The Preservation and Postglacial History of Ice Age Holarctic Beetles, as Inferred from Museum and Ancient DNA

Peter David Heintzman, MSc

Supervised by Dr. Ian Barnes and Prof. Scott A. Elias.

A thesis submitted to Royal Holloway College, University of London in accordance with the requirements of the degree of Doctor of Philosophy in the School of Biological Sciences, Faculty of Science.

December 2012

Declaration of Authorship

I, Peter David Heintzman, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Peter David Heintzman

Signed: _____

Date: _____

Ian Barnes

Signed: _____

Date: _____

Abstract

Degraded DNA can be recovered from specimens that are preserved in museums and the natural environment. Data generated from such DNA have provided valuable evidence for the assessment of a suite of biologically important questions. However, research of this nature is limited for invertebrate taxa, despite their diversity and ecological necessity. Using DNA data from dry-stored museum and permafrost-preserved ancient specimens, this thesis greatly extends the study of degraded DNA from invertebrates. The thesis focuses on two arctic ground beetle species (*Amara alpina*, *Pterostichus brevicornis*), which are abundant in museum collections and permafrost deposits.

A lack of data that characterises the preservation and potential of degraded beetle DNA, and thereby assessment of future possibilities for this emerging field, provided the impetus for the first three results chapters. Using two different sequencing approaches, the preservation of DNA in museum and ancient specimens was investigated. In addition, the taxonomic utility of DNA extracted from these specimens was assessed. These chapters demonstrate that DNA could be routinely recovered from museum specimens. DNA from ancient specimens could be recovered from *A. alpina* but not *P. brevicornis*. In most cases therefore, degraded DNA from these beetles could be used to address further questions.

The final two results chapters focus on the response of the two study species to a past period of climatically driven change, using DNA data from museum and ancient specimens. In these chapters, the mode of postglacial colonisation of Canada at the end of the last ice age was investigated. It was found that existing models of this process were broadly, but not wholly, correct. This may have implications for models of how beetles will respond to future climatic change.

Although some challenges lie ahead, this thesis demonstrates the potential for museum and ancient permafrost-preserved beetle specimens in future, DNA-based, large-scale investigations.

Acknowledgements

First and foremost, I thank my supervisors, Ian Barnes and Scott Elias, for their advice, guidance, and inspiration throughout this project. Ian has provided hugely valuable insight and encouragement throughout. Additionally, Selina Brace, Jessica Thomas, and Mark Brown provided valuable discussion throughout project development. Selina Brace, Meirav Meiri, and Amber Teacher taught me the basics of working with degraded DNA. Roland Preece provided technical assistance, and Edwin van Leeuwen provided assistance with, and discussion about, statistics. Rob Prouse and Lyn Wright ensured the timely delivery and dispatch of samples and consumables. Modern and ancient specimens were kindly donated by Svetlana Kuzmina, Philip Thomsen, and Scott Elias. Moreover, Yves Bousquet and Anthony Davies (both CNC), as well as Johannes Bergsten (NRM), provided access to, and permission to sample, the museum specimens. I thank the Exeter Sequencing Service for conducting the next-generation sequencing. Jessica Thomas, Aurelien Ginolhac, Matthias Meyer, and Martin Kircher provided helpful advice on NGS and associated data analyses. I also thank the latter two, as well as Tom Gilbert and Susana Gonçalves, for donating literature. Ian Barnes, Scott Elias, Selina Brace, and Jessica Thomas helpfully commented on earlier versions of this thesis, but I take sole responsibility for any remaining errors and omissions.

With regard to funding, a RHUL Reid Scholarship covered my living expenses. Furthermore, the trip to the CNC was made possible through a RHUL Research Strategy Fund grant, and the trip to the NRM, including molecular analyses, was funded by a SYNTHESYS grant (SE-TAF-1185; SYNTHESYS receives funding from the EC-RIA under FP7). Additional expenses for the latter trip were kindly covered by Love Dalén, whom I also thank for being my host. Ian Barnes financed the majority of the consumable costs associated with this research.

I wish to thank numerous other people not directly involved in this project, but who have made the last three years at RHUL memorable and enjoyable: the EEB, SCR, Railway, and Happy Man social groups. Lastly, I thank Chloe Marquart who has been very supportive and endured countless hours listening to me talk about this research.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

General Abbreviations

28S: 28S ribosomal DNA (*nuclear DNA marker*)

aDNA: ancient DNA

BC: barcode

bp(s): base pair(s)

BSA: bovine serum albumin

Cal. yrs BP: calibrated/calendar years before the present

COI: cytochrome *c* oxidase subunit I (*mitochondrial DNA marker*)

COII: cytochrome *c* oxidase subunit II (*mitochondrial DNA marker*)

d.f.: degrees of freedom

DNA: deoxyribonucleic acid

Ht: haplotype

Hg: haplogroup

ITS1: internal transcribed spacer 1 (*nuclear DNA marker*)

ka: thousand years ago

Ma: million years ago

mtDNA: mitochondrial DNA

NGS: next-generation sequencing

nuDNA: nuclear DNA

NUMT: nuclear-inserted mitochondrial DNA (*pseudogene*)

PCR: polymerase chain reaction

RC yrs BP: radiocarbon years before the present

USA: United States of America

yr(s): year(s)

Institutional Abbreviations

CNC: Canadian National Collection of Insects, Arachnids and Nematodes (Ottawa, Canada)

NRM: Swedish Museum of Natural History (Stockholm, Sweden)

RHUL: Royal Holloway University of London (Egham, UK)

Chapter 1. Introduction

1.1. The Importance of Beetles

Insects are the most diverse class of organisms on Earth and are of importance economically, scientifically, and epidemiologically, as well as being imperative to the healthy functioning of terrestrial ecosystems (Gullan & Cranston 2010). The beetles (Insecta: Coleoptera) are the most speciose insect order (40%; >350,000 species) and are found in nearly all terrestrial ecosystems, fulfilling a great variety of niches (Grove & Stork 2000; New 2007). Beetles are therefore the focus of countless studies investigating a suite of biological questions, such as those related to evolution, ecosystem function, and conservation. Many of these studies require a genetic approach, as this is a proven way to robustly assess key indicators, such as relatedness, population structure, and the interactions between and within species, which are used to assess these questions. To conduct genetic studies, specimens need to be collected. However, some regions, such as the arctic - an area that is particularly vulnerable to predicted future climatic and associated ecosystem changes (Post *et al*. 2009), are difficult, dangerous, and expensive to sample due to their inhospitable nature.

1.2. The Genetic Potential of Museum Collections

Fortunately, due to more than two centuries of collection effort, hundreds of millions of insect specimens have been deposited in museum collections (Suarez & Tsutsui 2004), thereby removing the need to sample directly from difficult regions in many cases (Schaefer *et al*. 2009). In addition to being an easily accessible source of beetle specimens, museum collections can also add a temporal perspective, on a decadal to centennial time scale (Rowe *et al*. 2011), to genetic studies. Museum specimen based genetic studies of other insects have assessed issues such as the preadaptation to recent pesticide-induced selection pressures (Hartley *et al*. 2006) and extreme genetic drift (Harper *et al*. 2006). Although some researchers have utilised DNA from dry-stored museum specimens of beetles (Castalanelli *et al*. 2010; Gibson *et al*. 2012; Gilbert *et al*. 2007; Goldstein & Desalle 2003; Thomsen *et al*. 2009), these studies are

generally small in scale and/or limited in scope. DNA from museum specimens is highly degraded (Wandeler *et al*. 2007), especially in old (>50yr) specimens, which has no doubt hampered or dissuaded researchers from utilising them in other past studies. Data are now becoming available that begin to characterise the DNA present in old museum specimens of other insect orders (Andersen & Mills 2012; Strange *et al*. 2009; Tagliavia *et al*. 2011; Ugelvig *et al*. 2011; van Houdt *et al*. 2010; Watts *et al*. 2007). However, these studies vary in their quality and large-scale, standardised data are usually limited. Furthermore, there are no standardised data that characterise the DNA present in old museum beetle specimens. A large-scale study that robustly assesses both the proportion and preservation of endogenous DNA in these remains is therefore required. Data from such a study would provide either reassurance or a warning to researchers keen to exploit these potentially vast untapped genetic resources, as well as museum curators who may be hesitant about allowing their specimens to being used for potentially destructive genetic analyses (Mandrioli 2008).

