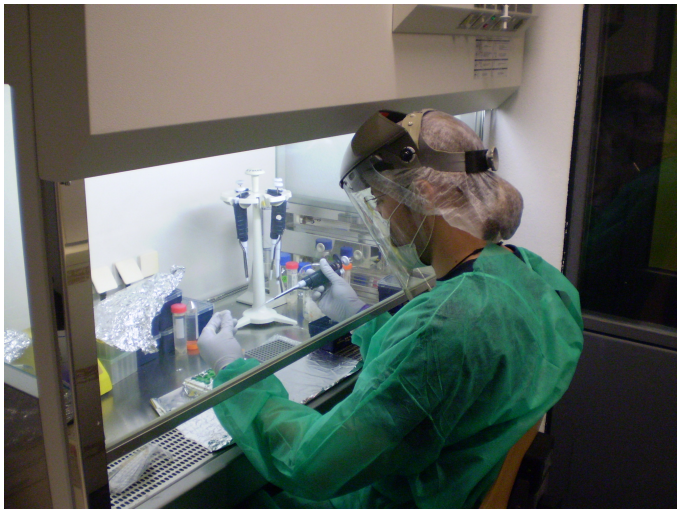

Figure 1.3: **Work within a cleanroom.**

1.3 Ecology

1.3.1 Climate of the late Pleistocene and Holocene

The history of the Pleistocene was strongly influenced by cyclic climate changes. Whereas the period of this glacial-interglacial cycles comprised 41.000 y during early Pleistocene it changed to 100.000 y at about 850.000 y BP (e.g. Hays *et al.* (1976); Paillard (1998); de Garidel-Thoron *et al.* (2005)). This periodicities are correlated with Earth's main orbital parameters: obliquity (41.000 y) and eccentricity (100.000 y) (Paillard, 1998). As a crucial energy transporter the ocean circulation system has strong effects on the stability of the climate (Hewitt, 2000). Paleoclimate simulations showed the influence of ocean heat transport on the climate of the Last Glacial Maximum (Webb *et al.* , 1997). Evidence for this periodical climate cycles comes from isotope (e.g.

oxygen,...) and chemical combound (e.g. methan, nitrous oxide,...) measurements from ice cores, cores of the sea bed and lake bottoms (e.g. Brook *et al.* (1996); Hewitt (2000); Spahni *et al.* (2005)).

However, given the durability of ancient DNA we are most interested in the climate of the late Pleistocene and Holocene. Furthermore this period comprises drastic climate shifts that had major effects on Earth's biological and ecological systems as we know them today.
Figure 1.3.1 shows glacials and interglacials of the last 250.000 y of Earth's history, with the Last Glacial Maximum at about 25.000 y cal. BP, the Bolling/Allerod Interstadial at approximately 14.700 y cal. BP and the transition from Pleistocene to Holocene at approximately 11.500 y cal. BP. In this section I will focus mostly on climate history of the last 25.000 years of Earth's history. The last glacial maximum took place at the end of the Greenland Stadial 3 roughly 25.000 y cal. BP ago and lasted until 18.000 y cal. BP. This period was characterized by Greenland temperatures 21 °C below the present (Cuffey & Clow, 1997). Climate reconstruction of the Taimyr peninsula revealed temperatures about 3-6 °C colder then today (seasonally dependent) for northern Taimyr (Andreev *et al.* , 2003) and temperatures 4-5 °C below modern values for lower Taimyr (Andreev *et al.* , 2002). The vegetation comprised open steppe-like - and tundra-like plant communities in more mesic areas. Temperature increased during the following years and reached its maximum at the Bolling/Allerod interstadial with average year temperatures close to modern or slightly below (Andreev *et al.* , 2002; Klimanov, 1997). During the Allerod the Eurasian climate was strongly influenced by occurrence of the ice-sheet and the regression of the sealevel (Klimanov, 1997).

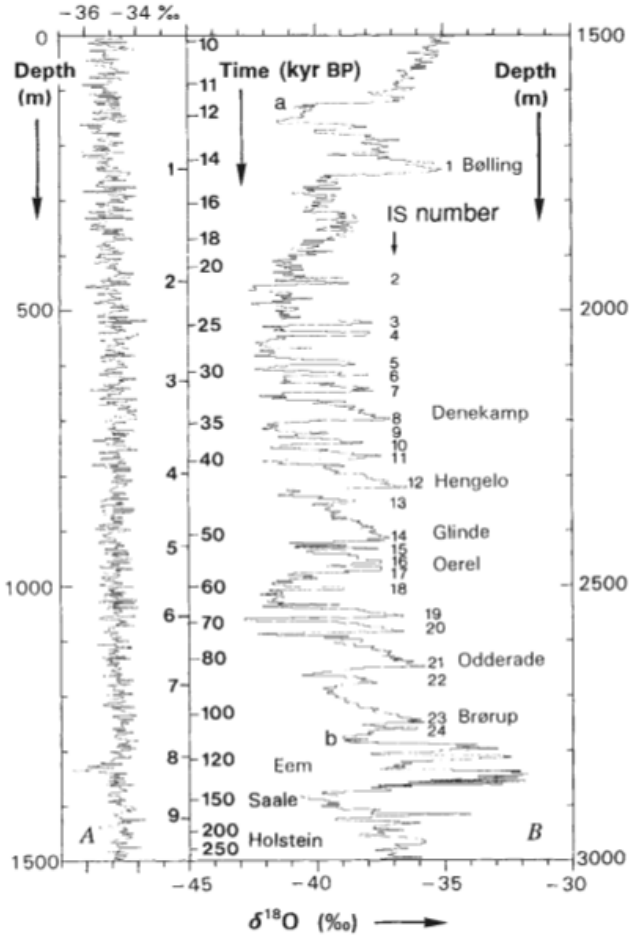

Figure 1.4: **Climate Reconstruction based on the GRIP record from (Dansgaard *et al.* , 1993).**

The climate changed abruptly around 12.900 y cal. BP at the beginning of the Younger Dryas. The cause of this short cooling event is still debated. One major theory is a drastic melt water influx into the northern Atlantic (Kennett & Shackleton, 1975; Broecker *et al.* , 1989; Dansgaard *et al.* , 1989; Broecker, 2006). During the Younger Dryas the temperature decreased to 3-4 °C below the present (Velichko *et al.* , 1997; Andreev *et al.* , 2002). This cooling phase lasted until approximately 11.500 y cal. BP, when average year temperature increased by roughly 7 °C within 5 decades only (Dansgaard *et al.* , 1989).
The vegetation of the Pleistocene Holocene transition at the end of the Younger Dryas was characterized by a drastic decrease in herb pollen proportion especially from steppe and tundra taxa (Andreev *et al.* , 2002). This early Preboreal period had temperatures of about 1.5 °C higher than today (Andreev *et al.* , 2002). Periglacial vegetation degraded and forest formation expanded rapidly throughout northern Eurasia (Velichko *et al.* , 1997).
The Boreal period (10.600 - 9.100 y cal. BP) was characterized by dark coniferous spruce forests (Velichko *et al.* , 1997). The atlantic period (9.100 - 5.750 y cal. BP) was characterized by a development of mixed forest vegetation (Velichko *et al.* , 1997). Temperatures were about 1.5 °C below present (Andreev *et al.* , 2002). The subatlanticum (2.500 y cal. BP until present) fea-

tured a cooling and vegetation similar to present (Velichko *et al.* , 1997; Andreev *et al.* , 2002, 2003). This latest period of the Earth's climate history showed brief warming events intruding the present-like-conditions (Andreev *et al.* , 2003).

1.3.2 Radiocarbon Dates and Calibration

A critical step to link temporal genetic data to climate reconstructions is the dating of the samples or layers, respectively. Radiocarbon dating is the most common technique used today. This method is based on the decay of the concentration of the unstable isotope ^{14}C in an organism after death. In principle there are three natural isotopes of carbon ^{12}C, ^{13}C (both stable) and ^{14}C (unstable or radioactive). ^{14}C is formed in the atmosphere by thermalization of neutrons and its reaction with nitrogen (^{14}N) (Currie, 2004) and enters the organisms as $^{14}CO_2$ after oxidation. It is continuously incorporated into the body via food, reflecting its concentration in the atmosphere. After the death of an organism no further ^{14}C will be included and the concentration starts to decline. The rate of this decay was first measured by Libby *et al.* (1949). They discovered the so called "half-time" of ^{14}C, the time during which half of the isotopes in the sample decay (5568$\pm$30 years). Since little of the isotope remains after 10 half-lives this approach is restricted to dates up to approximately 50.000 years. Since the discovery of ^{14}C itself and its use for dating of subfossil remains different refinements and extensions led to better estimations and a widely acceptance of this method.
One major problem of radiocarbon dating is that the actual proportion of ^{14}C to ^{12}C in the atmosphere is not stable, mostly because of changes in the magnetic field and the effects of sunspots on the amount of cosmic radiation (Voelker *et al.* (1998); Kitagawa & Van der Plicht (1998); Beck *et al.* (2001); Hughen *et al.* (2004); Van der Plicht *et al.* (2004); Fairbanks *et al.* (2005), etc.). Thus radiocarbon dates do not match calendar years. To overcome this problem calibration curves have been estimated based on tree-ring chronology (Reimer *et al.* , 2004), deep-sea sediments (Hughen *et al.* , 2004), stalagmite formation (Van der Plicht *et al.* , 2004) and fossil coral formations (Fairbanks *et al.* , 2005). Table 2.11 shows calibrated and uncalibrated dates of the samples used in this study.

1.4 Primer to Computational Biology

In the following I will briefly describe the theoretical background of the computational approaches applied in this study and provide simple schemes of important facts.

1.4.1 Model-based Data Analysis in Population Genetics

" All models are wrong, some are useful" George Box in Box et al. *(1979)*

The advent of the "Post-Sanger Sequencing Area" led to a high availability of bigger and bigger data sets due to fast and cost-efficient DNA sequencing methods. High-through-put sequencing as Roche's 454 technology or Illumina's SOLEXA generates data up to 3 giga bases in one run

(for Illumina SOLEXA) providing sufficient and fast data processing platforms. This growth in the amount of molecular data intensifies the needs for efficient data analysis approaches all along with an increasing need for computational power. Recent developments in computational genetics focus on different aspects of this problems. This section will give an overview of popular model-based approaches that use different make-ups to deal with huge data amounts and different strategies to overcome computational limitations.

The most popular models in population genetics are so called *stochastic models*, which allow for many possible outcomes rather than a pre-determined (as in *deterministic models*) (Marjoram & Tavaré, 2006). For further details see Primer 1.1. Kingman invented the most widely used stochastic model in population genetics, the so called *coalescent theory* (see chapter 1.4.2) in 1982 (Kingman, 1982). The coalescent characterizes the genealogy (”family tree”) of a DNA sample following the Wright-Fisher model of inheritance (see chapter 1.4.2). Given certain assumptions it describes the probability of randomly taken samples of DNA fragments to share a common ancestor at time *t* in the past. One major advantage of the coalescent is that it models the genealogical process backwards in time therefore just counting the actual lineages that gave rise to the sample in question. Thus this approach is very efficient when it comes to computational power. All approaches described in this chapter are based on the *coalescent* using the wright-fisher model of inheritance.

Primer 1.1 (Deterministic vs. Stochastic Models)

One major aspect of a model is whether it includes uncertain parameters or not. A model in which a certain state is defined at any given time is called a ***Deterministic Model****. Let us imagine a very simple model of a ball moving along a straight line, as shown in figure 1.6. Every* ***n's*** *minute it advances with a fixed length toward the right site. Given a fixed moving length its actual position at any time point (t'$_n$) is determined. Therefore e.g. at t_2 the ball will be 2n's further to the right.*
In general, biological systems are not definite. Temporal demographic changes depend on population factors (e.g. the reproductive success of individuals) as well as on external factors (e.g. changing living conditions). So that, when modeling demographic changes over time one has to include probabilistic aspects. A model including probabilistic characteristics is called a ***Stochastic - or Probabilistic Model****. Inferences from stochastic models are based on probable outcomes rather then certain results.*

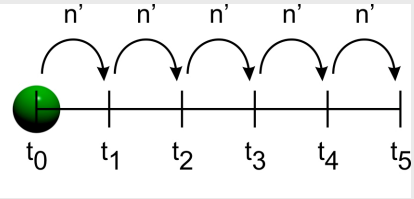

Figure 1.5: Deterministic Model. © by Prost and Finstermeier

The main question one wants to address when using model-based analysis is: ”How was the underlying data generated?”. Thus trying to infer accurate estimates of parameters shaping the genealogical process.

As mentioned above *stochastic models* include probabilistic aspects to predict parameter values of the model. Given the sequence data and a chosen model one can calculate the *posterior probability* in a *bayesian framework* using conditional probabilities. Let *D* denote the observed data and θ be a vector of the unknown model parameter then the *posterior probability distribution* (that is the distribution of the probabilities of the parameter given the data) $f(\theta|D)$ will be:

$$f(\theta|D) = \frac{P(D|\theta)\pi(\theta)}{P(D)} \tag{1.1}$$

$\pi(\theta)$ denotes the prior distribution, *P(D|θ)* the likelihood function of the data and *P(D)* a normalization constant. The likelihood is the probability *P* of the data *D* over possible parameter values θ, as conveyed in *P(D|θ)*.
As the data amount and the model complexity increases the likelihood becomes unresolvable (concerning reasonable computational power and time). Two simplifications can be made to overcome this problem: (1) to use a summary of the data to approximate the posterior density or (2) to use simplified models. The next subsection will focus on the approximation of the data by using summary statistics (1) as used in the *Approximate Bayesian Computation (ABC)* approach (Beaumont *et al.* , 2002). Briefly, in the ABC approach the data is summarized by summary statistics and the posterior probability distribution is approximated using rigorous simulations instead of calculating the likelihood directly from the data and the prior distribution (includes prior knowledge about the parameter θ of interest).

In general the posterior probability distribution *P(θ|D)* is proportional to the product of the prior distribution *π(θ)* and the likelihood of the data *P(D|θ)*. Since the normalization constant depends only on the data we can simplify the equation:

$$P(\theta|D) \propto P(D|\theta)\pi(\theta) \tag{1.2}$$

In figure 1.6 a simplificated scheme of model-based analysis is provided. The key parts of this approach include prior knowledge that is expressed in the prior distribution *π(θ)* and the likelihood that is calculated based on the data *D* and the coalescent model.

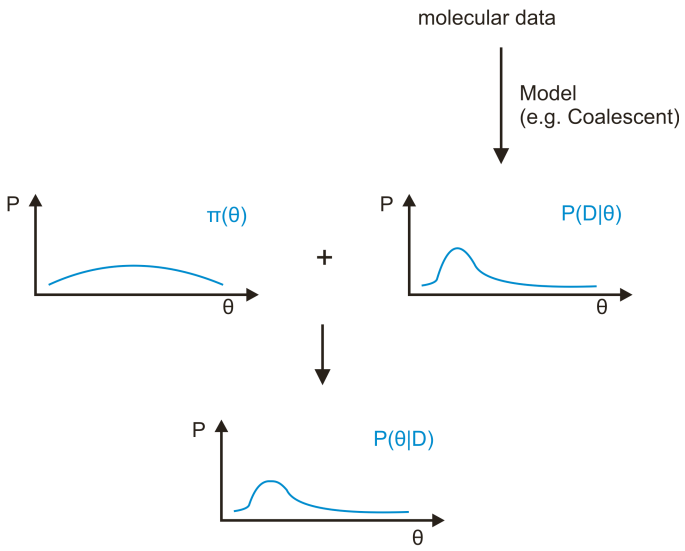

Figure 1.6: **Principle of model-based analysis.** The posterior probability distribution *P(D|θ)* is proportional to the product of the prior distribution *π(θ)* and the likelihood of the data *P(D|θ)*. The likelihood is calculated using sequence data *D* and a model (e.g. Coalescent Model).

1.4.1.1 Approximate Bayesian Computation

The approximate bayesian computation (ABC) approach was invented by Beaumont *et al.* (2002) and is based on the rejection-sampling method from Tavare *et al.* (1997). Tavare *et al.* (1997) used simulations to approximate the posterior probability and replaced the data by summary statistics, because of the difficulties to compute the likelihood of the data. The drawback of their method is that it accepts only sufficient summary statistics when

$$P(S = S_i|\theta') \geqslant cU \tag{1.3}$$

where U is a uniform variable and *c* is a constant that applies to $c \geqslant max_\theta P(S = S_i|\theta)$. This approach was further improved by Fu & Li (1997), Weiss & von Haeseler (1998) and Pritchard *et al.* (1999). Weiss & von Haeseler (1998) first introduced a tolerance δ instead of an exact match $(S = S_i)$, that it accepts θ' when $||\ S_i - S\ || \leqslant \delta$. Beaumont *et al.* (2002) further improved these methods by applying a spherical acceptance region using the Euclidean norm $||\ S_i\ || = \sqrt{\sum_{j=1}^{q} s_j^2}$, where $S_i \equiv (S_1, ..., S_q)$ and the introduction of (local) linear regression adjustment and smooth weighting. Weighting corresponds to the distance between the simulated data and the empirical observed data and can therefore improve properties of the posterior probability estimates (Marjoram & Tavaré, 2006).

Approximate Bayesian Computation Algorithm - rejection method:

A1 Sample parameter θ_i from a prior distribution $\pi(\theta)$
A2 Simulate data D_i using the model *M* with parameter θ_i
A3 Summarize D_i with summary statistics S_i
A4 Summarize empirical data D with summary statistics S
A5 Accept θ_i whenever $|| S_i - S ||\leqslant \delta$, otherwise reject
A6 repeat A1 - A5 until *k* acceptances have been obtained.

Approximate Bayesian Computation Algorithm using regression adjustment and smooth weighting:

B1 Sample parameter θ_i from a prior distribution $\pi(\theta)$
B2 Simulate data D_i using the model *M* with parameter θ_i
B3 Summarize D_i with summary statistics S_i
B4 Summarize empirical data D with summary statistics S
B5 Apply smooth weighting and local linear regression
B6 Accept θ_i whenever $|| S_i - S ||\leqslant \delta$ otherwise reject
B7 repeat B1 - B6 until *k* acceptances have been obtained.

As can be seen in figure 1.7 weighting θ_i (that is the value for the parameter θ randomly drawn from the prior distribution $\pi(\theta)$) according to $|| S_i - S ||$ and the weakening of discrepancy effects between S_i and S using local-linear regression improve the posterior probability distribution.

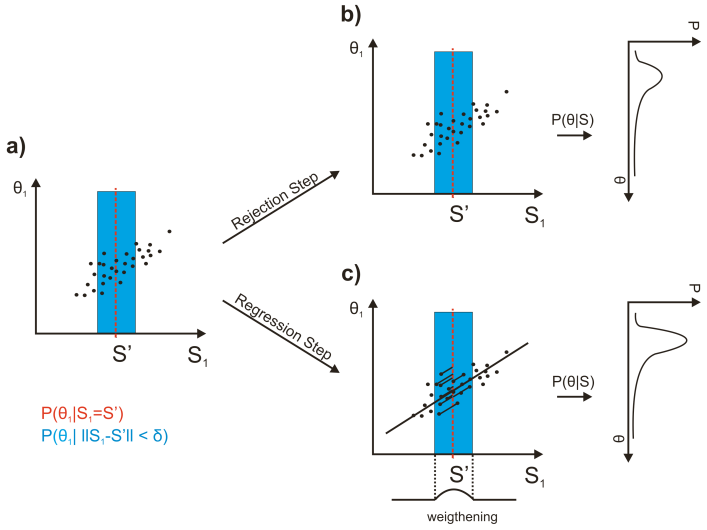

Figure 1.7: **Schematic overview of the rejection - and the local linear regression and smooth weighting step of the Approximate Bayesian Computation (ABC) algorithm.** a) Correlation of the parameter estimates of the simulated samples (θ_1) to the summary statistics (S'...observed and S_1...simulated). The blue area marks the division of the 1000 closest parameter estimates. b) Rejection step and the inferred probability *P(θ|S)* c) Local linear regression and smooth weighting step and the inferred probability *P(θ|S)*. © by Sollari and Prost

When accepting θ_i according to a tolerance $|| S_i - S ||\leqslant \delta$ one should not exceed *k*=10.000, because inaccurateness will be introduced when adding acceptances more and more deviating from the true value of θ_i . Beaumont *et al.* (2002) used a quantile, P_δ, for the empirical distribution function of the simulated $|| S_i - S ||$, that when simulating a model *M* with 1.000.000 iterations and a tolerance of $P_\delta = 0.001$ it will refer to an acceptance of 0,01% of the data namely the 1000 closest estimates. According to Beaumont (personal communication) one should accept about 1.000th to 10.000th of the closest θ_i estimates.

1.4.1.2 MCMC based bayesian inference

For some very simple models the *posterior probability distribution* can be derived analytically. However, many models are far to complicated and thus approximation methods have to be used. A very popular approach, which estimates posterior probability distribution directly from data is the Markov Chain Monte Carlo (MCMC) method. Those MCMC techniques are often used to estimate joint posterior probability distributions (that is the posterior density of all possible combination of the different parameter values describing the model) (Excoffier & Heckel, 2006). Instead of exploring the whole parameter space MCMC concentrates on the high-likelihood proportion of the space (Excoffier & Heckel, 2006).

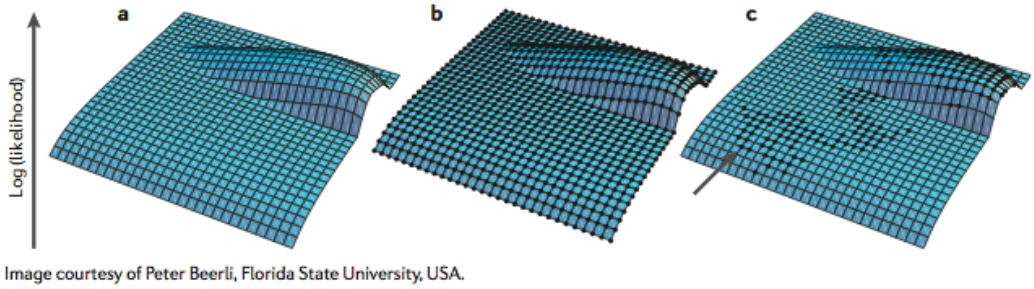

Figure 1.8: **Likelihood surface of the parameter space.** a) Likelihood surface of the parameter space describing a model. As can easily be seen not all points of the parameter space have the same likelihood. Goal of the MCMC method is to find the maximum of this likelihood surface. b) The black dots represent the whole parameter space, which might be way to big to explore. c) Thus, instead of exploring the complete space MCMC starts at a random point in the parameter space and moves around the surface. An actual step is taken under a certain requirement that is that the next point has to have a higher likelihood than the previous one. The markov chain makes sure that the next step is only dependent on the current one and not any previous states.

The principle of the MCMC method is summarized in figure 1.9. MCMC explores just parts of the parameter space and thus facilitates the estimation of the posterior density. This is enabled by the feature of the MCMC chain that, when exploring a likelihood surface a new step is just taken if the new state has a higher likelihood than the actual one. The markov chain thereby makes sure that the next step is only dependent on the current one and not any previous states. The quality of the results are dependent on the starting point and the length of a chain (Excoffier & Heckel,

2006). A very popular algorithm used in MCMC methods is the Metropolis-Hastings algorithm (described in Metropolis *et al.* (1953) and Hastings (1970)). Thereby, a new state is accepted with a probability *h*, known as the Hastings ratio:

$$h = min\{1, \frac{P(D|\eta')\pi(\eta')q(\eta' \rightarrow \eta)}{P(D|\eta)\pi(\eta)q(\eta \rightarrow \eta')}\} \tag{1.4}$$

Where η' is the new set of parameter values, $q(\eta \rightarrow \eta')$ stands for the probability to nominate a new state η' from the current state η, π is the prior distribution, *D* the data, *P* the probability distribution and *min* is the minimum. If the new state is rejected the chain persist in the current state. The 'proposal kernel' *q* is crucial for the efficiency of the MCMC run (Marjoram & Tavaré, 2006). If large changes are suggested the algorithm will be much less likely to accept a step (the denominator will be much bigger than the numerator and thus the probability of accepting the new state will be much less than 1) (Marjoram & Tavaré, 2006). Thus, chages are typically small (Marjoram & Tavaré, 2006).

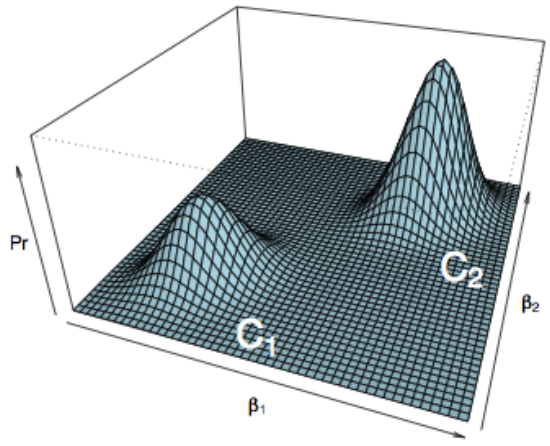

Figure 1.9: **Likelihood surface with two local maxima after Hadfield (MasterBayes Tutorial).** This example shows a joint posterior probability distribution of the parameters β_1 and β_2. This parameter space bears two local maxima. To avoid the MCMC chain to get stuck in C_1 different chains have to be run starting from randomly chosen starting points to reach the maximum at C_2.

1.4.2 Coalescent Theory

Purpose of this subsection is to summarize the basic concept of coalescent theory underlying the inference of the likelihood *P(D|θ)* of the data (MCMC methods) or the simulations to approximate the posterior distribution *P(θ|D)* (ABC method).

In order to describe the basics of coalescent theory we first have to define the reproduction model underlying the inheritance relationship (gene genealogy).

Definition 1.2 (gene genealogy)

Ancestral relationship for a given gene.

Models used for theoretical inference should be as simple as possible but at the same time capture the essential biological features of the population they are applied to. The most commonly used model was invented in the early 30s by Fisher (1930) and Wright (1931). The so called Wright-Fisher model is a very simple model describing an idealized population.

1.4.2.1 Wright-Fisher Model

Assumptions of the Wright-Fisher model:

- *Discrete non-overlapping generations*
  All individuals of a population die after one generation and are replaced by offspring. All individuals of the population share the same generation time (e.g. in lemmings 1 year). One major simplification of this reproduction model is assuming a simultaneous replacement (all individuals or genes respectively give birth and die at the same time). In most cases this assumption will be violated in natural population and models including overlapping generations have been formulated later on (e.g. Moran (1958)). However, it has been proven that these models are of little practical consequence, meaning that they produce quite similar genealogies.

- *Haploid individuals or population-subdivision (male and female)*
  This simplification is useful in the majority of cases. We assume a population size of $2N$ genes for haploid models.

  Definition 1.3 (haploid individuals)

  Haploid individuals exhibit only a single copy of a gene.

  When working with mitochondrial DNA this assumption is sufficient, since mtDNA of animals is exclusively inherited on the maternal line.

- *Constant and finite Population Size*
  The population size $2N$ is assumed to be constant over time and finite. Violations of this assumption have major impact on important quantities describing the genealogy. Anyhow, the coalescent framework can be extended to allow for simple deviations. Examples for genealogy shapes under different demographic scenarios are given in figure 1.10.

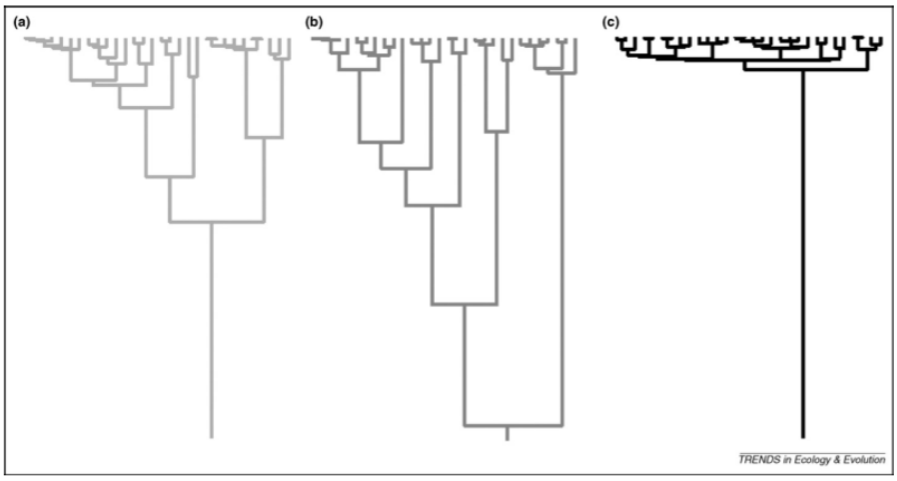

Figure 1.10: Genealogies under different demographic scenarios: (a) constant-populationsize, (b) exponential decline and (c) exponential growth. From Kuhner (2009)

- *No Recombination*
  Recombination can occur in most DNA sequences apart from Y-chromosme and mitochondrial DNA. When implementing recombination the genealogy will become a graph rather than a tree.

- *Random mating*
  Meaning that each parent can be chosen randomly assuming no population-subdivision. *Random mating* is an important feature of the wright-fisher model since violations have a strong effect on the shape of the genealogy. When studying natural populations one has to consider the possibility of population structure or -subdivision.

- *No Selection*
  In the wright-fisher model all individuals are equally fit, meaning that the probability of an individual to have progeny is the same for all individuals of the sample.

1.4.2.2 Discrete-time Coalescent

The coalescent theory was invented by Kingman (1982) and generalized by Griffiths & Tavare (1994). In general the coalescent theory describes the probability of randomly taken DNA fragments from a sample to share a common ancestor at time t in the past. As mentioned above the coalescent infers the genealogy of randomly sampled DNA fragments backwards in time and thus just counting the actual lineages that gave rise to the sample in question based on the wright-fisher model of inheritance.

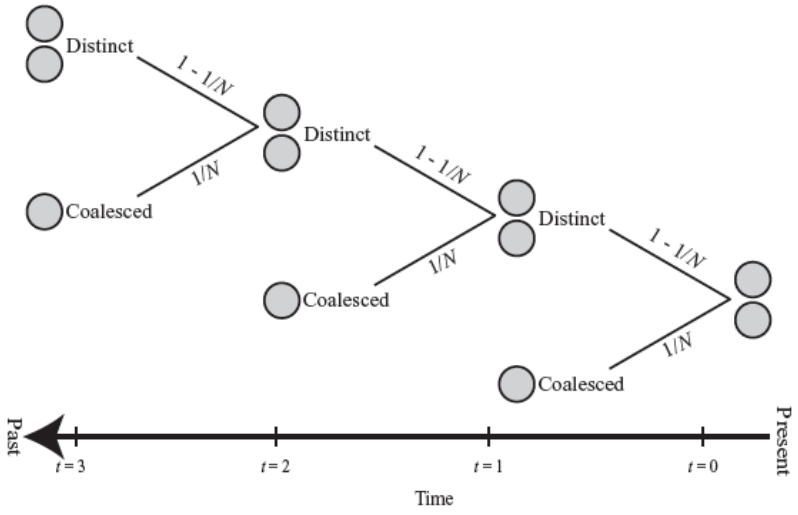

Figure 1.11: **Decision Tree for the coalescent model after Otto *et al.* (2007).** This tree shows the path needed to have a coalescent event at *t=n* in the past. E.g. the probability to have a coalescent event at *(t = 2)* is $P(T_2 = 2) = (1 - 1/2N)^{2-1}\ 1/2N$, because in order to have a coalescent event at *(t=2)* the samples have to have different ancestors at time *(t = 1)*.

Imagine a population of 2N individuals. The probability of 2 randomly chosen samples to descend from the same ancestor is $p = 1/2N$ and therefore the probability of having different ancestors is $q = 1 - p = 1 - 1/2N$.
In our example (figure 1.12) we have a population of 10 individuals thus the probability of 2 randomly picked sequences to coalesced in the previous generation is $p = 1/2N = 1/10 = 0.1$ and $q = 0.9$. If we want to calculate the probability of a sample to coalesced at time *t* in the past we can follow the decision tree provided in figure 1.11. Thus the chance to coalesced at time *(t = 1)* is $P(T_2 = 1) = 1/2N$ and the probability at *(t = 2)* is $P(T_2 = 2) = (1 - 1/2N)^{2-1}1/2N$, because in order to have a coalescent event at *(t=2)* the samples have to have different ancestors at time *(t = 1)*. We now can formulate the probability distribution of 2 randomly sampled sequences to share a common ancestor at time *t* in the past as follows:

$$P(T_2 = t) = (1 - \frac{1}{2N})^{t-1}\frac{1}{2N} \tag{1.5}$$

Where $(1 - 1/2N)$ refers to unsuccessful - and $1/2N$ to successful coalescent events, respectively. We can then calculate the expected time until coalescence of the 2 sequences , that is the mean time until coalescence of the 2 sequences using equation 1.9:

$$E[T_2] = \frac{1}{p} = \frac{1}{1/2N} = 2N \tag{1.6}$$

Based on equation 1.6 we can infer that the probability of the 2 sequences sharing a common ancestor in the previous generation gets smaller the bigger the population size. Thus, the larger

the population size the less related are its individuals. Again in our example of 10 individuals the mean coalescent time would be 10 generations in the past.