1.3. Insects and aDNA Research

In addition to museum specimens, ancient specimens - defined here as >100yr old individuals that have been preserved in the natural environment and are therefore likely to have degraded aDNA, are also potentially large untapped genetic resources with the capacity to extend the temporal aspect of insect DNA studies to millennial time scales. With this extended temporal range, it would be possible to investigate novel questions related to historical events, such as evolutionary changes and the response of species to past episodes of climatic and environmental change.

The first attempts to retrieve aDNA from insects were conducted on specimens preserved in 25 to 135 million year old (Ma) ambers (Cano *et al.* 1993; Desalle *et al.* 1992). However, these studies were met with intense scepticism, based on the stability and degradation rate of *post mortem* DNA (Lindahl 1993). Studies that attempted to replicate these findings either failed or identified the DNA source as a contaminant (Austin *et al.* 1997a; Gutierrez & Marin 1998; Walden & Robertson 1997). These amber-based studies have been dismissed as an example of the contamination and reproducibility issues that are inherent to aDNA studies (Austin *et al.* 1997b; Hebsgaard *et al.* 2005). To combat these issues, handling and analysis protocols are

now a component of the design of any significant aDNA study (Cooper & Poinar 2000; Gilbert *et al.* 2005; Hofreiter *et al.* 2001).

The maximum age of DNA-bearing fossil material has also been the subject of significant discussion. However, there is some agreement that, under ideal conditions (very cold and consequently dry - see below for examples), the theoretical age limit for the recovery of aDNA is ~0.1 to 1Ma (Lindahl 1993; Poinar *et al.* 1996; Smith *et al.* 2003; Willerslev & Cooper 2005).

The first insect aDNA study, using specimens of suitable age (<1Ma), was on glacially-preserved (400yr) grasshoppers (Chapco & Litzenberger 2004). However, the authors did not present contamination reduction protocols or other support for the authenticity of the recovered DNA. The first researcher to recover insect aDNA of suitable age, and follow strict aDNA protocols, was Reiss (2006) who explored the retrieval of beetle aDNA from 20 thousand year old (ka) packrat middens; a deposit-type characterised by exceptionally low moisture content. Willerslev and colleagues (2007) then recovered the oldest accepted insect aDNA from the basal section of a Greenland ice core, which dates to between 450 and 800ka. A further two studies, based on the extraction of beetle aDNA from exoskeletal sclerites, were published in 2009. King and colleagues (2009) isolated weevil aDNA from Roman and Medieval non-frozen sediments, whereas Thomsen and colleagues (2009) recovered aDNA from 26 to 10ka permafrost deposits. In the latter study, insect aDNA was also retrieved directly from non-frozen cave sediments that were between 1800 and 3280yrs old (Thomsen *et al.* 2009).

These proof of concept studies have demonstrated the presence of insect aDNA from a variety of environments and ages, with permafrost-preserved specimens considered to have great potential (Reiss 2006). Therefore, a next step in insect aDNA research is to characterise the DNA present in permafrost-preserved remains. This includes an assessment of DNA preservation and the taxonomic composition of the DNA extracted from these remains. Together, these data would assess the potential of permafrost-preserved insect remains as a genetic resource for future studies.

1.4. Recovering Museum and Ancient DNA

The repercussions from the aforementioned amber-based studies demonstrate that the recovery of reliable DNA data from museum and ancient specimens is a non-trivial endeavour. Endogenous aDNA (which originates from the organism) is degraded by fragmentation, cross-linking, and other damage-inducing processes that are influenced by environmental factors such as temperature, moisture, and exposure to ultraviolet (UV) radiation and free oxygen (Lindahl 1993; Paabo *et al.* 2004). Increasing the level of these environmental factors will degrade DNA, and can prevent aDNA from being recoverable (Lindahl 1993; Paabo *et al.* 2004). Hydrolytic reactions, which occur in the presence of free water, can cause depurination (removal of purine bases). This can increase the chance of DNA strand breaks (producing fragmentation). Additionally, hydrolysis can cause miscoding lesions, whereby a base is chemically modified into an analogue of a different base, which can introduce errors into any recovered aDNA sequences. Oxidation can produce blocking lesions and UV radiation can cause intra- and inter-strand crosslinks in DNA; both of these processes can prevent aDNA amplification (Hofreiter *et al.* 2001; Lindahl 1993; Paabo *et al.* 2004; Willerslev & Cooper 2005). Altogether, these processes severely reduce the concentration of endogenous aDNA, which can lead to the erroneous recovery of exogenous, or contaminated, sequences (Lindahl 1993; Paabo *et al.* 2004). Strict protocols require that aDNA handling, including DNA extraction, be performed in sterile conditions, physically isolated from potential contaminant sources. Additionally, any recovered genetic data should be replicable, make phylogenetic sense, and display appropriate biomolecular behaviour such as only recovering short DNA fragments (almost invariably less than 1000bp, and generally less than 200bp), due to fragmentation (Cooper & Poinar 2000; Gilbert *et al.* 2005; Hofreiter *et al.* 2001).

Post mortem damage will reduce the amount of DNA that can be recovered and characterised from a sample. Animal cells have two types of DNA present: mitochondrial (mt) and nuclear (nu) DNA. MtDNA, located within the mitochondria, is maternally inherited with up to several thousand copies per cell (Moraes 2001). NuDNA, found within the nucleus, is paternally and maternally inherited with two copies per cell. However, some genes within the nuclear genome are multi-copy (e.g. 28S, ITS1, 18S), thereby increasing their copy number per cell. The probability of

preservation for any given DNA locus may increase with copy number (Hofreiter *et al.* 2001). Therefore, mtDNA and multi-copy nuDNA would have a greater chance of preservation than single-copy nuDNA.

At present, there are two major methodological approaches for the retrieval of aDNA sequences from extracted DNA: Sanger sequencing, and next-generation sequencing. The former involves the targeted amplification of short (<1000bp) DNA loci (fragments) by PCR, followed by Sanger sequencing of the resultant PCR products to obtain a consensus of the amplified DNA. Problems can arise if the amplified DNA fragments are damaged or contaminated. However, multiple short overlapping fragments can be combined to aid detection of these potential error sources, as well as increasing the amount of usable genetic data (Krings *et al.* 1997). Sanger sequencing data are appropriate for detecting the presence of aDNA, as well as assessing the preservation (such as maximum recoverable fragment length) of aDNA.