We can further calculate the variance in coalescent times using equation 1.10:

$$Var[T_2] = \frac{(1-p)}{p^2} = \frac{1-1/2N}{1/(2N)^2} = (2N)^2(1-1/2N) = 2N(2N-1) \tag{1.7}$$

This clearly demonstrates the extreme variability of the exact coalescent time.

Primer 1.2 (Geometric Distribution)

$$P(X=k) = p(1-p)^{k-1} \tag{1.8}$$

for k= 0,1,2,...,n

$$E[X] = \frac{1}{p} \tag{1.9}$$

$$Var[X] = \frac{(1-p)}{p^2} \tag{1.10}$$

In our case we use the geometric distribution to measure the number of independent trials (k) until the first coalescent event.

The coalescent can be extended to *n* samples $(n \geqslant 2)$ and scaled in continuous time to improve the computational efficiency, but it is out of the scope of this chapter. For further extensions on the coalescent read Hein *et al.* (2005) and Wakeley (2008) .

In the following I will give a simplified overview of the actual approach to infer a genealogy of randomly sampled DNA sequences using the coalescent theory.

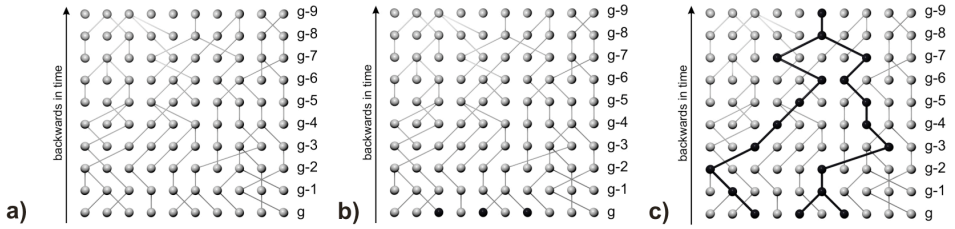

Figure 1.12: **Schematic reconstruction of a population of 10 individuals over 10 generations under the Wright-Fisher Model.** a) Shows the relationship of the individuals, when assuming random mating. b) One picks 3 individuals randomly and c) trace back the lineages until the most common ancestor is reached (here in the eighth generation backwards in time).

Imagine a population of 10 individuals as shown in figure 1.12a. Following the assumption of the wright-fisher model we can then apply lineages (lines of descent) randomly to our sample, because of an equal chance of the individuals to give birth and no population subdivision. In the next step we sample randomly from the present generation *(t = 0)*. In our example we choose 3 individuals (remember that individuals refers to DNA fragments rather then the organisms). Following the lineages back in time we infer the time until the *most recent common ancestor* of our 3 samples 1.12 c).
In the next step we superimpose mutations forward in time according to a mutation model (infinite - or finite sites model) as can be seen in figure 1.13. On the contrary to phylogenetic trees a branching event or a coalescent event (when thinking backwards in time) refers to a replication event of a sequence rather then a divergence event. A replication event does not mandatory include a mutation event, which means that the two coalescing DNA fragments can share the same nucleotide sequence.
Based on the shape of the genealogy we can then infer the underlying demography using the mathematical formulation of the coalescent. Genealogies constructed using different demographical scenarios are shown in figure 1.10. When working with temporal sampling we use serial-sampling coalescent as formulated by Rodrigo & Felsenstein (1999) rather then the original formulation of the coalescent by Kingman (1982) and its generalization by Griffiths & Tavare (1994).

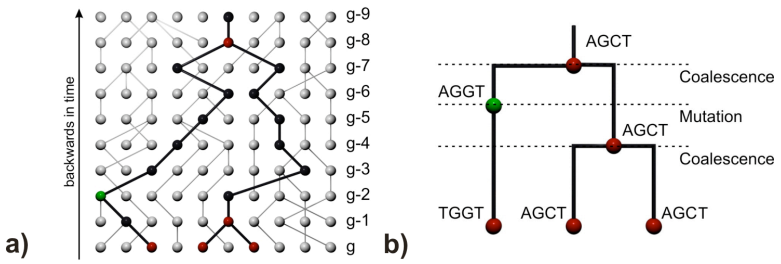

Figure 1.13: **Schematic reconstruction of a population of 10 individuals over 10 generations under the Wright-Fisher Model.** a) Mutations (green dot) are superimposed forward in time. b) This genealogy can also be shown in a so called gene tree. Note that branching events are independent from mutation events.

1.4.3 Selection of the best fitting Substitution Model

A critical step in population- and phylogenetical analysis is the choice of the right nucleotide substitution model. Posada and Crandall developed a model testing software called ModelTest 3.7 Posada & Crandall (1998). It uses log likelihood scores (as calculated by the software Paup* version 4.0b10 Swofford & Sullivan (2003).) to evaluate the best fitting model using Akaike Information Criterion (AIC) and the Hierarchical Likelihood Ratio Test (hLRT). The likelihood is the probability *P* of the data *D* over the model of evolution and possible parameter values as conveyed in

$$P(D|M, \theta, \tau, v) \tag{1.11}$$

Where, *M* is the model of evolution, θ a vector of K model parameters θ=(θ_1,θ_1,...,θ_K), τ a Tree topology and *v* a vector of S branch lengths Posada & Buckley (2004).
The AIC is calculated using maximized log-likelihood (*l*) and the number of parameters estimated (*K*):

$$AIC = -2l + 2K \tag{1.12}$$

The AIC can be seen as the loss of information when approximating the empirical process of nucleotide substitution with a model (e.g. HKY85). Thus, the smallest AIC is the best fitting one (Posada & Buckley, 2004).
The hLRT performs pairwise likelihood ratio tests in a specific sequence until the best fitting model is found.

$$LRT = 2(l_1 - l_0) \tag{1.13}$$

Where, l_0 is the null and l_1 refers to the alternative model, respectively. If the calculated P-value is smaller than a certain threshold (significance level) the alternative model fits significantly better then the null model. Posada & Buckley (2004)
According to Posada and Buckley 2004 the AIC is superior in the possibility to find the right model of evolution to the commonly used hLRTs and moreover offers important advantages (see Posada and Buckley 2004, page 84, table 4). The importance of the right model choice is crucial for phylogenetic reconstructions (Lemmon & Moriarty, 2004; Lockhart *et al.* , 1996).

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Buffers and Solutions

Type	Company	Country
Binding Buffer	self-kade	-
0.5M EDTA, pH 8.0	self-kade	-
Extraction Buffer	self-kade	-
1x TE	self-kade	-
10x Buffer Tango	Fermentas	St.Leon-Rot, DE
10x ligation buffer	New England Biolabs (NEB)	Frankfurt am Main, DE
10x ThermoPol buffer	NEB	Frankfurt am Main, DE

Table 2.1: **Buffers and Solutions**

Binding Buffer(50ml)
29.5g GuSCN → 5M
5ml Sodium acetate, 3M

0.5M EDTA, pH 8.0(500ml)
93.06g EDTA → 0.5M
fill up with H_2O to 450ml
≈ 11g NaOH
measure pH
fill up with water to 500ml

Extraction Buffer(50ml)
45ml 0.5M EDTA, pH 8.0 → 0.45M
1.25ml 10mg/ml Proteinase K → 0.25mg/ml
3.75ml H_2O
always to be prepared fresh

Silica Suspension
4.8g Silica
filled up with H_2O to 40ml, vortex and let sit for 1h
take of 39ml, transfer into a new tube (discard supernatant)
let sit for another 4h
take off 35ml (discard supernatant)
add 48μl 30% HCl
vortex, aliquot and store in the dark

1xTE(50ml)
100μl 0.5M EDT, pH 8.0 → 1mM
500μl 1M Tris, p.H 8.0 → 10mM
fill up with water to 50ml
always to be prepared fresh

2.1.2 Chemicals

Type	Company	Country
Ethylendiamintetraacetic acid (EDTA)	Sigma	Taufkirchen, DE
Guanidiniumthiocyanate (GuSCN)	Sigma	Taufkirchen, DE
Sodium acetate Buffer, 3M, pH 5.2	Sigma	Taufkirchen, DE
Sodium Hypochlorite	Merck	Dresden, DE
Agarose SeaKem LE	Cambrex	Rockland, USA
Ampicillin	Sigma	Taufkirchen, DE
5-Brom-4-Chlor-3-indoxyl-β-D-galactosid (X-Gal)	Loewe Biochemica	Sauerlach, DE
Ethanol (absolute)	Merck	Dresden, DE
Ethidiumbromid	Roth	Karlsruhe, DE
$MgCl_2$ (25 mM)	Applied Biosystems	Foster City, USA
Silica	Sigma	Taufkirchen, DE
ATP 100mM	Fermentas	St.Leon-Rot, DE
dNTPs (25mM each)	GE Healthcare	Munich, DE
BSA	Sigma-Aldrich	Taufkirchen, DE
Water, HPLC-grade	Sigma	Taufkirchen, DE
DNA-loading dye	Fermentas	St.Leon-Rot, DE
Ethidium bromide	Sigma	Taufkirchen, DE
50% PEG-4000	Fermentas	St.Leon-Rot, DE
PEG 8000	Promega	Madison, USA

Table 2.2: **Chemicals**

2.1.3 Enzymes

Type	Company	Country
Proteinase K	Sigma	Taufkirchen, Germany
AmpliTaq Gold DNA Polymerase	Applied Biosystems	Foster City, USA
T4 DNA polymerase	Fermentas	St.Leon-Rot, DE
T4 polynucleotide kinase	Fermentas	St.Leon-Rot, DE
T4 ligase	Fermentas	St.Leon-Rot, DE
Bst DNA polymerase, large fragment	NEB	Frankfurt am Main, DE

Table 2.3: **Enzymes**

2.1.4 Oligonucleotides

All primer oligonucleotides were purchased from Invitrogen (www.invitrogen.com).

Name	Primer Sequence 5'-3'	Fragment length (including Primers)
H2Dt1F	GGATGCTATGACTCACCA	
H2Dt1R	TGGAATTATGTGGATGAATG	95bp
H2Dt2F	GTCTAGCTGGACTTATCATTCA	
H2Dt2R	AGTCCTTGAAGGTGGAATA	131bp
H2Dt3F	ATATAACAGTACTAATCATGCAAA	
H2Dt3R	GATAGGCGATTTCAGGT	148bp
H2Dt4F	AAAACAAGCTTAGCCAAA	
H2Dt4R	CCTGATACTGGTTTTTAAGTCA	142bp
H2Dt5F	CAGGACCTCTCATCCATT	
H2Dt5R	AAGCTACATCAACATTTATCA	152bp
CBDt1F	AGCAACAGCATTCTCGT	
CBDt1R	CCTCGTCCTACGTGTAAG	126bp
CBDt2F	GGCTAATTCGCTACATACA	
CBDt2R	ACAGCGAATAGGAGGATAAT	139bp
CBDt3F	GGCTCCTATAACATAATTGAA	
CBDt3R	AAGGATATTTGTCCTCATGG	107bp
CBDt4F	AGCAACAGCATTCATAGG	
CBDt4R	CTCAGATTCATTCTACTAGGG	125bp

Table 2.4: **Primer-Oligonucleotide Sequences.** Primers used in the 2-step multiplex approach (Roempler *et al.* , 2006) and the 60cycle PCR approach (Hofreiter *et al.* , 2002). Fragment length of the amplicons including primers did not exceed 152bp.

2.1.5 Kits

Type	Company	Country
DNAeasy Blood and Tissue Kit (50)	QUIAGEN	Hilden, Germany
TOPO TA Cloning Kit	Invitrogen	Karlsruhe, Germany
QIAquick PCR purification Kit	QUIAGEN	Hilden, Germany
AMPure SPRI PCR purification kit	Agencourt	Beverly, USA
DyNAmoł SYBR Green qPCR Kit with ROX	Finnzymes	Espoo, Finland

Table 2.5: **Kits**

2.1.6 Machines and Computers

Type	Name	Company	Country
Gel Visualisation	Eagle Eye II	Stratagene	Heidelberg, Germany
Capillary Sequencer	ABI 3730	Applied Biosystems	Foster City, USA
Pyrosequencer	454 FLX	Roche	USA
Biorobot	Biorobot 9600	Qiagen	Hilden, Germany
Biorobot	epMotion	Eppendorf	Hamburg, Germany
Spectrometer	Nanodrop-Spektrometer	Kisker-Biotech	Steinfurt, Germany
Thermocycler	PTC 200 / PTC 225	MJ Research	Watertown, USA
Quantitative PCR	MX 3005P QPCR System	Stratagene	La Jolla, USA
Thermomixer	Thermomixer compact	Eppendorf	Hamburg, Germany

Table 2.6: **Machines used in the laboratory.**

Computer	CPU number	CPU speed	Memory total
bionc04	16	2260MHz	48,9 GB
bionc05	16	2261MHz	48,9 GB
bionc06	2	2004MHz	16,2 GB
MacBook	2	2130MHz	2 GB

Table 2.7: **Computers used for the bioinformatic analysis.** All bionc's used in this study are part of the Bionc Cluster of the Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

2.1.7 Plasticware and Filters

Type	Company	Country
MobiCols Columns	MoBiTec	Goettingen, Germany
Filter Large $10\mu m$	MoBiTec	Goettingen, Germany
GF/B Fiberglassmembrane $1\mu m$	OMNILAB	Munich, Germany

Table 2.8: **Plasticware and Filters**

2.2 Samples

All samples used in this study are summarized in table 2.9 and table 2.10.

Number	Sample	Age	Purpose	Locality	Origin
267	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
275	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
299	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
340	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
341	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
342	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
344	tissue	modern	primer design	Dikson, Western Taimyr	Dorothee Ehrich
PSX1	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX2	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX3	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX4	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX5	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX6	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX7	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX8	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX9	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PSX10	left jaw	modern	analysis	Pymva Shor	Nickolay Smirnov
PS10	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS1 1	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS12	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS13	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS14	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS15	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS16	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS17	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS18	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS19	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS30	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS31	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS32	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS33	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS34	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS36	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS38	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS39	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS49	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov
PS50	left jaw	11.500	analysis	Pymva Shor	Nickolay Smirnov

Table 2.9: **Sample List - PartI**

Number	Sample	Age	Purpose	Locality	Origin
PS3	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS4	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS5	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS20	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS1	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS21	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS22	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS23	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS24	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS25	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS26	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS27	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS28	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS29	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS51	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS52	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS53	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS54	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS55	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS56	left jaw	15.200	analysis	Pymva Shor	Nickolay Smirnov
PS2	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS6	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS7	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS8	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS9	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS40	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS41	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS42	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS43	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS44	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS45	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS46	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS47	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
PS48	left jaw	25.200	analysis	Pymva Shor	Nickolay Smirnov
JP01	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP02	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP03	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP04	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP05	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP06	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP07	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP08	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP09	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP10	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP11	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP12	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov
JP13	left jaw	< 1.000	migration analysis	Yangana-Pe-4	Nickolay Smirnov

Table 2.10: **Sample List - PartII**

Locality	Layers	Age/Radiocarbon dates	Calibrated dates (Fairbanks0107)
Pymva-Shor	surface	modern	
Pymva-Shor	L4	10.000 $\pm$ 250 (GIN-9004)	11.538 $\mp$ 423
Pymva-Shor	L6 up	13.090 $\pm$ 60 (CAMS-38221)	15.246. $\mp$ 122
Pymva-Shor	L6 low	21.910 $\pm$ 250 (TUa-11501)	25.244 $\mp$ 359
Yangana-Pe-4	L2	approx. 1000	

Table 2.11: **Sampling.** All samples were taken from two sample sites (Pymva Shor, Yangana-Pe-4) in the northern Ural. Radiocarbon dates and Layer assignment according to Golovachov and Smirnov 2008 Golovachov & Smirnov (2008) are shown as well as calibrated radiocarbon dates using the Fairbanks0107 curve Fairbanks *et al.* (2005).

2.3 DNA Extraction

2.3.1 Modern DNA extraction from tissue and bones

The 10 modern lemming samples for primer design were extracted using the DNAeasy Blood and Tissue Kit from QUIAGEN. About 20mg to 25mg tissue (muscles) was cut into small pieces and treated with 180μl Buffer ATL in a 1,5ml Eppendorf tube. Before vortexing for 10" 20μl of Proteinase K was added to digest proteins. The mixtures were incubated at 56 °C on a rocking platform for 2h. After incubation the tubes were vortexed for 15", 200μl AL Buffer was added to the samples and again vortexed thoroughly for additional 10". This step was repeated using 200μl of absolute Ethanol. The mixture was then pipet into a DNAeasy Mini spin column, placed in a 2ml collection tube and centrifuged at 8.000 rpm for 1'. The flow-through was discarded and the DNAeasy Mini Spin column placed into a new 2ml collection tube. Followed by the addition of 500μl AW1 Buffer and an additional centrifugation step at 8.000 rpm for 1'. This step was repeated using 500μl AW2 Buffer but this time centrifuged at 13.500 rpm for 3' to dry the membrane in the DNAeasy Mini Spin column. The extracted and purified DNA was eluted afterwards using 200μl AE Buffer by centrifuging for 1' at 8.000 rpm after incubation for 1' at RT. The elution step was repeated for maximum DNA yield.

All modern bone samples included in the analysis were extracted using the Ancient DNA Extraction Protocol (see subsection 2.3.2) in the post PCR laboratory, but under a special hood to avoid contamination with PCR products (in the air).

2.3.2 Ancient DNA Extraction from Bones and Teeth

The 77 samples (64 from Pymva Shor and 13 from Yangana-Pe-4) were ground to fine powder using a mortar and a pistil. For each sample a separate mortar and pistil were used, which were treated with Sodium Hypochlorite o/n after usage to destroy the DNA. We used only left jaws for DNA extraction to rule out the possibility of sampling the same individual twice. Quantities ranging from 50 to 150mg of bone powder then were mixed with 5ml of Extraction Buffer. The tubes (15ml Falcon tubes) were sealed with parafilm and rotated o/n at RT in the dark. 24 samples

including 3 negative controls (no bone powder) were processed in parallel. The next day the tubes were centrifuged for 2' at 4000rpm using a Macrofuge centrifuge. The supernatant from the centrifugation step was then added to a fresh labeled tube including 100μl silica suspension and 2.5ml Binding Buffer to bind the DNA to the silica particles. The bone pellets were kept in the freezer for possible reextractions. The tubes were once again sealed with parafilm and rotated for 3h at RT in the dark. After an additional centrifugation step for 2' at 4000rpm the supernatant was taken of and stored in fresh labeled tubes in the fridge also for possible reextractions. The silica pellets then were resuspended in 400μl Binding Buffer and transfered onto the extraction columns and centrifuged for 1' at 13000rpm using a Microfuge centrifuge with closed lids. The remaining silica particles in the tubes were then obtained by washing the tubes with an additional amount of 400μl Binding Buffer and transfered to the extraction columns. After an additional centrifugation step (1' at 13000rpm) the silica with the bound DNA was washed twice using 400μl of Washing Buffer and centrifuged (1' at 13000rpm) again. In case the silica particles were still brown additional washing steps were performed. In order to dry the silica particles the columns were centrifuged once again for 1' at 13000rpm. In order to elute the DNA the columns were put into a new labeled 1.5ml Eppendorf tube, 50μl of 1xTE were added and incubated for approximately 10'. This was done once more with additional 50μl of 1xTE. The extracts were stored in a freezer at -20°. To obtain the maximum of recovery the extracts should be stored in siliconized tubes (to prevent binding of the DNA to the tube walls).
The basic approach was published by Rohland & Hofreiter (2007) and further improvements were developed by Rohland in 2008 (unpublished).

2.4 DNA Amplification

2.4.1 Primer Design and DNA Amplification

Amplification of the modern *Dicrostonyx torquatus* DNA for primer design were carried out using universal primers (*L15926* and *HN00651* from Kocher *et al.* (1989), *stacy2H* from Stacy *et al.* (1997), *cwm4L* and *cwn5H* after Matson & Baker (2001)). The PCR mix contained 17.6μl HPLC purified water, 2.5μl PCR Buffer II, 2.5μl $MgCl_2$ (25mM), 0.2μl dNTP's (25mM each), 0.2μl Taq Gold Polymerase, 0.5μl primer forward and 0.5μl primer reverse. 1μl template were amplified in a 30 cycles reaction (94 °C for 5' to initiate the Taq Gold Polymerase and 30 cycles of 94 °C for 30", 55 °C for 45" and 72 °C for 45" followed by 72 °C for 7'. The reaction was cooled down afterwards to 10 °C). Primers for targeted PCR amplification were designed using the web-based tool *Primer 3.0* (Rozen & Skaletsky, 2000) applying mostly default parameters using the sequenced modern samples. Due to the fragmented nature of ancient DNA the maximum amplicon size should not exceed about 150bp (including primers) depending on the quality of the aDNA in the extract but at the same time has to assure confident species assignment not to amplify contaminants in the reaction chemicals used (for further details see Leonard *et al.* (2007)).The maximum melting temperature difference between forward and reverse primer was reduced to 3 °C. Amplifications were performed using the Multiplex PCR after Roempler *et al.* (2006) with a 15 min initial Taq activation step at 95 °C, followed by 40 cycles of 94 °C for 30s, 53 °C for 40s, 72 °C for 40s and a final incubation for 10 min at 72 °C. The composition of the PCR mix used in phase 1 and 2 are shown in FIG.2.1 and 2.2. The retrieved amplification products of phase 1 was subsequently diluted 1:30

and 5μl of the dilution were used as template in phase 2 of the Multiplex PCR.

Replication of the fragments was performed either using the same set-up for the 2-step multiplex or a 60 cycle PCR approach described in Hofreiter et al.(Hofreiter *et al.* , 2002) (with a 15 min initial Taq activation step at 95 °C, followed by 40 cycles of 94 °C for 30s, 53 °C for 40s, 72 °C for 40s and a final incubation for 10 min at 72 °C). The PCR mix contained 2 units Taq Gold, 10x Taq Gold Buffer, dNTP's 25mM each, primer 10μM each, 10mg/ml BSA and 4,0mM MgCl2 in a total reaction volume of 20μl. Along with each amplification 2 to 3 PCR blanks were carried out to monitor for contamination and crosscontanimnation during the PCR setup.

20µl total volume		final conc.
3.2µl	$MgCl_2$ (25mM)	4mM
2.0µl	Gold Buffer10x	1x
2.0µl	BSA (10mg/ml)	1mg/ml
0.2µl	dNTP's (25mM each)	250µM each
0.4µl	Taq Gold	2U
3.0µl	Primer Mix 1µM each outer primer in primer Mix	150nM each primer
1.5 -5.0µl	water, blank, DNA extract or dilution	
up to 20µl	ddH_2O	

Figure 2.1: Composition of the PCRmix for the first step (Phase1) of the multiplex PCR using 5μl of template

20ul total volume		final conc.
3.2µl	$MgCl_2$	4mM
2.0µl	Gold Buffer10x	1x
2.0µl	BSA (10mg/ml)	1mg/ml
0.2µl	dNTP's (25mM each)	250µM each
0.05µl	Taq Gold	0.25U
3.0µl	Primer Mix (respective inner primer pair 10µM fwd + rvs each)	1.5µM
5.0µl	dilution of Phase I	
up to 20µl	ddH_2O	

Figure 2.2: Composition of the PCRmix for the second step (Phase2) of the multiplex PCR using 5μl of the diluted template of phase 1.

2.5 Sequencing

The amplified fragments were sequenced using either the direct Sanger sequencing method (samples for primer design), 454 pyrosequencing (samples included in the analysis) or Sanger sequencing after Topo-TA Cloning (for replication of the last fragments). The Sanger sequencing was performed on an ABI Capillary Sequencer 3730 either directly or after cloning using the Topo - TA Cloning Kit from Invitrogen. Pyrosequencing was performed after DNA-tagging and library

preparation on a 454 FLX platform (ROCHE).

In the following I will describe the methods used to prepare the amplified DNA fragments for the different sequencing methods in detail.

2.5.1 Direct Sanger Sequencing

All modern samples for the primer design step were sequenced directly on an ABI Capillary Sequencer 3730 using universal primers from Kocher *et al.* (1989); Stacy *et al.* (1997); Matson & Baker (2001) after a purification step using the Millipore Purification System and thus, extricating all PCR's from small molecules like salt, unincorporated dNTP's and primers before sequencing.
All PCR's were diluted to a volume of 100μl with water and transfered to the purple Millipore plate using a multichannel pipet. The plate then was placed on the vacuum manifold and vacuum was applied (approximately 600 mbar) until all wells have emptied plus additional 30". The dried plate was removed from the manifold and 20μl and 80μl HPLC purified water was added to the wells, respectively, depending on the amplified DNA region (HVR1 and HVR2). After sealing with the plate cover the DNA was resuspended by shaking vigorously. The purified PCR products were transfered to a new PCR plate using the multichannel pipet.
The recovered DNA amount was measured using the Nano-Drop. All samples were normailzed to a DNA amount of 5ng. 6.5μl of diluted DNA was mixed with 0.5μl BigDye mix, 2μl Sequencing Buffer and 1μl primer (2.5μM each). The reactions were cycled for 25 times (96 °C for 1' (inititation step), 96 °C for 20", 53 °C for 30" and 60 °C for 4'). The cycle sequencing reaction was cleaned up using an Ethanol/Sodiumacetate precipitation. 1μl 3M Ac and 25μl EtOH (absolut) were added to each sample (10μl template). The plate was sealed with a cover, mixed carefully and spinned briefly. The plate was incubated for 30' at RT. Afterwards the reactions were centrifuged at 4.000rpm for 30' using a Megafuge centrifuge. The supernatant was dicarded by inverting the plate. In the next step the inverted plate was centrifuged at 1.800rpm for 1' to dry the plate. After the centrifugation a washing step (using 150μl 70 % EtOH) was performed. The samples along with the EtOH were spinned down at 4.000rpm for 10' at RT. The supernatant was discarded by in verting the plate and centrifuged again at 1.800rpm for 1' to dry. The plate was incubated at RT in the dark for at least 30' before sealing with tape foil and storage at -20 °C (for sequencing).

2.5.2 Pyrosequencing Approach

2.5.2.1 Direct Multiplex Tagging

I used the Direct Multiplex Sequencing (DMPS) method after Stiller *et al.* (2009) for targeted high-throughput pyrosequencing. On the contrary to the original method I used this tagging approach in the second step of the two-step multiplex PCR in order to tag all 2nd step amplicons for each individual and mix the different specimens just for sequencing.

- First the 2nd step PCRs were quantified with the plateread method after Meyer *et al.* (2008b) (Online appendix), but using 1:500 dilution of SYBR Green instead of 1:20 dilution of Pico Green. I used an Eppendorf epMotion Robot to normalize the DNA amount of all individual fragments. All amplicons of the same specimens were then mixed to get a pool of DNA fragments with up to 500ng (100ng minimum) for the tagging.

- In the next step the pooled PCR fragments were purified using SPRI Beats (see Meyer *et al.* (2008b)). A custom buffer (2.5g PEG-8000, 1.5g 10x TE, 3g 5M NaCl, filled up with dH_2O to 10g) was used instead of the provided one. In order to change the buffer 2ml of ready to use SPRI beats were put on a magnetic rack until all beats were sticking on the tube wall. The buffer was taken off and discarded. The beats were washed twice with 200μl 1x TE. The SPRI beats were stored in 200μl 1x TE in the fridge. 2μl of the beats were used per 10μl template along with 20μl of custom PEG buffer in each purification step.

Tech 2.1 (Ampipure SPRI Beat Purification)

2μl of SPRI beats were mixed per 10μl template with 20μl of custom PEG buffer. After vortexing for 20" the mixture was incubated for 5' in a rack on the labor bench, followed by a second incubation step on the magnetic rack for 3'. After clearance of the mixture the liquid was taken off and discarded. The captured beats were washed twicce with 150μl ethanol (70%) and dried for 10' on the Thermocycler (37 °C). After checking if all the ethanol has evaporated the beats were mixed with 15μl 1x TE on a vortexer for 20". After a 2' incubation step the plate was put on the magnetic rack again and the supernatant taken off and transfered to a new labeled plate. The old plate was sealed and stored in the freezer at -20 °C.

I determined the right amount of PEG-8000 to add in the custom buffer for the different purification steps using gelelectrophoretic separation. Using a Fermentas ladder I checked the different PEG-8000 concentration for their ability to get rid of DNA fragments under a certain length threshold. Goal of the SPRI purification steps is to get rid of all unbound DNA fragments in the mix to avoid cross reaction in the different tagging steps.

- After the purification step the DNA fragments were blunt-ended and phosphorylated to ensure sufficient ligating of the tags. In the blunt-ending and phosphorylation step 15μl of the template were mixed with 9.48μl HPLC purified water, 3μl 10x Buffer Tango, 0.12μl dNTP's (25mM each), 0.3μl ATP (100 mM), 1.5μl T4 Polynucleotide Kinase (T4 PNK) and 0.6μl T4 Polymerase (T4 Pol). The solution was mixed carefully by pipetting up and down and incubated in a Thermocycler for 5' at 12 °C and additional 5' on 25 °C.

- After the incubation a SPRI beat purification step (with an 25% PEG Buffer) was performed.

- In the next phase the adapter oligomers were ligated to the blunt-ended and phosphorylated DNA. Therefore 13.5μl of template were first mixed with the adapters A and B. The A-adapter includes a 7bp long barcode to identify the fragments after the sequencing run, wheareas the B-adapter is universal in all ligation steps. After vortexing the template and the adapters 8.5μl water, 3μl 10x T4 Ligase Buffer, 3μl 50% PEG-4000 and 1μl T4 Ligase were added. The ligation reaction was performed in the Thermocycler at 22 °C for 20'.

- Two SPRI beat purification steps were performed imediatley after the incubation to avoid

unspecific ligation. The purification was performed using an 18% PEG Buffer. I used a lower concentrated buffer to get rid of adapter-dimers (80bp).

- To finish the tagging phase a BST-adapter fill-in step was performed. 15μl of template was incubated at 37 °C for 20' in a Thermocycler after additing of 10.7μl water, 3μl 10x Thermopol Buffer, 0.3μl dNTP's (25mM each) and 1μl BST Polymerase.

- 27μl were eluted after an additional SPRI purification step.

- In the next step the adapters were extended via PCR amplification. Therefore 10μl template were mixed with 5μl water, 2μl 10x Buffer Gold, 1.5μl $MgCl_2$ (25mM), 0.5μl make 454 forward and 0.5μl make 454 reverse primers (10μM each), 0.25μl dNTP's (25mM each) and 0.25μl Taq Gold Polymerase. 10 cycles were performed in the "make-454" PCR with a 95 °C for 12' (initiation step) followed by 10 cycles of 94 °C at 30", 60 °C for 40" and 72 °C at 40". A renaturation step at 72 °C was performed for 10' followed by a storage at 10 °C.

- A final purification step was carried out after the amplification to get rid of all dNTP's.

- In the very last step all libraries were quantified using the quantitative PCR approach after Meyer *et al.* (2008a), normalized and pooled. At last the sequencing library was quantified again using the quantitative PCR approach. A final amount of 6x10^5 copies per library were used for the 454 sequencing.

Tech 2.2 (Quantification: quantitative PCR method)

Quantification of the DNA amount in the sequencing libraries was performed using the quantitative PCR method as described in Meyer et al. *(2008a).*
2μl template was mixed to 8.77μl HPLC purified water, 10μl Dynamo Mastermix (from the DyNAmoł SYBR Green qPCR Kit with ROX), emPrimer forward and reverse 0.1μl each and 0.03μl ROX. 2μl of a self-made standard was used in the reaction (in replicates) comprising 10^2 to 10^9 copies. The copy number is determined according to the intensity of the fluorescence signal (dependent on the amount of incorporated SYBR Green in the DNA).

2.5.3 Classical Sequencing Approach

2.5.3.1 Topo-TA-Cloning

Step1: Cloning and Transformation

At first a master mix of 0.25μl salt-solution and 0.25μl vector (pCR-2.1-TOPO vector) was prepared. 1.5μl template DNA was mixed along with 0.5μl master mix and interspersed by pipetting up and down. This solution was incubated for 15' at RT. Competent cells (E.coli) were carefully unfreezed on ice and 8μl of the cells were mixed with the solution by careful tossing. After an incubation at RT for additional 15' the cells were put into a water-basin for the heat-shock. This heat-shock comprises 30" at 42 °C (crucial step!). The cells were stored in ice immediately after the heat treatment. 100μl SOC Medium was added to each sample and the cell growth afforded

by incubation at 37 °C for 1h on a thermal shaker (500rpm). After the growth phase the cells were plated on LB amp agar plates and incubated at 37 °C for 17h.