Next-generation sequencing (hereafter NGS) can be employed to sequence millions of extracted DNA fragments in parallel (the shotgun approach), thereby massively increasing the amount of DNA sequence data retrieved compared to Sanger sequencing (Knapp & Hofreiter 2010). Although various NGS technologies are available (Shokralla *et al.* 2012), the Illumina sequencing-by-synthesis system has been a popular method for many recent aDNA studies [e.g. Allentoft *et al*. (2012), Meyer *et al.* (2012), Rasmussen *et al*. (2011), Sawyer *et al*. (2012)], due in part to a low error rate and superior throughput (Knapp & Hofreiter 2010; Shokralla *et al.* 2012). The NGS procedure has three main steps: DNA library construction, sequencing, and data analysis. Library construction prepares the DNA for sequencing through the ligation of adapters. During this step, a barcode, in the form of a short unique nucleotide sequence, can be incorporated into the adapter in order to identify which specimen the DNA originated from (Meyer & Kircher 2010). This allows for multiple specimens to be sequenced at the same time (multiplexing). Paired-end sequencing, in which each DNA fragment is sequenced from both ends, can be used to ensure higher quality data (Kircher *et al*. 2011). After sequencing, the data are quality filtered; a process which includes the removal of adapter and poor quality sequence, and the merging of the paired-end sequences (Kircher 2012), ahead of downstream

analysis (see chapter two). NGS data derived from a shotgun-based approach are suitable for characterising the preservation, such as average fragment length, of aDNA, as well as assessing the taxonomic composition of the DNA present in museum and ancient remains.

1.5. The Late Quaternary, Beringia, and Permafrost-Preserved Insects

The late Quaternary, which includes the Late Pleistocene stage (126 to 11.7ka) and Holocene epoch (11.7ka to present), was a period of major climatic and environmental change and encompassed the most recent full glacial cycle or ice age (fig. 1.1) (Gradstein *et al*. 2012). During glaciations, sea levels were up to 130m lower than they are today, exposing large areas of now inundated land (Hopkins 1973; Lambeck *et al*. 2002).

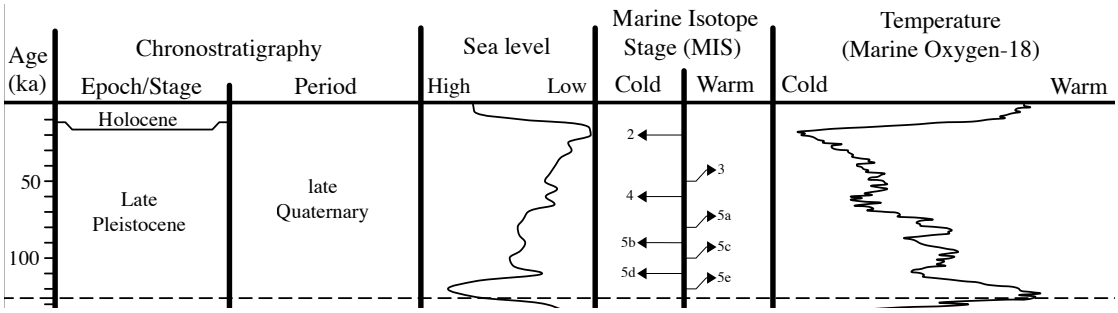


Figure 1.1 The late Quaternary subdivisions and climate proxies. A single glacial cycle has occurred during this period (126 to 0ka). Modified from Gradstein and colleagues (2012).

Beringia was a region that existed at these times and consisted of northeastern Siberia, the now inundated Bering Land Bridge, lowland Alaska, and the unglaciated regions of the Yukon (fig. 1.2) (Elias & Crocker 2008; Hultén 1937). It connected the Palearctic (Europe and Asia) and Nearctic (North America) biogeographic realms, which together constitute the Holarctic, and therefore provided a major terrestrial migration route at times of low sea level (Sher 1999). In addition, and despite its northerly latitude, lowland Beringia remained free from glaciers throughout the ice ages and therefore provided the arctic fauna with a refuge during glaciations (Hultén 1937; Pielou 1991). Permafrost, which is excellent for both morphological and molecular preservation, has persisted in Beringia since these glaciations and has

resulted in an excellent palaeontological record of the Beringian flora and fauna through multiple ice ages (Shapiro & Cooper 2003).

Figure 1.2 Map of Beringia, with the maximum extent of the Bering Land Bridge (lightest green) during the last glaciation. Relief indicated by the shade of green (darker represents greater altitude). In East Beringia (Alaska, Yukon), only the lowland areas were unglaciated. Modified from Elias and Crocker (2008), after Hoffecker and Elias (2007).

The late Quaternary permafrost deposits of Beringia are a rich source of preserved subfossil insect remains, of which beetles are the most commonly recovered (Elias 2010). This is due to their hard exoskeleton, which is usually recovered as isolated sclerites, increasing their chance of preservation (Coope 2004). Nearly all elements of the Quaternary permafrost-preserved beetle fauna are extant, with many species being stenothermic (narrow temperature range tolerance) and adapted to specific habitats (Elias 2010). As permafrost-preserved sclerites are often identifiable to the species level using exoskeletal morphology (Coope 2004), these beetle remains can therefore provide valuable evidence for palaeoclimatic and palaeoenvironmental interpretations of Beringia and how the climate and ecosystem changed through time (Elias 2010).

Despite their abundance and importance, the possibility of aDNA investigation using permafrost-preserved beetles has only recently been demonstrated (Thomsen *et al.* 2009). The vast majority of previous Beringian aDNA studies have focused on the large mammal (mega) fauna, and have assessed a broad suite of important questions related to extinction, past population history, and the effect of past climatic changes

on these events and processes (Barnes *et al.* 2002; Barnes *et al.* 2007; Barnett *et al.* 2009; Campos *et al.* 2010; Lorenzen *et al.* 2011; Shapiro *et al.* 2004). Therefore, the use of an aDNA approach for the assessment of biologically meaningful questions in beetles is both timely and warranted. Considering their role in maintaining properly functioning terrestrial ecosystems, one such suitable and pressing area for investigation is the response of beetle species to past climatically driven change.

1.6. Beetle Responses to Climatically Driven Change

Based on scenarios of near-future climatic change, the present distributions of habitats and ecosystems are likely to continue migrating toward the poles for at least the next century (Bellard *et al.* 2012; Lawler *et al.* 2009; Parmesan 2006). In order to survive, species with narrow climatic and/or ecological tolerances (stenotherms) will need to respond by tracking these changing distributions. Understanding how species will respond to these challenges is of major importance for conservation and to ensure the health of future ecosystem function (Bellard *et al.* 2012; Botkin *et al.* 2007; Meyers & Bull 2002). In order to assess how species will respond to climatic change in the future, known responses from past episodes of climatic and environmental change can be used to inform potential future responses.

Although studies can rudimentarily assess species responses based on present species distributions, through phylogeographic or regional analyses, a rigorous study requires genetic information not only from the present, but also from the past when these climatic changes were actually occurring. This is because the modern record can be misleading, due to past demographic processes such as population bottlenecks (Willerslev & Cooper 2005). The only way to obtain genetic information from the past is by using aDNA.

The last ice age provides an ideal time period for this type of investigation, as it was a time of major climatic fluctuations in the geologically recent past and previous studies have shown a general trend of major species level responses to these environmental changes, based on both modern and past distributions (Brace *et al.* 2012; Dalén *et al.* 2007; Hewitt 1999; Stewart *et al.* 2009). North America provides an ideal geographic setting because, during the height (26,500 to 19,000 cal. yrs BP; fig. 1.1) of the last

ice age, the uninhabitable Laurentide and Cordilleran ice sheets covered Canada, with the exception of the Beringian areas of the Yukon, and the ice margin extended south to the northernmost areas of the lower 48 states of the USA (fig. 1.3) (Clark *et al*. 2009; Dyke 2004). At this time, the cold-adapted fauna persisted in refugia: Beringia (western refugium; figs. 1.2, 1.3) and the lower 48 states of the USA (southern refugium; fig. 1.3) (Hultén 1937; Pielou 1991). After deglaciation initiated, around 15,500 cal. yrs BP (Dyke 2004), the fauna colonised the newly exposed landmasses, although the mode of colonisation differed between species (Anderson *et al.* 2006; Aubry *et al.* 2009; Beatty & Provan 2010, 2011; Eidesen *et al.* 2007; Fedorov *et al.* 2003; Weckworth *et al.* 2010).

Based on a synthesis of modern and subfossil distributional evidence, the Schwert-Ashworth (1988) model describes how the cold-adapted beetle fauna colonised Canada during deglaciation (fig. 1.3).

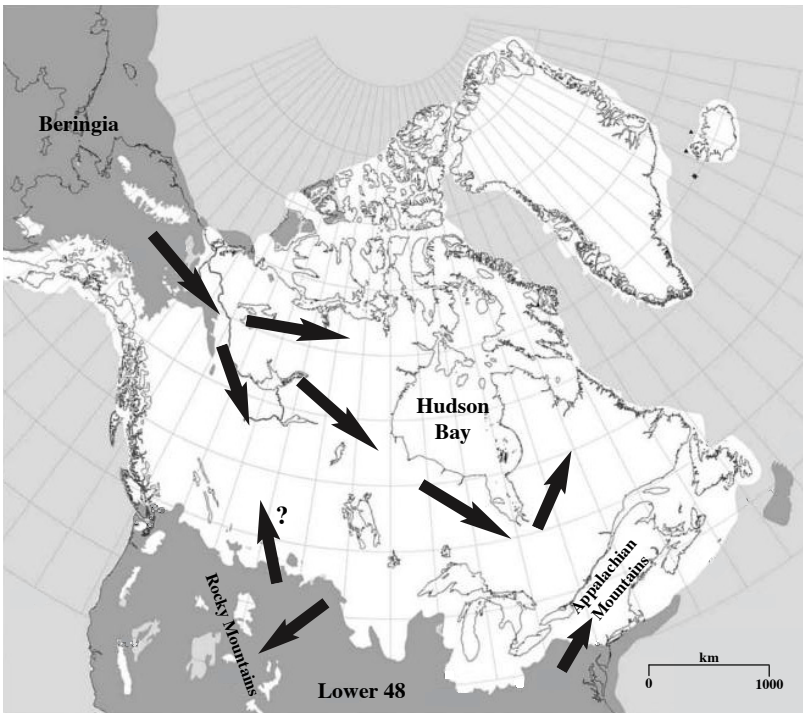


Figure 1.3 The Schwert-Ashworth (1988) model of the postglacial colonisation mode of Canada by the arctic beetle fauna, as exemplified by *Amara alpina*. Western refugium: Beringia, southern refugium: lower 48 of the USA, dispersal barrier: Hudson Bay. White: extent of the Laurentide and Cordilleran ice sheets at 18ka, just after maximal extent. Light grey: water bodies. Dark grey: exposed land. Map modified from Dyke (2004). Colonisation routes are based on Ashworth (1996), Reiss *et al*. (1999), and Schwert and Ashworth (1988).

The model states that populations in the southern refugium (lower 48) survived warming by colonising high altitude mountains (of the Rockies and Appalachians), and that Canada was either primarily or exclusively colonised by eastward dispersal from Beringia. A problem with this model, identified by Schwert and Ashworth (1988), is Hudson Bay. This major physiographic barrier to eastward dispersal was the last area of mainland North America to deglaciate, around 8,000 cal. yrs BP (Carlson *et al*. 2008). The only way for flightless ground beetles to cross this barrier would have been via the regions south of Hudson Bay, which may have been problematic for the cold-adapted beetle fauna. Distributional evidence of some ground beetle species strongly suggests that Hudson Bay was a barrier to dispersal (Schwert & Ashworth 1988), although the model maintains that species now found east of Hudson Bay must have crossed this barrier. An alternative scenario is that this region was colonised at least partially from the southern refugium, which would suggest that Hudson Bay was a stronger barrier to dispersal than is currently thought. Resolution of these scenarios may be informative for future migration route models of the North American ground beetle fauna in response to predicted climatic change.

1.7. Study Taxa

Two closely related ground beetle (Coleoptera: Carabidae) species were selected for investigation in this thesis, *Amara alpina* and *Pterostichus brevicornis*. These species are both well represented in museum collections and the permafrost deposits of Beringia (Elias *et al.* 2000). They have a north Holarctic, near-circumpolar distribution and are also found on mountain peaks of the lower 48 states of the USA (Ball 1966; Lindroth 1966, 1968). In addition, these species are flightless (Ashworth *et al.* 1996; Lindroth 1968, 1969; Schwert & Ashworth 1988), important for palaeoclimatic and palaeoenvironmental reconstruction due to their stenothermic nature (Elias 2010), and are cold adapted, with *A. alpina* being the most cold-adapted ground beetle (Bennike *et al.* 2000). The latter two points would indicate that both species are especially likely to respond to even minor levels of future climatic change. These species differ in their ecological requirements; *A. alpina* is found in both xeric (dry) and mesic (moist) habitats, whereas *P. brevicornis* is adapted to mesic habitats (Elias & Crocker 2008; Lindroth 1968). These two species could therefore be used to

assess whether potential inferences were specific, or could be generalised across the arctic ground beetle fauna.

1.8. Thesis Aims

There were two overarching aims to this thesis:

1. First was to assess the potential of DNA from recent (<50yr) and old (>50yr) museum, as well as ancient permafrost-preserved, specimens of the Holarctic ground beetles *A. alpina* and *P. brevicornis*. This included:
    a. Investigating the proportion of specimens in which DNA is preserved using a Sanger sequencing based approach (chapter three).
    b. Characterising the state of DNA preservation in these remains, as inferred from Sanger sequencing (chapter three) and NGS (chapter four).
    c. Assessing the proportion of endogenous, and taxonomically characterising the remaining, DNA that could be recovered from these remains using a shotgun-based NGS approach (chapter five).
2. Second was to assess how Canada was colonised after the last glaciation by the ground beetles, *A. alpina* (chapter six) and *P. brevicornis* (chapter seven), using a combined museum and ancient DNA approach.

In addition, several chapter-specific questions were investigated. These are introduced and discussed in their respective results chapters and included:

- Assessing the misidentification rate for permafrost-preserved specimens (chapter three).
- Exploring some of the bioinformatic aspects for the analysis of shotgun-generated NGS data from a taxon lacking a suitable reference genome (chapter five).
- Examining whether there was population turnover in *A. alpina* during the ice age (chapter six).
- Investigating the taxonomy of modern *P. brevicornis* (chapter seven).

The overarching and specific aims were investigated in chapters three to seven, with the overall findings outlined and discussed in chapter eight.

1.9. References

Allentoft ME, Collins M, Harker D, *et al*. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society of London Series B-Biological Sciences* **279**, 4724-4733.

Andersen JC, Mills NJ (2012) DNA extraction from museum specimens of parasitic hymenoptera. *PLoS One* **7**, e45549.

Anderson LL, Hu FS, Nelson DM, Petit RJ, Paige KN (2006) Ice-age endurance: DNA evidence of a white spruce refugium in Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12447-12450.

Ashworth AC (1996) The response of arctic Carabidae (Coleoptera) to climate change based on the fossil record of the Quaternary Period. *Annales Zoologici Fennici* **33**, 125-131.

Ashworth AC, Schwert DP, Reiss RA (1996) Biogeography of Arctic beetles: Integrating the results of paleontological and molecular population genetic studies. *Arctic Insect News* **7**, 12-15.

Aubry KB, Statham MJ, Sacks BN, Perrine JD, Wisely SM (2009) Phylogeography of the North American red fox: vicariance in Pleistocene forest refugia. *Molecular Ecology* **18**, 2668-2686.