Step2: Lysis and colony PCR

Eight colonies per plate were picked and transfered into a water (30μl HPLC purified water) filled well of a 96-well plate. After brief shaking the pipet spikes were removed and the plate sealed with cap-strips. In order to lysate the cells the plate was put into a thermal cycler (95 °C for 10'). Afterwards the plate was centrifuged in a Megafuge centrifuge for 5' at RT at 4.000rpm. A colony PCR (94 °C for 3' initiation and 40 cycles of 94 °C for 20", 55 °C for 30" and 72 °C for 30" followed by 72 °C for 5') was performed subsequently. Therefore 37μl of the master mix (comprising 31.3μl HPLC purified water, 4μl 10x CB Buffer, 1mul primer M13 forward and reverse (10 μM each), 0.2μl dNTP's and 0.5μl Taq-Polymerase) was mixed with 3μl cell lysate.

Step3: Purification after colony PCR

In order to not chocking the capillaries of the ABI 3730 capillary sequencer the PCR's were purified using a QIAquick PCR Purification Kit (QUIAGEN) and the QUIAGEN Biorobot 9600.

Step4: Cycle-Sequencing and Cleaning-up of the cycle-sequencing reactions

The DNA amount of all samples were measured using the Nano-Drop technology subsequently after purification. For DNA fragments of roughly 300 - 400bp the samples were normalized to an amount of 5ng/μl. For the cycle-sequencing 6.5μl of diluted DNA were mixed with 0.5μl BigDye mix, 2μl Sequencing Buffer and 1μl primer (2.5μM each). The reactions were cycled for 25 times (96 °C for 1' (inititation step), 96 °C for 20", 55 °C for 30" and 60 °C for 4'). The cycle sequencing reaction was cleaned up using an Ethanol/Sodiumacetate precipitation. 1μl 3M Ac and 25μl EtOH (absolut) were added to each sample (10μl template). The plate was sealed with a cover, mixed carefully and spinned briefly. The plate was incubated for 30' at RT. Afterwards the reactions were centrifuged at 4.000rpm for 30' using a Megafuge centrifuge. The supernatant was dicarded by inverting the plate. In the next step the inverted plate was centrifuged at 1.800rpm for 1' to dry the plate. After the centrifugation a washing step (using 150μl 70 % EtOH) was performed. The samples along with the EtOH were spinned down at 4.000rpm for 10' at RT. The supernatant was discarded by in verting the plate and centrifuged again at 1.800rpm for 1' to dry. The plate was incubated at RT in the dark for at least 30' before sealing with tape foil and storage at -20 °C (for sequencing).

2.6 Computational Biology

2.6.1 Statistical Parsimony Networks

I calculated statistical parsimony networks for each timepoint (modern, 11.500y, 15.200y and 25.500y cal. BP) separately using the software TCS 1.21 (Clement *et al.* , 2000). These networks were then connected to a 3D network by combining haplotypes that were present at different time-

points. Furthermore, I constructed two other statistical parsimony networks using additional modern sequences from different regions surrounding PS and sequences from a second paleontological sampling site (Yangana-Pe-4) to look for possible gene flow. All additional modern sequences where obtained from complete mitochondrial genomes, which will be published elsewhere (Fedorov in prep).

2.6.2 Markov Chain Monte Carlo based bayesian inference

To infer demographic history directly from the data I used a markov chain monte carlo (MCMC) based bayesian approach as implemented in the software BEAST 1.4.8 (Drummond & Rambaut, 2007). I displayed the temporal demographic changes using the bayesian skyline plot (Drummond *et al.* , 2005). The estimates of the female effective population size were obtained from the lineage coalescent rate through time. I used the software ModelTest 3.7 (Posada & Crandall, 1998) to identify the best fitting nucleotide substitution model. The akaike information criterion (AIC) supported the GTR+I model (given the models provided by BEAST 1.4.8) (see table 3.4).This nucleotide substitution model was used to estimate the genealogy of the data. Three independent MCMC runs of 20.000.000 iterations each sampling every 1000th step with a burn-in of 2.000.000 were performed. I used the program TRACER 1.4 (Drummond & Rambaut, n.d.) to verify the effective sample sizes (ESS) and the trace of the MCMC runs. Moreover, Tracer 1.4 was used to display the bayesian skyline plot. To validate the support for the BSP reconstruction I used a constant size tree prior in BEAST 1.4.8 (Drummond & Rambaut, 2007). I further specified an uniform prior for the population size (from 0 to 200.000). Then I calculated log10 bayes factors (see table 3.1) for the two models using TRACER 1.4 (Drummond & Rambaut, n.d.).

2.6.3 Approximate Bayesian Computation

To further test my hypothesis I used an alternative method, namely the Approximate Bayesian Computation approach (Beaumont *et al.* , 2002). It approximates the posterior distribution by using information from a prior distribution and extensive simulations rather then calculating the likelihood of the data directly.
I used the software Bayesian Serial Simcoal (BayeSSC) (Anderson *et al.* , 2005) to simulate temporal data under a single-population-bottleneck model and R to perform the rejection algorithm and the local linear regression adjustment and smooth weighting after Beaumont *et al.* (2002). Bayesian Serial Simcoal is a modification of the program Simcoal 1.0 (Excoffier *et al.* , 2000) which allows for temporal sampling. Thus, it is a specially useful tool when working with serial ancient DNA data. In addition I simulated the data under a single-population-constant size model to reject the null hypothesis of a constant fNe over time and population cycles. I assumed a generation time of one year for *Dicronstonyx torquatus* as suggested in Tomilin (1957).

Single-population-bottleneck model
An advantage of Bayesian statistics is the possibility to incorporate previous knowledge in a so called prior. A uniform prior specifies a range of possible values of a given parameter assigning equal probabilities. Using uniform priors I specified a certain range of possible parameter values

to investigate in the simulation. By applying only uniform priors one minimizes the possibility to bias the results in a certain direction, but on the other hand one looses power in the analysis. So the right prior choice is crucial.
I incorporated uniform priors for the modern fNe, growth rate, event timing and severity, mutation rate and fNe before and after the bottleneck. Modern fNe ranges from 1 to 50.000 individuals and 1 to 200.000 and 1 to 5.000 for ancient fNe before and after the bottleneck, respectively. I simulated the data under possible values of the mutation rate ranging from 1 to 40%. The software Treepuzzle 5.2 (Schmidt *et al.* , 2002) was applied to investigate the shape parameter of the gamma distribution of mutation rates. 4 sample groups were used to summarize the data. I used the same range of possible values for implied parameters in the constant size model.

Summary statistics
I chose number of Segregating Sites, Pairwise Difference and Nucleotide Diversity because of their ability to unveil different aspects of population history (Ramakrishnan & Hadly, 2009).

Simulation and Analysis
I performed simulations of both models (constant size and bottleneck) comprising 5.000.000 iterations. BayeSSC provides a separate output file containing all calculated summary statistics and parameter values. All required summary statistics of the empirical data were estimated using the software DNAsp v5 (Rozas *et al.* , 2003) and can be seen in table 3.3.
All R 2.8.1 functions specific for the ABC approach were kindly provided by Mark Beaumont on his homepage. I first applied the rejection algorithm to test the two models against each other. Local linear regression and smooth weighting can then be applied using the function calmod(). This function estimates the posterior probability of a particular model using categorial regression. However, the rejection step revealed that 100% of the 1000 closest parameter estimates of my simulations affiliated to the bottleneck model.

In the next step I used rejection, local linear regression and smooth weighting to estimate the posterior probability distributions of the demographic parameters of the bottleneck model using the locfit function. The posterior probability distributions were estimates using the following algorithm:

A1 Sample parameter θ_i from a prior distribution $\pi(\theta)$
A2 Simulate data D_i using the model *M* with parameter θ_i
A3 Summarize D_i with summary statistics S_i
A4 Summarize empirical data D with summary statistics S
A5 Accept θ_i whenever $|| \ S_i - S \ ||\leqslant \delta$, otherwise reject
A6 repeat A1 - A5 until *k* acceptances have been obtained.

I applied a tolerance of $p_\delta = 0.001$ for parameter estimation and $p_\delta = 0.005$ for the model-testing, accepting the 5000 and 1000 closest estimates, respectively.

Chapter 3

Results

Climate changed repeatedly in the past. In order to properly understand its effects on mammal populations I extracted mitochondrial DNA from 64 ancient and modern collared lemmings from a single sample site in the norther Ural (Russia). I successfully amplified and sequenced 282bp of cytochrome B (CytB) and 426bp of the control region (CR) from all extracts.

Inference of temporal demographic changes

At first I calculated statistical parsimony networks for each time point (comprising all samples from a specific layer) separately and combined them to a 3 dimensional network to show the diversity change over time (see figure 3.1). As can easily be seen in figure 3.1 the diversity decreased drastically over time, showing the highest amount of variation at the LGM and the lowest in the modern population. Nearly all distinct haplotypes (more than one substitution in between) were lost between 15.200y and 11.500y cal. BP. The network shows two major haplotypes, which can be tracked in all or in all but the modern population, respectively. Between 11.500y cal BP and today the number of haplotypes decreased further from 8 to 3.
Then I constructed a Bayesian Skyline Plot (BSP) using MCMC based bayesian inference as implemented in the Software BEAST (Drummond & Rambaut, 2007). The software ModelTest 3.7 Posada & Crandall (1998) was used to identify the best fitting nucleotide substitution model for the data. The results can be seen in figure 3.4. Given the models supported in the BEAST analysis I used the GTR+I in the following analysis. The BSP (see figure 4.1) shows a decrease in the female effective population size (fNe) starting at the LGM. This fNe decline reaches its maximum between 14.000y and 15.000y cal. BP. After a short recovery phase it gets to another inflection point at around 12.000y cal BP and starts slightly to decrease again. The highest posterior density (HPD) intervals broaden between 11.500y cal. BP and today because of a loss of information in the data. The mutation - or clock rate, respectively calculated by BEAST was 9%.
I further used TRACER 1.4 (Drummond & Rambaut, n.d.) to calculate log10 bayes factors (BF) (Kass & Raftery, 1995) to compare the BSP reconstruction to a simple constant population size model. The analysis revealed a BF of 1.009 for the BSP reconstruction and a BF of -1.009 for the constant size model and thus, suggesting a higher support for the BSP reconstruction.

Bayes Factors	BSP	Constant Size
log10 Bayes Factors	1.009	-1.009

Table 3.1: **Model comparison using Bayes Factors.** I further used Tracer 1.4 (Drummond & Rambaut, n.d.) to compare the results of the Bayesian Skyline Plot reconstruction to a simple constant population model. The calculated log10 bayes factors suggest a higher support for the BSP reconstruction. A log10 BF of 0 implies an equal probability of both models, whereas a value bigger than 2 suggests a very strong and a values less than -2 a very poor support of the models (Kass & Raftery, 1995).

In order to verify the results of the MCMC method I used the approximate bayesian computation (ABC) approach (Beaumont *et al.* , 2002) to test a single-population-bottleneck model. The posterior probability distribution of the 'event timing', 'fNe after and before the bottleneck' and 'mutation rate' can be seen in figure 4.2. The parameter modes and quantiles are provided in table 3.2.

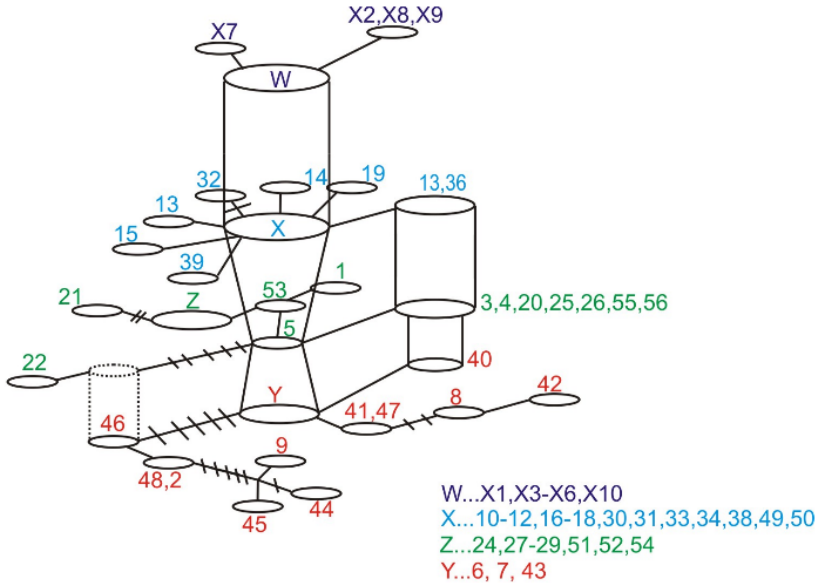

Figure 3.1: **Temporal haplotype network.** Statistical parsimony network comprising sampled haplotypes from each time point. Identical haplotypes between the time points were combined to form a three dimensional network to show the diversity change over time. The numbers indicate the haplotypes found and the colors the ages of the samples. Red...25.200 y cal. BP, blue...15.200 y cal. BP, light blue...11.500 y cal. BP and dark blue...modern.

The posterior probability distributions showed clear peaks apprising that the genetic data contained enough information to differ non-ambigiously from the uniform priors. This approach revealed a strong population decline of about 87% at 14.000 generations or y cal. BP (assuming a generation time of one year), respectively. Starting from an estimated ancient female effective population size of 28.331 it declined to about 3.690 and improved afterwards to a modern fNe of 11.090, given a mutation rate of 8.5% (Which is consistent with the mutation rate of 9% calculated in the Bayesian Skyline Plot). The bottleneck model was supported to 100% in the rejection step

of the model-testing approach. Thus, the single-population bottleneck was preferred clearly over the single-constant population size model as can be seen in figure 3.2

Parameter	**Mode**	**Quantile**					
		5%		**25%**		**50%**	
		Llim	**ULim**	**LLim**	**ULim**	**Llim**	**ULim**
Bottleneck time	13998.5	13392.9	14571.37	10784.98	16813.75	7227.734	20250.96
fNe after the Bn	28331.13	25535.92	31170.01	14507.96	42918.61	269.8741	59449.63
fNe before the Bn	3690.576	3591.244	3785.518	3117.085	4114.795	2251.086	4383.704
Mutation rate	5.5985e-05	5.2241e-05	5.9791e-05	3.7232e-05	7.5500e-05	1.6987e-05	9.7145e-05

Table 3.2: **Parameter - Mode and Quantiles (5%, 25%, 50%).** This table summarizes the results from the Approximate Bayesian Computation approach. Bn refers to Bottleneck, Llim to lower limit and ULim to upper limit. The mode and the 5%, 25% and 50% Quantiles are provided.

Age (y cal. BP)	**Seg.Sites**	**Haplotypes**	**Pairw. Diff.**	**Nucleo. Div.**	**Hapl. Div.**
Cytochrome B - 282bp					
Modern	1	2	0.467	0.00165	0.467
11.500	6	6	0.689	0.00244	0.516
15.200	8	6	1.868	0.00663	0.737
25.200	12	7	3.736	0.01325	0.901
Control Region - 426bp					
Modern	1	2	0.200	0.00047	0.200
11.500	1	2	0.100	0.00023	0.100
15.200	5	4	0.679	0.00159	0.437
25.200	6	8	2.451	0.00575	0.824
Combined - 708bp					
Modern	2	3	0.667	0.00094	0.600
11.500	7	7	0.789	0.00112	0.584
15.200	13	7	2.547	0.00360	0.774
25.200	18	10	6.187	0.00874	0.945

Table 3.3: **Statistics summarizing the data.** I estimated different statistics for the two genes separately and for the combined region. The number of Segregating Sites directly reflects the genetic variability and therefore the effective population size. Additionally I provide frequency based statistics such as Pairwise Difference. (Seg.Sites...Segregating Sites, Haplotypes...Number of Haplotypes, Pairw. Diff....Pairwise Difference, Nucleo. Div....Nucleotide Diversity, Hapl.Div....Haplotype Diversity).

Model	**-lnL**	**K**	**AIC**	δ	**weight**	**cumWeight**
TrN+I	1163.5972	6	2339.1943	0.0000	0.3322	0.3322
TrN+I+G	1163.3665	7	2340.7329	1.5386	0.1539	0.4861
TIM+I	1163.4777	7	2340.9553	1.7610	0.1377	0.6238
GTR+I	1162.0063	9	2342.0127	2.8184	0.0812	0.7049
TIM+I+G	1163.2518	8	2342.5037	3.3093	0.0635	0.7684

Table 3.4: **Five best Modeltest results for the data used in the Bayesian Skyline Plot.** *-lnL* is the negative log likelihood, *K* the number of parameters, δ the information difference, *weight* is the Akaike weight and *cumWeight* the cumulative Akaike weights.