Austin JJ, Ross AJ, Smith AB, Fortey RA, Thomas RH (1997a) Problems of reproducibility - Does geologically ancient DNA survive in amber-preserved insects? *Proceedings of the Royal Society of London Series B-Biological Sciences* **264**, 467-474.

Austin JJ, Smith AB, Thomas RH (1997b) Palaeontology in a molecular world: the search for authentic ancient DNA. *Trends in Ecology and Evolution* **12**, 303-306.

Ball GE (1966) A revision of the North American species of the subgenus *Cryobius* Chaudoir (*Pterostichus*, Carabidae, Coleoptera). *Opuscula Entomologica Supplementum* **28**, 1-166.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **295**, 2267-2270.

Barnes I, Shapiro B, Lister A, *et al*. (2007) Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Current Biology* **17**, 1072-1075.

Barnett R, Shapiro B, Barnes I, *et al*. (2009) Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* **18**, 1668-1677.

Beatty GE, Provan J (2010) Refugial persistence and postglacial recolonization of North America by the cold-tolerant herbaceous plant *Orthilia secunda*. *Molecular Ecology* **19**, 5009-5021.

Beatty GE, Provan J (2011) Phylogeographic analysis of North American populations of the parasitic herbaceous plant *Monotropa hypopitys* L. reveals a complex history of range expansion from multiple late glacial refugia. *Journal of Biogeography* **38**, 1585–1599.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters* **15**, 365-377.

Bennike O, Björck S, Böcher J, Walker IR (2000) The Quaternary arthropod fauna of Greenland: a review with new data. *Bulletin of the Geological Society of Denmark* **47**, 111–134.

Botkin DB, Saxe H, Araujo MB, *et al*. (2007) Forecasting the effects of global warming on biodiversity. *Bioscience* **57**, 227-236.

Brace S, Palkopoulou E, Dalén L, *et al*. (2012) Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20532–20536.

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

Cano RJ, Poinar HN, Pieniazek NJ, Acra A, Poinar GO (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature* **363**, 536-538.

Carlson AE, Legrande AN, Oppo DW, *et al*. (2008) Rapid early Holocene deglaciation of the Laurentide ice sheet. *Nature Geoscience* **1**, 620-624.

Castalanelli MA, Severtson DL, Brumley CJ, *et al*. (2010) A rapid non-destructive DNA extraction method for insects and other arthropods. *Journal of Asia-Pacific Entomology* **13**, 243–248.

Chapco W, Litzenberger G (2004) A DNA investigation into the mysterious disappearance of the Rocky Mountain grasshopper, mega-pest of the 1800s. *Molecular Phylogenetics and Evolution* **30**, 810-814.

Clark PU, Dyke AS, Shakun JD, *et al*. (2009) The Last Glacial Maximum. *Science* **325**, 710-714.

Coope GR (2004) Several million years of stability among insect species because of, or in spite of, Ice Age climatic instability? *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 209-214.

Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* **289**, 1139-1139.

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

Desalle R, Gatesy J, Wheeler W, Grimaldi D (1992) DNA-Sequences from a Fossil Termite in Oligomiocene Amber and Their Phylogenetic Implications. *Science* **257**, 1933-1936.

Dyke AS (2004) An outline of North American deglaciation with emphasis on central and northern Canada. In: *Quaternary Glaciations-Extent and Chronology — Part II: North America* (eds. Ehlers J, Gibbard PL), pp. 373-406. Elsevier, Amsterdam.

Eidesen PB, Carlsen T, Molau U, Brochmann C (2007) Repeatedly out of Beringia: *Cassiope tetragona* embraces the arctic. *Journal of Biogeography* **34**, 1559-1574.

Elias SA (2010) *Advances in Quaternary Entomology* Elsevier, Oxford, UK.

Elias SA, Berman D, Alfimov A (2000) Late Pleistocene beetle faunas of Beringia: where east met west. *Journal of Biogeography* **27**, 1349–1363.

Elias SA, Crocker B (2008) The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota? *Quaternary Science Reviews* **27**, 2473-2483.

Fedorov VB, Goropashnaya AV, Jaarola M, Cook JA (2003) Phylogeography of lemmings (*Lemmus*): no evidence for postglacial colonization of Arctic from the Beringian refugium. *Molecular Ecology* **12**, 725-731.

Gibson CM, Kao RH, Blevins KK, Travers PD (2012) Integrative taxonomy for continental-scale terrestrial insect observations. *PLoS One* **7**, e37528.

Gilbert MTP, Bandelt HJ, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology and Evolution* **20**, 541-544.

Gilbert MTP, Moore W, Melchior L, Worobey M (2007) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One* **2**, e272.

Goldstein PZ, Desalle R (2003) Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. *Molecular Ecology* **12**, 1993-1998.

Gradstein FM, Ogg JG, Schmitz MA, Ogg GM (2012) *A Geologic Time Scale 2012* Elsevier, Oxford, UK.

Grove SJ, Stork NE (2000) An inordinate fondness for beetles. *Invertebrate Taxonomy* **14**, 733-739.

Gullan PJ, Cranston PS (2010) *The Insects: An Outline of Entomology*, 4th edn. Wiley-Blackwell, UK.

Gutierrez G, Marin A (1998) The most ancient DNA recovered from an amber-preserved specimen may not be as ancient as it seems. *Molecular Biology and Evolution* **15**, 926-929.

Harper GL, Maclean N, Goulson D (2006) Analysis of museum specimens suggests extreme genetic drift in the adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society* **88**, 447–452.

Hartley CJ, Newcomb RD, Russell RJ, *et al*. (2006) Amplification of DNA from preserved specimens shows blowflies were preadapted for the rapid evolution of insecticide resistance. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 8757-8762.

Hebsgaard MB, Phillips MJ, Willerslev E (2005) Geologically ancient DNA: fact or artefact? *Trends in Microbiology* **13**, 212-220.

Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**, 87-112.

Hoffecker JF, Elias SA (2007) *Human Ecology of Beringia* Columbia University Press, New York.

Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. *Nature Reviews Genetics* **2**, 353-359.

Hopkins DM (1973) Sea level history in Beringia during the past 250,000 years. *Quaternary Research* **3**, 520-540.

Hultén E (1937) *Outline of the history of arctic and Boreal biota during the Quaternary period* Bokförlags Aktiebolaget Thule, Stockholm.

King GA, Gilbert MTP, Willerslev E, Collins MJ, Kenward H (2009) Recovery of DNA from archaeological insect remains: first results, problems and potential. *Journal of Archaeological Science* **36**, 1179-1183.

Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. In: *Ancient DNA: Methods and Protocols, Methods in Molecular Biology* (eds. Shapiro B, Hofreiter M), pp. 197-228. Springer.

Kircher M, Heyn P, Kelso J (2011) Addressing challenges in the production and analysis of illumina sequencing data. *Bmc Genomics* **12**, 382.

Knapp M, Hofreiter M (2010) Next generation sequencing of ancient DNA: requirements, strategies and perspectives. *Genes* **1**, 227-243.

Krings M, Stone A, Schmitz RW, *et al*. (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19-30.

Lambeck K, Esat TM, Potter E-K (2002) Links between climate and sea levels for the past three million years. *Nature* **419**, 199-206.

Lawler JJ, Shafer SL, White D, *et al*. (2009) Projected climate-induced faunal change in the Western Hemisphere. *Ecology* **90**, 588-597.

Lindahl T (1993) Instability and Decay of the Primary Structure of DNA. *Nature* **362**, 709-715.

Lindroth CH (1966) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 4. *Opuscula Entomologica, Supplementum* **29**, 409-648.

Lindroth CH (1968) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 5. *Opuscula Entomologica, Supplementum* **33**, 649-944.

Lindroth CH (1969) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 1. *Opuscula Entomologica, Supplementum* **29**, I—XLVIII.

Lorenzen ED, Nogues-Bravo D, Orlando L, *et al*. (2011) Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359-364.

Mandrioli M (2008) Insect collections and DNA analyses: how to manage collections? *Museum Management and Curatorship* **23**, 193 -199.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb prot5448.

Meyer M, Kircher M, Gansauge MT, *et al*. (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222-226.

Meyers LA, Bull JJ (2002) Fighting change with change: adaptive variation in an uncertain world. *Trends in Ecology and Evolution* **17**, 551-557.

Moraes CT (2001) What regulates mitochondrial DNA copy number in animal cells? *Trends in Genetics* **17**, 199-205.

New TR (2007) Beetles and conservation. *Journal of Insect Conservation* **11**, 1-4.

Paabo S, Poinar H, Serre D, *et al*. (2004) Genetic analyses from ancient DNA. *Annual Review of Genetics* **38**, 645-679.

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics* **37**, 637-669.

Pielou EC (1991) *After the Ice Age: The return of life to glaciated North America* University of Chicago Press, Chicago.

Poinar HN, Hoss M, Bada JL, Paabo S (1996) Amino acid racemization and the preservation of ancient DNA. *Science* **272**, 864-866.

Post E, Forchhammer MC, Bret-Harte MS, *et al*. (2009) Ecological dynamics across the Arctic associated with recent climate change. *Science* **325**, 1355-1358.

Rasmussen M, Guo X, Wang Y, *et al*. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98.

Reiss RA (2006) Ancient DNA from ice age insects: proceed with caution. *Quaternary Science Reviews* **25**, 1877-1893.

Reiss RA, Ashworth AC, Schwert DP (1999) Molecular genetic evidence for the post-Pleistocene divergence of populations of the arctic-alpine ground beetle *Amara alpina* (Paykull) (Coleoptera: Carabidae). *Journal of Biogeography* **26**, 785-794.

Rowe KC, Singhal S, Macmanes MD, *et al*. (2011) Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* **11**, 1082-1092.

Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131.

Schaefer H, Heibl C, Renner SS (2009) Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous oversea dispersal events. *Proceedings of the Royal Society of London Series B-Biological Sciences* **276**, 843-851.

Schwert DP, Ashworth AC (1988) Late Quaternary History of the Northern Beetle Fauna of North-America - a Synthesis of Fossil and Distributional Evidence. *Memoirs of the Entomological Society of Canada*, 93-107.

Shapiro B, Cooper A (2003) Beringia as an Ice Age genetic museum. *Quaternary Research* **60**, 94-100.

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

Sher A (1999) Traffic lights at the Beringian crossroads. *Nature* **397**, 103-104.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* **21**, 1794-1805.

Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**, 203-217.

Stewart JR, Lister AM, Barnes I, Dalen L (2009) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London Series B-Biological Sciences* **277**, 661-671.

Strange JP, Knoblett J, Griswold T (2009) DNA amplification from pin-mounted bumble bees (*Bombus*) in a museum collection: effects of fragment size and specimen age on successful PCR. *Apidologie* **40**, 134–139.

Suarez AV, Tsutsui ND (2004) The value of museum collections for research and society. *Bioscience* **54**, 66-74.

Tagliavia M, Massa B, Albanese I, La Farina M (2011) DNA extraction from Orthoptera museum specimens. *Analytical Letters* **44**, 1058–1062.

Thomsen PF, Elias S, Gilbert MT, *et al*. (2009) Non-destructive sampling of ancient insect DNA. *PLoS One* **4**, e5048.

Ugelvig LV, Nielsen PS, Boomsma JJ, Nash DR (2011) Reconstructing eight decades of genetic variation in an isolated Danish population of the large blue butterfly *Maculinea arion*. *BMC Evolutionary Biology* **11**, 201.

van Houdt JK, Breman FC, Virgilio M, de Meyer M (2010) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources* **10**, 459-465.

Walden KKO, Robertson HM (1997) Ancient DNA from amber fossil bees? *Molecular Biology and Evolution* **14**, 1075-1077.

Wandeler P, Hoeck PE, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution* **22**, 634-642.

Watts PC, Thompson DJ, Allen KA, Kemp SJ (2007) How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses? *Journal of Insect Conservation* **11**, 195–198.

Weckworth BV, Talbot SL, Cook JA (2010) Phylogeography of wolves (*Canis lupus*) in the Pacific Northwest. *Journal of Mammalogy* **91**, 363-375.

Willerslev E, Cappellini E, Boomsma W, *et al*. (2007) Ancient biomolecules from deep ice cores reveal a forested Southern Greenland. *Science* **317**, 111-114.

Willerslev E, Cooper A (2005) Ancient DNA. *Proceedings of the Royal Society of London Series B-Biological Sciences* **272**, 3-16.

Chapter 2. Materials and methods

This chapter describes the general procedures that were used for multiple chapters (three to seven). Methods outlined in sections 2.1 to 2.3 were used for all results chapters, those in section 2.4 (Sanger sequencing) were used for chapters three, six, and seven, whereas those in section 2.5 (NGS) were employed for chapters four and five.

2.1. Specimen Collection

Specimens fell into three age classes: modern (<10 years old), museum (from dry-stored collections; >10 years old), and ancient (from permafrost deposits). All specimens, including age and provenance, are listed in appendix 1.

Modern specimens were collected from seven localities between 2002 and 2004 by S. Kuzmina. Dry-stored museum specimens, that had either been pinned or glued to a card mount, were gathered from the Canadian National Collection of Insects, Arachnids, and Nematodes (CNC) in April 2010 (*Amara alpina*, *Pterostichus brevicornis*) and the Swedish Museum of Natural History (NRM) in May 2011 (*A. alpina*). These specimens ranged in age from 137 to 11yrs old (as of 2012), and originated from 105 localities. These localities represent the entire modern distribution of *A. alpina*, with the exception of an isolated Scottish population (Lindroth 1968; Luff 2007), and the complete North American range of *P. brevicornis* (Ball & Currie 1997). Although complete dry-stored museum specimens can be used for DNA extraction in a non-destructive manner (Castalanelli *et al*. 2010; Gilbert *et al*. 2007b), the long-term effects on the structural integrity of specimens, through weakening of desiccated soft tissues connecting the sclerites, is not known (M.T.P. Gilbert, pers. comm.). Therefore, in order to preserve specimens (Mandrioli 2008), a single non-morphologically identifiable hind leg (including femur, tibia, tarsi; fig. 2.1a-c) was removed for analysis using tweezers, which were cleaned after each collection with bleach and water to minimise sample cross contamination.

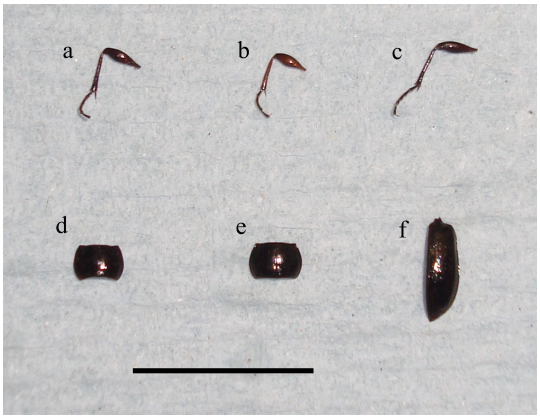


Figure 2.1 Museum and ancient specimens of *A. alpina*. All specimens have undergone DNA extraction. a-c: museum hind leg specimens from the CNC (a: Quebec, b: NWT, c: Alaska), d-f: ancient sclerite specimens (d: TR Pleistocene, e-f: TR Holocene). d-e: pronota, f: elytron. a-d yielded DNA. Scale bar is 10mm.