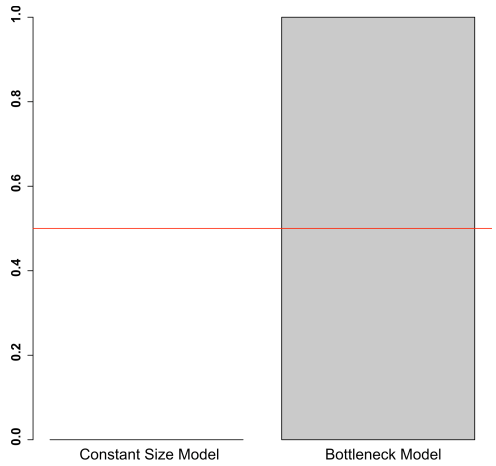

Figure 3.2: **Model-testing using the rejection algorithm.** All 1000 closest parameter estimates rank among the bottleneck model estimates. Thus, clearly favoring the bottleneck model over the constant size model.

A parameter that has a strong effect on the reconstruction of temporal demographic changes is gene flow. In order to address the question whether the data is affected by gene flow or not I used modern phylogeographical data. I constructed parsimony networks comprising haplotypes from different geographical regions. Furthermore, I included 13 samples from another site (less than 1000 y BP), namely Yangana-Pe4. This locality is situated between Pymva Shor and the Yamal peninsula. This data suggests modern gene flow only on a small regional scale. Whereas, all the modern sites revealed only regionally fixed haplotypes (1 for Pechora, 1 for Yamal, 3 for Pymva Shor and 3 for Taimyr) (see figure 4.3), the sampling from Yangana-Pe4 comprises haplotypes from Pymva Shor (10 individuals) , Yamal (2) and one that is unique to YP4 (see figure 4.4).

Chapter 4

Discussion

A population bottleneck is a rapid decline in the census population size having strong effects on the genetic variation within the population and therefore on the effective population size (Ne). When a population genetically decreases the rate of random genetic drift (Wright, 1931) and inbreeding (Saccheri *et al.* , 1999) increase, which then can lead to a reduction in genetic variation. If a widespread population undergoes a strong bottleneck local populations might survive and evolve independent of each other (depending on migration rates). In this case the size reduction causes a decrease in the genetic variation within a population but at the same time lead to an increase of the genetic variation between populations due to separate evolution. On the contrary, it was shown that in many small mammals gene flow increases following to a population decline and therefore reducing variation between populations. Individuals of those species show a negative density-dependent dispersal, meaning that migration rates are higher when the population density is low (e.g. lemmings and voles (Blackburn *et al.* , 1999; Ims & Andreassen, 2000; Andreassen & Ims, 2001; Lin & Batzli, 2001; Ehrich & P.E., 2005)).
During the Late Pleistocene *Dicrostonyx* inhabited wide parts of the Holarctic, especially in northern Eurasia (Database, 1995). Today, the collared lemming is present only in the high arctic with a nearly circumpolar distribution (e.g. (Fedorov *et al.* , 1999)). It is restricted to dry and treeless tundra and only this genus of rodents inhabits the polar desert of the northernmost parts of the Eurasian and Canadian Arctic (Rodgers & Lewis, 1986; Pitelka & Batzli, 1993). Fossil records show that during the glacial advances these Arctic adapted rodents expanded their distribution ranges thousands of kilometers to the south and contracted ranges to the north during warm interglacials (Markova *et al.* , 1995; Group, 1996).
In addition to the paleontological record genetic data of modern *Dicrostonyx* indicates that this genus underwent a very recent bottleneck (e.g. (Fedorov *et al.* , 1999; Fedorov & Goropashnaya, 1999; Fedorov, 1999)). Modern DNA sequences of *Dicrostonyx torquatus* show low nucleotide and haplotype diversity within phyleogeographical groups (Fedorov *et al.* , 1999). Furthermore, haplotype clusters clearly correspond to geographical regions indicating regional bottleneck events (Fedorov *et al.* , 1999).
Based on the paleontological record and modern genetic data I chose *Dicrostonyx torquatus* as a model organism to study the effects of strong recent bottlenecks using temporal DNA sampling. I successfully extracted ancient and modern DNA from a single site (Pymva Shor) in the northern Ural. The temporal network (figure 3.1) showed a drastic decline in genetic variation over time with its minimum at the LGM and its maximum in the modern population. Nearly all distinct haplotypes (more than one substitution in between) were lost between 15.200y and 11.500y cal BP indicating a strong demographic event such as a drastic bottleneck at that time. This loss of

variation could alternatively be explained by selection. Neutrality of the mitochondrial genome is still debated and indeed several studies have confirmed inconsistency of the mtDNA variation and evolution with neutrality (Meiklejohn *et al.* , 2007). However, selection is nevertheless unlikely to be the cause since X and Y chromosome data of modern collared lemmings also exhibit reduced genetic variation (Fedorov unpublished data).
Temporal demographic reconstructions further support the hypothesis of a very severe and recent bottleneck event. Both MCMC based bayesian inference and ABC indicate a severe population decline correlating with the abrupt temperature increase around 14.700y cal. BP. This observation is further supported by small 95% confidence intervals in the ABC method (13.393 to 14.571 generations). However, the BSP reveals some ambiguity in the data, due to the small amount of mutations present in the data. To get a stronger support longer DNA fragments would be needed. Whereas the MCMC method predicted the maximum of the population decline prior to 14.700 cal. BP (approximately at 15.000 cal BP) the ABC method showed the highest probability afterwards (approximately 14.000 cal. BP). This slight discrepancy is not surprising when taking into account the different methods used to infer the temporal demographic history and stochastic process.
On the contrary the similarity of the outcomes support the accuracy of our results, which matches predictions from ecological - and paleontological data. The fossil record of the region surrounding Pymva Shor showed the highest density of *Dicrostonyx torquatus* within the coldest and most arid periods (40.000 - 45.000y and 21.000 - 13.000y uncal. BP) (Golovachov & Smirnov, 2008). In layers indicating warmer climate the proportions of collared lemming remains decrease. This layers were dominated by forest species such as the *Clethrionomys* voles (most prominent remains) (Golovachov & Smirnov, 2008). The density decrease of collared lemming remains along with the temperature increase and the faunal changeover from cold tundra adapted to forest species indicate severe effects of climate change on the biological systems during the Holocene-Pleistocene transition *sensu lato*. However, the fossil record has to be treated carefully since all lemming remains were brought in by predatory birds. This decrease in empirical numbers could also reflect a change in prey-pattern or more likely a bottleneck in the predator itself. Anyhow, paleontological - and genetic data both indicate a severe population decline in *Dicrostonyx torquatus* correlating to the abrupt climate change at the end of the Pleniglacial.
Based on the ecological requirements of the collared lemmings one would predict a second population decline at 11.500y cal. BP (Holocene-Pleistocene transition *sensu stricto*), which was not detectable in the MCMC method. Since all distinct haplotypes went extinct between 15.200 and 11.500 generations I would suggest that most of the power to detect a population decline or information content in the data respectively was lost. This suggestion is further supported by bigger highest posterior densities (HPD) in the Bayesian skyline bot after 11.500 cal. BP.
Additionally I ruled out the possibility of population cycles causing the genetic pattern I observed. Population cycles are more or less regular density fluctuations. They are very well known for small mammals especially voles and lemmings and have been extensively studied (e.g. (Pitelka, 1957; Pearson, 1966; MacLean *et al.* , 1974; Wilson *et al.* , 1999; Ims *et al.* , 2008; Kausrud *et al.* , 2008)). Although the actual cause is not determined evidence suggests climate to be of major importance (Ims *et al.* , 2008; Kausrud *et al.* , 2008). In principle population cycles are repeating bottlenecks and therefore are seen to decrease variation. As already mentioned above recent data indicates that an increase in variation due to higher gene flow might rather be the rule than the exception in small mammals (Ehrich & P.E., 2005). However, in absence of higher dispersal rates the overall effective population size of a fluctuating population can be estimated using the harmonic mean. The harmonic mean weights smaller values to an higher extend by virtue of summarizing

the inverses of the Ne. Thus, reflecting populations with lower Ne and therefore increasing the weight of stochastic processes. I addressed this topic when simulating a population with constant population size over time under different fNe's in the ABC approach.

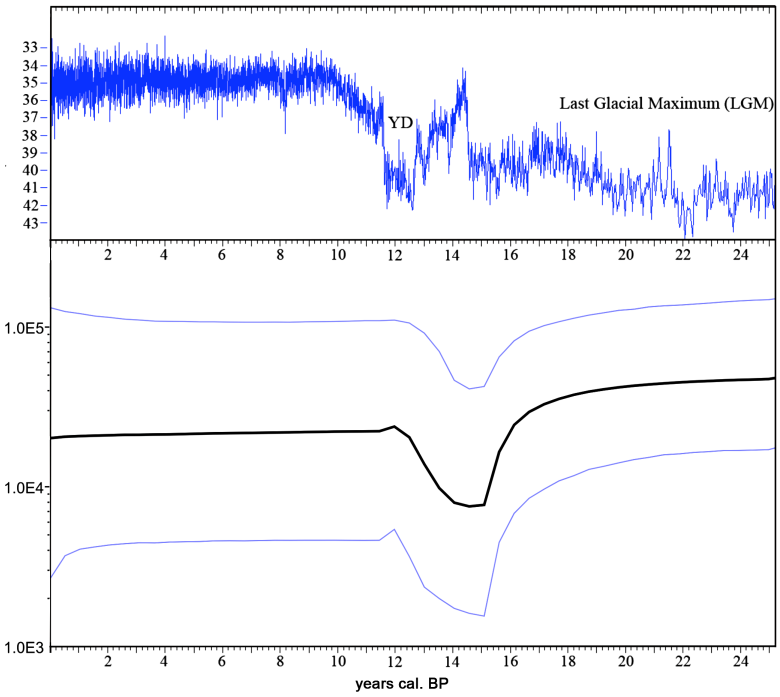

Figure 4.1: **Climate history as derived from the GISP2 Ice-core and demographic history of the collared lemming *Dicrostonyx torquatus* of the last 26.000yrs.** Top panel: oxygen isotope (σ^{18}O) derivations in ‰. Climate graph was calculated using the program CalPal. Bottom panel: Bayesian skyline plot reconstruction of the temporal demographic history of *Dicrostonyx torquatus*. The thick black line is the median estimate and the blue lines indicate the 95% highest posterior density (HPD). After the LGM the living conditions got less favorable and the population started to decline with its maximum at the Bolling-Allerod interstadial (approximately 14.700 to 12.700 y ca. BP). This estimate is congruent with the bottleneck timing using ABC of 14.000 ca. BP (mode). The female population size recovered slightly during the YD cooling event. A population decline at the pleistocene holocene transition is not strongly supported since all distinct haplotypes went extinct before and the data does not include much information. This lack of information further results in an increase of the HPD after 11.500 y ca. BP.

I addressed the question of whether the data is influenced by gene flow and if so to what extend by constructing two additional statistical parsimony networks including just modern and modern and up to 1000y old samples from Yangana-Pe-4, respectively. The modern haplotype network revealed no gene flow between the studied regional haplogroups. However, I found the most prominent haplotypes from PS and Yamal within the sampled data from Yangana-Pe-4. This suggests a contact zone between the two regional populations and therefore gene flow only on a small regional scale. However, one has to be careful, when applying conclusions based on modern findings to ancient data. This is even more true when the studied species underwent drastic demographical changes in the past. Thus, I extended the search for indications of gene flow by studying the temporal network.

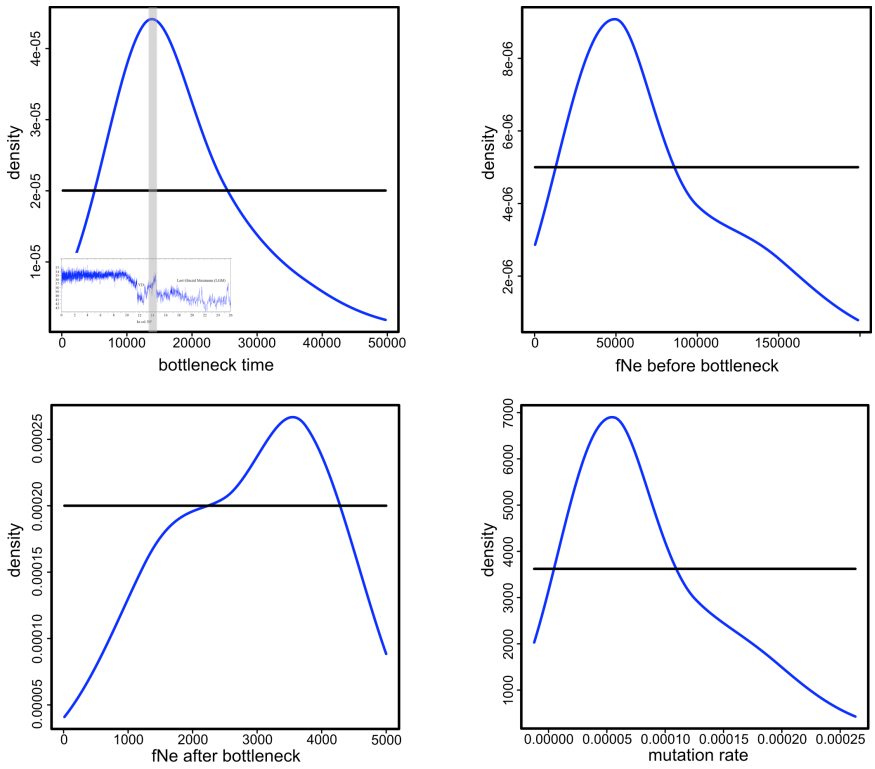

Figure 4.2: **Posterior probability distributions of different model parameters.** Upper left: Density curve of the bottleneck time. Grey bar shows the 5% highest posterior density. The 5% hpd matches to the first major temperature increase at the Pleistocene-Holocene transition (as indicated in the small climate reconstruction). Upper right: Density curve of the female effective population size prior to the bottleneck. Lower left: Density curve of the female effective population size after the bottleneck. Lower right: Density curve of the mutation rate.

On the contrary to the modern data I found evidence for gene flow within the ancient sampling. Many haplotypes (one mutation away from the most common one) accumulated within the sampling from 11.500 y cal. BP. All of those haplotypes were unique. Given the uniqueness of the haplotypes and a suggested mutation rate of 8.5% (in the ABC method) and 9% (in the MCMC method) this finding points toward migration from very close regions into PS, since about 3.000 y are not enough time to accumulate such an amount of unique mutations (given the mutation rate of 8.5% (ABC) and 9 %(BSP) respectively). This pattern would be coherent with the hypothesis that certain small mammals tend to migrate on a higher rate when the population density is low (Blackburn *et al.* , 1999; Ims & Andreassen, 2000; Andreassen & Ims, 2001; Lin & Batzli, 2001; Ehrich & P.E., 2005)). Hadly *et al.* (2004) showed an increase of genetic variation following a severe decrease for *Thomomys talpoides* using the "phylochronological approach". However, we found no evidence for or against gene flow within the older sample points.

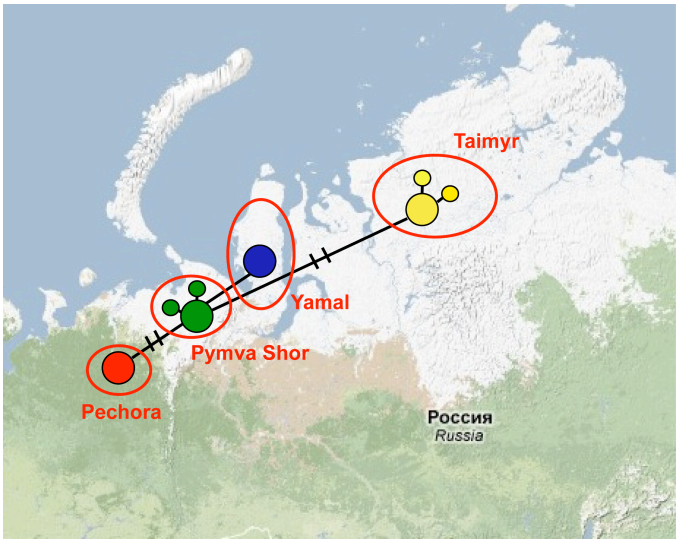

Figure 4.3: **Phylogeographical Reconstruction of the sampled modern haplotypes** Pechora (3 samples), Pymva Shor (10), Yamal (10), NW-Taimyr (14).

I further studied the genetic variation of the two mitochondrial regions (Cytochrome B (CytB) than Control Region (CR)) separately. Surprisingly, I found higher variation within CytB than CR. To show the difference in genetic variation between the two I calculated the average number of segregating sites per nucleotide of the different genes separately. Based on the Neutral theory (Kimura, 1983) one would expect to find more genetic variation with a neutral non-coding gene. This ”upside down” pattern has also been found in other mammals like mammoth (unpublished). However, further studies are needed to clarify its cause.