Ancient specimens, consisting of individual complete or broken sclerites [pronota (thoracic shields) or elytra (wing cases); fig. 2.1d-f], were collected from 30 Beringian sites between 1983 and 2006 (table 2.1). All remains represent adults. Specimens ranged in age from >560,000 to 5,800 cal. yrs BP, based on either exact radiocarbon dating of associated plant remains, relative tephra dating of layers above or below beetle-bearing sediments, or through rough stratigraphic correlation. Samples were removed from the sediment by bucket sieving, using water and a 300µm screen. They were subsequently dry picked and stored at room temperature (Elias 1994, 2010). Kerosene, a potentially destructive agent to DNA (Reiss 2006), was not used to isolate specimens. Prior to DNA extraction, ancient specimens were stored at -20°C to reduce further degradation (Reiss 2006).

2.2. Contamination Reduction

DNA extraction and PCR reaction preparation of degraded samples (ancient and museum) was performed in a dedicated aDNA laboratory which was physically isolated from the laboratory in which PCR and subsequent downstream reactions were conducted [following Cooper, Poinar (2000); Gilbert *et al*. (2005); Hofreiter *et al*. (2001); Wandeler *et al*. (2007)]. To prevent contamination of degraded samples, all non-biological materials, equipment, and surfaces were sterilised with bleach. The laboratory atmosphere was sterilised with UV light and small equipment was

		Locality		Province/			Age		Collection		
No.	Sediment	Name	Abbreviation	State	Latitude	Longitude	RC	Cal.	Year	Species	N
1	GZ-04-44a	Goldbottom Creek	N/A	Yukon	63.97	-138.97	25300	28100	2004	*A. a.*	50
2	T03-B17	Titaluk River	TR Holocene	Alaska	69.70	-155.20	?Hol.	?Hol.	2003	*A. a.*	21
3	GZ-06-57	Old Crow Loc. 106	Old Crow 106	Yukon	67.43	-140.37	29600	32400	2006	*A. a./P. b.*	17/4
4	GZ-04-44b	Goldbottom Creek	N/A	Yukon	63.97	-138.97	25300	28100	2004	*A. a.*	15
5	DF-06-101	Old Crow Loc. 11	Old Crow 11	Yukon	67.48	-140.10	41000	42800	2006	*A. a./P. b.*	11/7
6	T03-B08b	Titaluk River	TR Pleistocene	Alaska	69.70	-155.20	LP	LP	2003	*A. a.*	10
7	DF-06-90	Rock River	N/A	NWT	67.80	-130.85	35000	38100	2006	*A. a./P. b.*	5/2
8	DF-04-67	Old Crow Loc. 94	N/A	Yukon	67.43	-140.37	40000	42100	2004	*A. a.*	3
9	DF-06-92	Rock River	N/A	NWT	67.80	-130.85	55000	53100	2006	*A. a.*	3
10	AL-3-05-B13	Stevens Village	N/A	Alaska	66.02	-149.08	7000	5900	2005	*A. a./P. b.*	3/2
11	AL-4-05-B22	Palisades	N/A	Alaska	65.10	-153.25	LP	LP	2005	*A. a.*	3
12	AL-5-05-B28	Chester Bluff	N/A	Alaska	64.47	-142.72	>560 ka	>560000	2005	*A. a.*	3
13	GZ-03-128	Dominion Creek	N/A	Yukon	63.75	-139.30	>100 ka	>100000	2003	*A. a.*	2
14	PT-CFx	Ledovy Obryv	N/A	Chukotka	65.10	171.18	26000	28800	2004	*A. a.*	1
15	GZ-05-48	Quartz Creek	N/A	Yukon	63.78	-139.22	25000	27900	2005	*A. a.*	1
16	83 IGU 11-3	Igushik	N/A	Alaska	58.90	-159.07	30000	32700	1983	*P. b.*	5
17	84 IGU 1-4	Igushik	N/A	Alaska	58.90	-159.07	16000	17200	1984	*P. b.*	4
18	84 IGU 1-5	Igushik	N/A	Alaska	58.90	-159.07	16000	17200	1984	*P. b.*	4
19	Al-4-05-B21	Palisades	N/A	Alaska	65.10	-153.25	LP	LP	2005	*P. b.*	4
20	83FF 3-M-13	Flounder Flat	N/A	Alaska	58.73	-158.38	12000	11900	1983	*P. b.*	3
21	K/BR-24	Kuskokwim River	N/A	Alaska	61.88	-157.85	37500	40200	N/A	*P. b.*	3
22	83 IGU 1-7	Igushik	N/A	Alaska	58.90	-159.07	35000	38100	1983	*P. b.*	2
23	K/BR-22	Kuskokwim River	N/A	Alaska	61.88	-157.85	37500	40200	N/A	*P. b.*	2
24	NBI 22-25	North Barter Island	N/A	Alaska	70.12	-143.67	Mid Hol.	5800	N/A	*P. b.*	2
25	CP 2-9	Coffee Point	N/A	Alaska	58.20	-157.45	13000	13700	N/A	*P. b.*	1
26	12-M4	Etolin Point	N/A	Alaska	58.62	-158.25	31000	33700	N/A	*P. b.*	1
27	83FF 1-5	Flounder Flat	N/A	Alaska	58.73	-158.38	12000	11900	1983	*P. b.*	1
28	85 IGU-M-1	Igushik	N/A	Alaska	58.90	-159.07	15000	16300	1985	*P. b.*	1
29	85 IGU-M-5	Igushik	N/A	Alaska	58.90	-159.07	13000	13700	1985	*P. b.*	1
30	85 IGU-M-5A	Igushik	N/A	Alaska	58.90	-159.07	13000	13700	1985	*P. b.*	1

Previous page: *Table 2.1* Data on the ancient localities used in this thesis. *A.a.*: *Amara alpina*, *P. b.*: *Pterostichus brevicornis*. N: number of samples from each locality (*A. alpina*/*P. brevicornis*). LP: Late Pleistocene. ?Hol.: ?Holocene. RC: Radiocarbon years before present. Cal.: Calendar years before present. RC ages are approximate and were calibrated using OxCal (https://c14.arch.ox.ac.uk/oxcal/OxCal.html).

irradiated in a UV cross-linker for 20 minutes prior to work commencing. Additionally, a full body suit, foot covers, gloves, and a facemask were worn at all times in the aDNA laboratory.

2.3. DNA Extraction

DNA from degraded specimens was extracted using the QIAamp DNA Micro extraction kit (Qiagen), with a modified version of the tissue extraction protocol. This included using the optional carrier RNA to increase DNA yield and conducting the final eluting step twice (50µL each). The lysis step was conducted for between 16 and 20 hours. Tween-20 (final concentration of 0.05%) was added to the extract to ensure long-term viability. Extraction controls were used in a ratio of one control to five samples. Specimens were not disintegrated prior to extraction as initial investigation indicated that this did not affect the likelihood of DNA recovery. Modern DNA was extracted by Mack (2008) using the same method, with the exception that complete specimens were digested instead of a single hind leg [following Gilbert *et al*. (2007b)]. Two extracts of *A. alpina* (one modern, one ancient), received from P.F. Thomsen, were extracted in a previously published study (Thomsen *et al*. 2009).

2.4. Sanger Sequencing and Quality Control

2.4.1. Genetic Markers Targeted and Primer Design

Four genetic markers were targeted: mitochondrial COI and COII, and multi-copy nuclear 28S and ITS1. Details of the targeted regions, including length, are listed in table 2.2. Novel species-specific primer sets were designed using Oligo (Rychlik & Rychlik 2005). Templates for primer design were based on data from Mack (2008),

Thomsen *et al*. (2009), Gilbert *et al*. (2007b) and Genbank. All primer sequences are listed in appendix 2a.