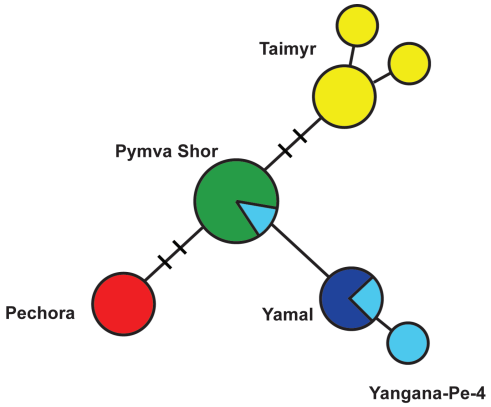

Figure 4.4: **Statistical parsimony network reconstruction of the sampled modern haplotypes and additional samples from Yangana-Pe-4.** red...Pechora (3 samples), green...Pymva Shor (10), blue...Yamal (10), yellow...NW-Taimyr (14) and cyan...Yangana-Pe-4 (13).

Like Chan *et al.* (2006) our results demonstrate that small rodent species can persist over thousands of generation with low genetic diversity. This finding might be interesting for future

conservation of endangered species and to build strategies to prevent extinction or loss of variation (Haig, 1998). The very specialized collared lemming *Dicrostonyx* will most likely face a dark future, when the average temperature further continues to rise. Ecological niche modeling predicts 60% habitats reduction for collared lemmings under a $4\,^{\circ}\mathrm{C}$ increase in average annual temperature on the Canadian mainland (Kerr & Packer, 1998). Given its actual circumpolar distribution not much refugia will be left to save this species from extinction.
In general, ancient DNA studies can provide important informations of possible effects and risks of climate change and therefore help to understand and predict the fate of future populations facing climate change.

Age (y cal. BP)	**Cytochrome B**	**Control Region**
Modern	0.00355	0.00235
11.500	0.02128	0.00235
15.200	0.02837	0.01174
25.200	0.04255	0.01408

Table 4.1: **Average number of Segregating Sites per Nucleotide.** I calculated the average number of Segregating Sites per nucleotide of the different genes separately. Surprisingly I found a higher degree of genetic variation within the coding-gene Cytochrome B and less in the neutral Control Region. Based on the Neutral theory (Kimura, 1983) one would expect to find more genetic variation with a neutral non-coding gene.

Chapter 5

Conclusion

I used temporal DNA sampling to study the demographic changes in the collared lemming (*Dicrostonyx torquatus*) from Pymva Shor in the northern Ural. I successfully extracted, amplified and sequenced DNA from 10 modern and 77 ancient lemming remains. Using makrov chain monte carlo (MCMC) based bayesian inference and approximate bayesian computation (ABC) I was able to show that the abrupt warming event at the Bolling/Allerod interstadial correlated with the timing of a severe population decline in the collared lemming. Given the strong adaptation to dry and cold climate of *Dicrostonyx* this finding suggests a coherence between the population bottleneck and this abrupt warming event. Using additional modern DNA sequences I showed that there are signs for gene flow in my sampled data, which is coherent with the assumption of a higher migration rate following a strong population decline in lemmings. A contact zone between individuals bearing the Yamal haplotype and the major haplotype from Pymva Shor was found in (approximately 1000 year old) samples from Yangana-Pe-4 and thus, suggesting modern gene flow only on a small geographical scale. My study underlines the usefulness of temporal ancient DNA data for a proper understanding of the effects of climate change on mammal populations.

Chapter 6

Zusammenfassung

Ich habe im Rahmen meiner Dimpolarbeit DNA von 10 modernen und 67 subfossilen Lemmingknochen (modern bis 25.200 Jahre kal. BP) aus dem noerdlichen Ural (Russland) extrahiert, amplifiziert und diese mittels 454 FLX "next-generation" Pyrosequencing (ROCHE) sequenziert. Auf Markov Chain Monte Carlo (MCMC) basierende Bayesian Analysen und Approximate Bayesian Computation (ABC) Analysen wurden benutzt um die zeitliche demographische Veraederung der Lemmingpopulation zu rekonstruieren. Der Plot der effektiven Populationsgroesse (Ne) zeigt ein deutliches "Bottleneck" (d.h. einen drastischen Abfall) der zeitlich mit der Klimaerwaermung im Boelling/Alleroed zusammenfaellt. Halsbandlemminge sind spezialisiert auf kaltes und sehr trockenes Klima. Diese hochspezialisierten oekologischen Anforderungen und paleontologische Untersuchungen unterstreichen die Glaubwuerdigkeit der genetischen Resultate weiters. Um den Einfluss von "gene flow" bzw. Migration auf meine Daten zu bestimmen habe ich ein Haplotypennetzwerk mit zusaetzlichen modernen Sequenzen errechnet. Die Ergebnisse dieser Analyse und die Interpretation des temporalen Netzwerke deuten auf eine erhoehte Migrationsrate nach dem Bottleneck hin. Aus publizierten Studien an modernen Lemmingen ist dieses Phenomaen (der Anstieg der Migrationrate wenn die Populationsdichte sinkt) bereits bekannt. Diese Erkenntnisse stehen im Einklang mit den erhaltenen Resultaten. Analysen mit zusaetzlichen subfossilen Lemmingen aus einer weiteren Fundstelle (Yangana-Pe-4) deuten auf das Vorhandensein einer Kontaktzone der regionalen Haplotypengruppen (Yamal und Pymva Shor) hin. Dies zeigt, dass Migration heute nur regional eingeschraenkt stattfindet. Die Ergebnisse dieser Arbeit unterstreichen die Nuetzlichkeit von temporaerem DNA sampling (moderne und subfossile DNA) um die Auswirkungen von Klimaeveraenderungen auf Saeugetierpopulationen zu studieren.

Chapter 7

Acknowledgements

Thanks to: Michael Hofreiter for being the local supervisor of this work at the Max Planck Institute in Leipzig. Doris Nagel for being my official advisor at the University of Vienna. Dorothee Ehrich for the modern samples. Nickolay Smirnov for all the ancient lemming samples and fruitful discussion about the project. Vadim Fedorov for modern sequences and advice. Marc Beaumont and Joao Sollari for the introduction to statistical methods used in Approximate Bayesian Computation. Yvonne Chan and Christian Anderson for help with BayeSSC. Rigo Schultz , Ines Buenger and Christoph Reiterich for support. Michael Knapp and Mathias Stiller for introduction to ancient DNA work and adivce. Nickolay and Galina Smirnov, Elena Kuzmina, Yvonne Chan and Marc Beaumont for friendly hosting. The Max Planck Society for funding.

Bibliography

Abbott, A. 2003. Anthropologists cast doubt on human DNA evidence. *Nature*, **423**(6939), 468.

Agadjanian, AK. 1976. The history of collared lemmings in the Pleistocene. *Beringia in the Cenozoic Era (VL Kontrimavichus, ed.). Russian translation series*, **28**, 379–388.

Alter, S.E., Rynes, E., & Palumbi, S.R. 2007. DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proceedings of the National Academy of Sciences*, **104**(38), 15162.

Anderson, C.N.K., Ramakrishnan, U., Chan, Y.L., & Hadly, E.A. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics*, **21**(8), 1733–1734.

Andreassen, H.P., & Ims, R.A. 2001. Dispersial in patchy vole populations: role of patch configuration, density dependence, and demography. *Ecology*, 2911–2926.

Andreev, A.A., Siegert, C., Klimanov, V.A., Derevyagin, A.Y., Shilova, G.N., & Melles, M. 2002. Late Pleistocene and Holocene vegetation and climate on the Taymyr lowland, northern Siberia. *Quaternary Research*, **57**(1), 138–150.

Andreev, A.A., Tarasov, P.E., Siegert, C., Ebel, T., Klimanov, V.A., Melles, M., Bobrov, A.A., Dereviagin, A.Y., Lubinski, D.J., & Hubberten, H.W. 2003. Late Pleistocene and Holocene vegetation and climate on the northern Taymyr Peninsula, Arctic Russia. *Boreas*, **32**(3), 484–505.

Beaumont, M.A., Zhang, W., & Balding, D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics*, **162**(4), 2025–2035.

Beck, J.W., Richards, D.A., Lawrence, R., Silverman, B.W., Smart, P.L., Donahue, D.J., Hererra-Osterheld, S., Burr, G.S., Calsoyas, L., Timothy, AJ, *et al.* . 2001. Extremely large variations of atmospheric 14C concentration during the last glacial period. *Science*, **292**(5526), 2453–2458.

Benton, M.J., & Twitchett, R.J. 2003. How to kill (almost) all life: the end-Permian extinction event. *Trends in Ecology & Evolution*, **18**(7), 358.

Blackburn, G.S., Wilson, D.J., & Krebs, C.J. 1999. Dispersal of juvenile collared lemmings (Dicrostonyx groenlandicus) in a high density population. *Canadian Journal of Zoology*, 2255–2261.

Box, GEP, Launer, RL, & Wilkinson, GN. 1979. Robustness in statistics. *New York: Academic*, 1–17.

Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Pr
"ufer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., *et al.* . 2007. Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences*, **104**(37), 14616–14621.

Broecker, W.S. 2006. *GEOLOGY: Was the Younger Dryas Triggered by a Flood?*

Broecker, WS, Kennett, JP, Flower, BP, Teller, JT, Trumbore, S., Bonani, G., & Wolfli, W. 1989. Routing of meltwater from the Laurentide ice sheet during the Younger Dryas cold episode. *Nature*, **341**(6240), 318–321.

Brook, E.J., Sowers, T., & Orchardo, J. 1996. Rapid variations in atmospheric methane concentration during the past 110,000 years. *Science*, **273**(5278), 1087.

Brotherton, P., Endicott, P., Sanchez, J.J., Beaumont, M., Barnett, R., Austin, J., & Cooper, A. 2007. Novel high-resolution characterization of ancient DNA reveals C¿ U-type base modification events as the sole cause of post mortem miscoding lesions. *Nucleic Acids Research*, **35**(17), 5717–5729.

Chan, Y.L., Anderson, C.N.K., & Hadly, E.A. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genet*, **2**(4), e59.

Clement, M., Posada, D., & Crandall, K. A. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**(10), 1657–1660.

Cooper, A., & Poinar, H.N. 2000. *Ancient DNA: do it right or not at all.*

Cuffey, K.M., & Clow, G.D. 1997. Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition. *Journal of Geophysical Research-Oceans*, **102**(C12), 26383–26396.

Currie, L.A. 2004. The remarkable metrological history of radiocarbon dating [II]. *Journal of Research-national Institute of Standards and Technology*, **109**, 185–218.

Dansgaard, W., White, JWC, & Johnsen, SJ. 1989. The abrupt termination of the younger dryas climate event. *Nature*, **339**(6225), 532–534.

Dansgaard, W., Johnsen, SJ, Clausen, HB, Dahl-Jensen, D., Gundestrup, NS, Hammer, CU, Hvidberg, CS, Steffensen, JP, Sveinbjörnsdottir, AE, Jouzel, J., *et al.* . 1993. Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature*, **364**(6434), 218–220.

Database, Paleophauna. 1995. Late Pleistocene Distribution and Diversity of Mammals in Northern Eurasia (PALEOPHAUNA Database). *Paleontologia I Evolucia*, 1–143.

de Garidel-Thoron, T., Rosenthal, Y., Bassinot, F., & Beaufort, L. 2005. Stable sea surface temperatures in the western Pacific warm pool over the past 1.75 million years. *Nature*, **433**(7023), 294–298.

Drummond, A.J., & Rambaut, A. *Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.*

Drummond, A.J., & Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**(1), 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**(5), 1185–1192.

Ehrich, D., & P.E., Jorde. 2005. High genetic variability despite high-amplitude population cycles in lemmings. *Journal of Mammology*, 380–385.

Excoffier, L., & Heckel, G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**(10), 745–758.

Excoffier, L., Novembre, J., & Schneider, S. 2000. SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. *The Journal of heredity*, **91**(6), 506.

Fairbanks, R.G., Mortlock, R.A., Chiu, T.C., Cao, L., Kaplan, A., Guilderson, T.P., Fairbanks, T.W., Bloom, A.L., Grootes, P.M., & Nadeau, M.J. 2005. Radiocarbon calibration curve spanning 0 to 50,000 years BP based on paired 230Th/234U/238U and 14C dates on pristine corals. *Quaternary Science Reviews*, **24**(16-17), 1781–1796.

Fedorov, V.B. 1999. Contrasting mitochondrial DNA diversity estimates in two sympatric genera of Arctic lemmings (Dicrostonyx: Lemmus) indicate different responses to Quaternary environmental fluctuations. *Proc. R. Soc. Lond. B*, 621–626.

Fedorov, V.B., & Goropashnaya, A. 1999. The importance of ice ages in diversification of Arctic collared lemmings (Dicrostonyx): evidence from the mitochondrial cytochrome b region. *Hereditas*, 301–307.

Fedorov, V.B., Fredga, K., & G.H., Jarrell. 1999. Mitochondrial DNA variation and the evolutionary history of chromosome races of collared lemmings (Dicrostonyx) in the Eurasian Arctic. *J. Evol. Biol.*, 134–145.

Fisher, RA. 1930. *The genetical theory of natural selection. Clarendon.*

Fu, YX, & Li, WH. 1997. *Estimating the age of the common ancestor of a sample of DNA sequences*.

Gilbert, M.T.P., Bandelt, H.J., Hofreiter, M., & Barnes, I. 2005. Assessing ancient DNA studies. *Trends in Ecology and Evolution*, **20**(10), 541–544.

Golovachov, I.B., & Smirnov, N.G. 2008. The Late Pleistocene and Holocene rodents of the Pre-Urals subarctic. *Quaternary International*.

Griffiths, RC, & Tavare, S. 1994. Ancestral inference in population genetics. *Statistical Science*, 307–319.

Group, Faunmap Working. 1996. Spatial response of mammals to late Quaternary environmental fluctuations. *Science*, 1601–1606.

Hadfield, JD. MasterBayes Tutorial. *MasterBayes: Maximum Likelihood and Markov chain Monte Carlo methods for pedigree reconstruction, analysis and simulation.*

Hadly, EA, Ramakrishnan, U., Chan, YL, Tuinen Mv, O.K.K., *et al.* . 2004. Genetic response to climatic change: insights from ancient DNA and phylochronology. *PLoS Biol*, **2**(10), e290.

Haig, S.M. 1998. Molecular contributions to conservation. *Ecology*, 413–425.

Hastings, WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**(1), 97–109.

Hays, JD, Imbrie, J., & Shackleton, NJ. 1976. Variations in the Earth's orbit: pacemaker of the ice ages. *Science*, **194**(4270), 1121–1132.

Hein, J., Schierup, M.H., & Wiuf, C. 2005. *Gene genealogies, variation and evolution: a primer in coalescent theory*. Oxford University Press, USA.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature*, **405**(6789), 907–913.

Hoek, W.Z. 2001. Vegetation response to the 14.7 and 11.5 ka cal. BP climate transitions: is vegetation lagging climate? *Global and planetary change*, **30**(1-2), 103–115.

Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., & Paeaebo, S. 2001a. Ancient DNA. *Nature Reviews Genetics*, **2**(5), 353–359.

Hofreiter, M., Jaenicke, V., Serre, D., Haeseler, A., & Paabo, S. 2001b. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**(23), 4793–4799.

Hofreiter, M., Capelli, C., Krings, M., Waits, L., Conard, N., Munzel, S., Rabeder, G., Nagel, D., Paunovic, M., Jambresic, G., *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**(8), 1244–1250.

Houghton, JT, Ding, Y., Griggs, DJ, Noguer, M., van der Linden, PJ, Dai, X., & Maskell, K. 2001. *Climate Change 2001: The Scientific Basis*. Cambridge University Press.

Hughen, K., Lehman, S., Southon, J., Overpeck, J., Marchal, O., Herring, C., & Turnbull, J. 2004. 14C activity and global carbon cycle changes over the past 50,000 years. *Science*, **303**(5655), 202–207.

Imbrie, J., Boyle, EA, Clemens, SC, Duffy, A., Howard, WR, Kukla, G., Kutzbach, J., Martinson, DG, McIntyre, A., Mix, AC, *et al.* . 1992. On the structure and origin of major glaciation cycles 1. Linear responses to Milankovitch forcing. *Paleoceanography*, **7**(6).

Ims, R.A., & Andreassen, H.P. 2000. Spatial synchronization of vole population dynamics by predatory birds. *Nature*, 194–196.

Ims, R.A., Henden, J-A., & S.T., Killengreen. 2008. Collapsing population cycles. *Trends in Ecology and Evolution*, 79–86.

Kass, R.E., & Raftery, A.E. 1995. Bayes factors. *Journal of the American Statistical Association*, **90**(430).

Kausrud, K.L., Mysterud, A., Steen, H., Vik, J.O., Ostbye, E., Cazelles, B., Framstad, E., Eikeset, A.M., Mysterud, I., Solhoy, T., & Stenseth, N.C. 2008. Linking climate change to lemming cycles. *Nature*, 93–98.

Kennett, J.P., & Shackleton, NJ. 1975. Laurentide ice sheet meltwater recorded in Gulf of Mexico deep-sea cores. *Science*, **188**(4184), 147–150.

Kerr, J., & Packer, L. 1998. The impact of climate change on mammalian diversity in Canada. *Environmental Monitoring and Assessment*, 263–270.

Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press.

Kingman, J.F.C. 1982. The coalescent. Stoch. *Pages 235–248 of: Proc. Appl*, vol. 13.

Kitagawa, H., & Van der Plicht, J. 1998. Atmospheric radiocarbon calibration to 45,000 yr BP: Late glacial fluctuations and cosmogenic isotope production. *Science*, **279**(5354), 1187.

Klimanov, VA. 1997. Late glacial climate in northern Eurasia: the last climatic cycle. *Quaternary International*, **41**, 141–152.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Paabo, S., Villablanca, F.X., & Wilson, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences*, **86**(16), 6196–6200.

Kowalski, K. 1995. Lemmings (Mammalia, Rodentia) as indicators of temperature and humidity in the European Quaternary. *Acta Zool. Cracov*, **38**, 85–94.

Kuhner, M.K. 2009. Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution*, **24**(2), 86–93.

Lehman, SJ, & Keigwin, LD. 1992. Sudden changes in North Atlantic circulation during the last deglaciation. *Nature*, **356**(6372), 757–762.

Lemmon, A.R., & Moriarty, E.C. 2004. The importance of proper model assumption in Bayesian phylogenetics. *Systematic Biology*, **53**(2), 265–277.

Leonard, JA., Shanks, O., Hofreiter, M., Kreuz, E., Hodges, L., Ream, W., Wayne, RK., & Fleischer, RC. 2007. Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archeological Science*, **34**(9), 1361–1366.

Libby, WF, Anderson, EC, & Arnold, JR. 1949. Age determination by radiocarbon content: worldwide assay of natural radiocarbon. *Science*, **109**(2827), 227–228.

Lin, Y.T.K., & Batzli, G.O. 2001. The influence of habitat quality on dispersal demography, and population dynamics. *Ecol. Monogr.*, 245–275.

Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature*, **362**(6422), 709–715.

Lockhart, P.J., Larkum, A.W.D., Steel, M.A., Waddell, P.J., & Penny, D. 1996. *Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis.*

MacLean, S.F. Jr., Fitzgerald, B.M., & Pitelka, F.A. 1974. Population cycles in arctic lemmings: winter reproduction and predation by weasels. *Arctic and Alpine Research*, 1–12.

Marjoram, P., & Tavaré, S. 2006. Modern computational approaches for analysing molecular genetic variation data. *Nature Reviews Genetics*, **7**(10), 759–770.

Markova, A.K., Smirnov, N.G., Kozharinov, A.V., Kazantseva, N.E., Simakova, A.N., & Kitaev, L.M. 1995. Late Pleistocene distribution and diversity of mammals in northern Eurasia. *Paleontologia I Evolucia*, 5–143.

Matson, C.W., & Baker, R.J. 2001. DNA sequence variation in the mitochondrial control region of red-backed voles (Clethrionomys). *Molecular Biology and Evolution*, **18**(8), 1494–1501.

Meiklejohn, C.D., Montooth, K.L., & Rand, D.M. 2007. Positive and negative selection on the mitochondrial genome. *Trends in Genetics*, 259–263.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., *et al.* . 1953. Equation of state calculations by fast computing machines. *The journal of chemical physics*, **21**(6), 1087–1092.

Meyer, M., Briggs, A. W., Maricic, T., Hober, B., Hoffner, B., Krause, J., Weihmann, A., & Paabo, S.and Hofreiter, M. 2008a. From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Res*, **36**(1), e5.

Meyer, M., Stenzel, U., & Hofreiter, M. 2008b. Parallel tagged sequencing on the 454 platform. *Nature Protocols*, **3**(2), 267–278.

Moran, PAP. 1958. A general theory of the distribution of gene frequencies. I. Overlapping generations. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 102–112.

Otto, S.P., Day, T., & Day, T. 2007. *A biologist's guide to mathematical modeling in ecology and evolution*. Princeton University Press.

Overpeck, J.T., & Webb, R.S. 1992. Mapping eastern North American vegetation change of the past 18 ka; no-analogs and the future. *Geology*, **20**(12), 1071–1074.

Paillard, D. 1998. The timing of Pleistocene glaciations from a simple multiple-state climate model. *Nature*, **391**(6665), 378–381.

Parmesan, C., & Yohe, G. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**(6918), 37–42.

Patz, J.A., Campbell-Lendrum, D., Holloway, T., & Foley, J.A. 2005. Impact of regional climate change on human health. *Nature*, **438**(7066), 310–317.

Pearson, O.P. 1966. The prey of carnivores during one cycle of mouse abundance. *J. Anim. Ecol.*, 217–233.

Pitelka, F.A. 1957. Some aspects of population structure in the short-term cycle of the brown lemming in northern Alaska. *Springs Harb. Symp. Quant. Biol.*, 237–251.

Pitelka, F.A., & Batzli, G.O. 1993. Distribution, abundance and habitat use by lemmings on the north slope of Alaska. *In The Biology of Lemmings (eds. N.C. Stenseth and R. A. Ims)*, 214–236.

Posada, D., & Buckley, T.R. 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology*, **53**(5), 793–808.

Posada, D., & Crandall, KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**(9), 817–818.

Pritchard, JK, Seielstad, MT, Perez-Lezaun, A., & Feldman, MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, **16**(12), 1791–1798.

Ramakrishnan, U., & Hadly, E.A. 2009. Using phylochronology to reveal cryptic population histories: review and synthesis of 29 ancient DNA studies. *Molecular Ecology*, **18**(7), 1310–1330.

Reimer, PJ, Baillie, MGL, Bard, E., Bayliss, A., Beck, JW, Bertrand, CJH, Blackwell, PG, Buck, CE, Burr, GS, Cutler, KB, *et al.* . 2004. IntCal04 terrestrial radiocarbon age calibration, 0-26 Cal Kyr BP. *Radiocarbon*, **46**(3), 1029–1058.

Rodgers, A.R., & Lewis, M.C. 1986. *Diet selection in Arctic lemmings (Lemmus sibiricus and Dicrostonyx groenlandicus): demography, home range, and habitat use.*

Rodrigo, A., & Felsenstein, J. 1999. Coalescent approaches to HIV population genetics. *The evolution of HIV*, 233–272.

Roempler, H., Dear, PH., Krause, J., Meyer, M., Rohland, N., Schoeneberg, T., Spriggs, H., Stiller, M., & Hofreiter, M. 2006. Multiplex amplification of ancient DNA. *Nature Protocols*, **1**(2), 720–728.

Rohland, N., & Hofreiter, M. 2007. Ancient DNA extraction from bones and teeth. *Nature Protocols*, **2**(7), 1756–1762.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C., & Pounds, J.A. 2003. Fingerprints of global warming on wild animals and plants. *Nature*, **421**(6918), 57–60.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., & Rozas, R. 2003. *DnaSP, DNA polymorphism analyses by the coalescent and other methods*.

Rozen, S., & Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*, **132**, 365–86.

Ruzzante, D.E., Walde, S.J., Gosse, J.C., Cussac, V.E., Habit, E., Zemlak, T.S., & Adams, E.D.M. 2008. Climate control on ancestral population dynamics: insight from Patagonian fish phylogeography. *Molecular Ecology*, **17**(9), 2234–2244.

Saccheri, I.J., Wilson, I.J., Bruford, M.W., & Brakefield, P.M. 1999. Inbreeding of bottlenecked butterfly populations: estimation using the likelihodd of changes in marker allele frequencies. *Genetics*, 1053–1063.

Schmidt, H.A., Strimmer, K., Vingron, M., & von Haeseler, A. 2002. *TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing*.

Spahni, R., Chappellaz, J., Stocker, T.F., Loulergue, L., Hausammann, G., Kawamura, K., Flueckiger, J., Schwander, J., Raynaud, D., Masson-Delmotte, V., *et al.* . 2005. Atmospheric methane and nitrous oxide of the late Pleistocene from Antarctic ice cores. *Science*, **310**(5752), 1317–1321.

Stacy, J.E., Jorde, P.E., Steen, H., Ims, R.A., Purvis, A., & Jakobsen, K.S. 1997. Lack of concordance between mtDNA gene flow and population density fluctuations in the bank vole. *Molecular Ecology*, **6**(8), 751–759.

Stiller, M., Green, RE, Ronan, M., Simons, JF, Du, L., He, W., Egholm, M., Rothberg, JM, Keates, SG, Ovodov, ND, *et al.* . 2006. Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences*, **103**(37), 13578–13584.

Stiller, M., Knapp, M., Stenzel, U., Hofreiter, M., & Meyer, M. 2009. Direct multiplex sequencing (DMPS) a novel method for targeted high-throughput sequencing of ancient and highly degraded DNA. *Genome Research*.

Swofford, D.L., & Sullivan, J. 2003. Phylogeny inference based on parsimony and other methods using PAUP*. *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*, 160.

Tavare, S., Balding, DJ, Griffiths, RC, & Donnelly, P. 1997. Inferring coalescence times from DNA sequence data. *Genetics*, **145**(2), 505–518.

Thalmann, O., Fischer, A., Lankester, F., Paabo, S., & Vigilant, L. 2007. The complex evolutionary history of gorillas: insights from genomic data. *Molecular biology and evolution*, **24**(1), 146.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., De Siqueira, M.F., Grainger, A., Hannah, L., *et al.* . 2004. Extinction risk from climate change. *Nature*, **427**, 145–148.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., & Prentice, I.C. 2005. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Sciences*, **102**(23), 8245–8250.

Tomilin, AG. 1957. Mammals of the USSR and adjacent countries. *Vol. IX. Cetacea. Moscow, Soviet Union (English translation, 1967, Israel Program for Scientific Translations, Jerusalem, Israel)*.

Tzedakis, PC, Andrieu, V., De Beaulieu, J.L., Crowhurst, S., Follieri, M., Hooghiemstra, H., Magri, D., Reille, M., Sadori, L., Shackleton, NJ, *et al.* . 1997. Comparison of terrestrial and marine records of changing climate of the last 500,000 years. *Earth and Planetary Science Letters*, **150**(1-2), 171–176.

Van der Plicht, J., Beck, JW, Bard, E., Baillie, MGL, Blackwell, PG, Buck, CE, Friedrich, M., Guilderson, TP, Hughen, KA, Kromer, B., *et al.* . 2004. NotCal04Comparison/Calibration 14 C records 26–50 cal kyr BP. *Radiocarbon*, **46**(3), 1225–1238.

Velichko, AA, Andreev, AA, & Klimanov, VA. 1997. Climate and vegetation dynamics in the tundra and forest zone during the Late Glacial and Holocene. *Quaternary international*, **41**, 71–96.

Voelker, AHL, Sarnthein, M., Grootes, PM, Erlenkeuser, H., Laj, C., Mazaud, A., Nadeau, M.J., & Schleicher, M. 1998. Correlation of marine 14 C ages from the nordic seas with the GISP 2 isotope record: Implications for 14 C calibration beyond 25 ka BP. *Radiocarbon*, **40**(1), 517–534.

Wakeley, J. 2008. *Coalescent theory - An introduction*.

Webb, R.S., Rind, D.H., Lehman, S.J., Healy, R.J., & Sigman, D. 1997. Influence of ocean heat transport on the climate of the Last Glacial Maximum. *Nature*, **385**(6618), 695–699.

Weiss, G., & von Haeseler, A. 1998. Inference of population history using a likelihood approach. *Genetics*, **149**(3), 1539–1546.

Wilson, D.J., Krebs, C.J., & Sinclair, A.R.E. 1999. Limitation of collared lemming populations during a population cycle. *OIKOS*, 382–398.

Wright, S. 1931. Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Zazhigin, VS. 1976. Early evolutionary stages of collared lemmings (Dicrostonychini, Microtinae, Rodentia) as characteristic representatives of Beringian Subarctic fauna. *Beringia in the Cenozoic Era(ed. V L. Kontrimavichus)*, 280–288.

Stefan Prost

Born: December 5th, 1982.
Citizenship: Austria

Address:
Kitaibelgasse 25
7210 Mattersburg
Austria

Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany
Deutscher Platz 6
04103 Leipzig
Germany

Institute of Microbiology and Genetics, Vienna Biocenter, Vienna Austria
Dr. Bohrgasse 9/4
01030 Vienna
Austria

Email: stefan_prost@eva.mpg.de
URL: http://www.eva.mpg.de/molecular-ecology/staff/prost/index.html

Education

- Magister rer. nat. *expected September 2009*.
- Magister Student at the University of Vienna, Vienna (Austria) for Biology (Microbiology and Genetics) and Earth Science *since Oct.2002*

  Diploma Thesis: *"DNA Analysis on late glacial mammal remains. - Reconstructing temporal demographic changes in the siberian collared lemming (Dicrostonyx torquatus) using approximate bayesian computation (ABC) and markov chain monte carlo (MCMC) based bayesian Inference."*
- Student Assistant at the Max-Planck Institute for Evolutionary Anthropology, Leipzig, Germany in the Department for Evolutionary Genetics, Junior Research Group Molecular Ecology *Feb.2007 - Feb.2009*
  Supervisor: Michael Hofreiter (hofreite@eva.mpg.de)
- General qualification for University entrance ('Reifepruefung') *June 2001*

Research Trips

- 5-days - Sampling - Institute of Plant and Animal Ecology (IPAE), Ekaterinburg (Russia) *Sept.2008*
- 1-month - Approximate Bayesian Computation- University of Hawai'i Manoa, Honolulu (USA) *May 2008*
- 10-days- Approximate Bayesian Computation- Marc A. Beaumont lab, University of Reading, Reading (UK) *April 2009*

Upcoming Publications

- **Prost, S.**, Smirnov, N., Fedorov, V., Sommer, R., Knapp, M., Stiller, M., Nagel, D. and Hofreiter, M. "Influence of climate warming on arctic mammals? New insights from ancient DNA studies of the arctic lemming Dicrostonyx torquatus." ***in prep***
- **Prost, S.**, Knapp, M., Stiller, M., Flemming, J. and Hofreiter, M. "Phylogenetic analyis of the russian don hare *Lepus tanaiticus* from late glacial deposits of the Ural mountain" ***in prep.***
- **Prost, S.**, Klietmann, J., Knapp, M., Stiller, M., Vrieling, K., Nagel, D., Rabeder, G, Hofreiter, M. and van Kolfschoten "Size doesn't matter. Notes on ecology and phylogenetic relationships of late glacial soricine shrews" ***in prep.***

Talks

Invited Talks

- *"Inference of Temporal Demographic Changes using Ancient DNA"* , University of Reading, Reading (UK) *April 6th 2009*
- *"Reconstructing Past Population Dynamics Using Temporal Ancient DNA Data"*, Natural History Museum, Berlin (Germany) *June 26th 2008*
- *"Ancient DNA and its use in population genetics and conservation biology"*, University of Hawai'i Manoa, Honolulu (USA) *May 2nd 2008*
- *"Ancient DNA: Tracing Past Population Dynamics"*, Universitat Pompeu Fabra, Barcelona (Spain) *Feb. 3rd 2008*

- *"Learning from the past - Inference of temporal demographic changes can help us to understand and predict possible effects of climate change"*, Faculty Centre of Biodiversity, University of Vienna, Vienna (Austria) *August 31st 2009*

Talks

- *"Tracing Past Population Dynamics Using a Coalescent Based Method Including Serial DNA Data"*, Max-Planck-Institute for Evolutionary Anthropology, Leipzig (Germany) *Feb. 21th 2008*
- *"Ancient DNA - Decoding the key of life."*, University of Vienna, Vienna (Austria) - in the frame work of the lecture *"Molekulare Mechanismen der Evolution - Was Darwin noch nicht wissen konnte" Dec. 1st 2008*

Conference Attendences

- Spymposium on Biodiversity (Pacific Bioscience Research Center), Hawaii (USA) *May 9th 2008*

Scholarships

- KWA (Short Research Stays Abroad), University of Vienna, Austria; for the maximal duration of 3 month

Languages

- German (native)
- English (fluent) [TOEFL iBT 108 points out of 120]
- French (very basics)