	Marker (bps)			
Species	**COI**	**COII**	**28S (D3)**	**ITS1**
Amara alpina	759	N/A	183	210
Pterostichus brevicornis	336	317	164 (172)	210

Table 2.2 Details of the genetic markers employed in chapters three, five, and six of this thesis. Figure in parentheses indicates several samples that had a longer D3 region of 28S rDNA.

2.4.2. PCR Amplification

Amplification of extracts was performed using a hot-start PCR procedure, with reactions consisting of 1x PCR buffer, 1mM of additional $MgCl_2$, 1mg/mL BSA [following (Rohland & Hofreiter 2007)], 200μM of each dNTP, 0.4μM of each primer (forward and reverse), 1U HotStar *Taq* (Qiagen), 1μL of DNA extract, and purified water to give a final volume of 25μL. Reactions were performed in a Peltier Thermal Cycler, using six sequential steps (step 1: 95°C for five minutes, step 2: 94°C for one minute, step 3: variable annealing temperature for one minute, step 4: 72°C for one minute, step 5: 72°C for ten minutes, step 6: 12°C for ten minutes), of which steps 2 to 4 were repeated a further 49 times. Primer sets and annealing temperatures are listed in appendix 2b. The presence of amplicons in PCR products was determined using electrophoresis of a 2% agarose, ethidium bromide stained gel.

2.4.3. DNA Sequencing

PCR products yielding amplicons were purified enzymatically using an Exonuclease I and Shrimp Alkaline Phosphatase (Exo-SAP) procedure. Each reaction consisted of 8U Exo, 0.4U SAP (both Thermo Scientific), purified water, and 10μL PCR product, to give a final volume of 12μL. Reactions were performed in a Peltier Thermal Cycler, using three sequential steps (step 1: 37°C for twenty minutes, step 2: 80°C for twenty minutes, step 3: 4°C for six minutes). Purified PCR products were sequenced in both directions using a high throughput ABi 3730xl genetic analysis capillary sequencer at Macrogen (Seoul, South Korea; Amsterdam, the Netherlands).

2.4.4. Sequence Data Quality Control

Sequence data were quality-checked manually and concatenated in Sequencher v4.7 (Gene Codes). Individual and concatenated sequences were compared against the Basic Local Alignment Search Tool (BLAST) database to ensure authenticity of results and detect contamination. Repeated amplifications and the application of multiple overlapping primer sets (appendix 2b) were used to infer sequence errors caused by potential NUMTs and miscoding lesions. If these were detected (in the form of mismatches or indels between overlapping fragments), then a majority rule approach was employed, with the base that occurred the most frequently being chosen. Consideration was also given to frequency data on various damage types (Gilbert *et al*. 2007a) and if the same base position was repeatedly problematic within and between samples (indicative of potential NUMT contamination). If ambiguity as to the correct base designation remained, bases were considered missing data. DNA sequences of protein coding regions (COI, COII) were translated in Se-Al v2.0 (Rambaut 2002) to detect potential unexpected stop codons, which may have arisen from miscoding lesions or NUMT contamination.

2.4.5. Independent Replication

To improve confidence in result authenticity, analysis of *A. alpina* ancient specimens was performed at two institutions (RHUL, NRM). Due to the small size of specimens, sub-sampling was not performed. Instead, different specimens from the same sample batches were analysed separately at the different institutions (appendix 1). Protocols used at RHUL were as described above. NRM protocols differed in the use of GelRed (Biotium) as the gel electrophoresis stain and Fast Alkaline Phosphatase for PCR product purification. Additionally, cleaned products were sequenced in-house at the NRM, using BigDye Terminator chemistry (v1.1) and an ABi 3100 capillary genetic analyser (Applied Biosystems). Results from both institutions were compatible, and in most cases identical, indicating that sequences obtained from ancient specimens are not the result of intra-laboratory contamination. Additionally, sequences were identical to those obtained independently in previously published (Thomsen *et al*. 2009) and unpublished (Mack 2008) studies.

2.5. Next-Generation Sequencing and Quality Control

2.5.1. Sample Selection

Six samples from the *A. alpina* dataset were used for NGS (table 2.3). DNA had been recovered from all of these samples using Sanger sequencing, with maximum amplification success (*sensu* chapter three). Samples were chosen to include the full age range of specimens [two each of modern (Mo), museum (Mu), ancient (An)] and at least one sample from each of the three mitochondrial haplogroups identified in chapter six. *P. brevicornis* was not used for NGS, as DNA could not be recovered from ancient specimens (chapter three).

2.5.2. Illumina DNA Library Construction and Sequencing

Prior to library construction, the concentration of DNA was crudely assessed through gel electrophoresis of 4μL of each DNA extract. All extracts lacked a visible band, demonstrating that DNA was of low concentration in all samples, and so all libraries were constructed using the same protocol. 20μL of DNA extract (14μL for Mo2, due to exhaustion of extract) was used for library construction.

DNA libraries were constructed using a modified version of the Meyer and Kircher (2010) protocol, without the initial DNA fragmentation step, and using MinElute PCR purification columns (Qiagen) for all cleanup steps. Modifications to the blunt-end repair step included not using Buffer Tango and ATP, but the addition of 0.1mg/mL BSA and 1x T4 DNA Ligase buffer (New England Biolabs). An additional enzyme inactivation step of 75C for ten minutes was used at the end of the blunt-end repair reaction, instead of using purification columns. The adapter ligation step was modified by the removal of T4 DNA Ligase buffer and the addition of 0.5mM ATP. The final volume for this step was 100μL (including 70μL from the blunt-end repair step). The adapter fill-in step was not modified, but the library was eluted in 30μL of EB buffer in the subsequent purification.

The index PCR, in which the barcodes are incorporated into the adapter, was conducted using *Pfu* Turbo Cx enzyme (Agilent). *Pfu* Turbo Cx was chosen for its ability to copy through uracils (for assessment of misincorporations), not bias against

Sample	Material	Age class	Age (yrs)	MtHg*	Locality	Latitude	Longitude	Source	Extraction lab.
Mo1	Whole specimen	Modern	8	3	Anadyr, Chukotka, Russia	64.73	177.48	Field collected	RHUL
Mo2	Whole specimen	Modern	9	1	Titaluk River, Alaska, USA	69.70	-155.20	Field collected	RHUL
Mu1	Hind leg	Museum	62	2	Repulse Bay, Nunavut, Canada	66.53	-86.23	CNC	RHUL
Mu2	Hind leg	Museum	137	1	Novaya Zemlya, Arkhangelsk, Russia	72.87	53.38	NRM	NRM
An1	Elytron	Ancient	28100	1	Goldbottom Creek, Yukon, Canada	63.97	-138.97	Permafrost	NRM
An2	Pronotum	Ancient	LP	1	Titaluk River, Alaska, USA	69.70	-155.20	Permafrost	RHUL

Table 2.3 Locality and age details of the six samples of *A. alpina* investigated using NGS. LP: Late Pleistocene (>11,700 cal. yrs BP), MtHg: Mitochondrial haplogroup.

*Based on assignment in chapter six. Extraction lab: laboratory in which DNA extraction was performed.

length (for assessment of fragment length), and not bias toward G-C rich sequences (for taxonomic assessment) (Dabney & Meyer 2012). Each library was tagged with a 6bp barcode developed at RHUL, which were modified from those of Meyer and Kircher (2010). The index PCR step was based on the Dabney and Meyer (2012) protocol, with the addition of 0.4mg/mL BSA. Three index PCR reactions were conducted per library, in order to reduce the number of PCR duplicates (Avila-Arcos *et al*. 2011). Each reaction used 10µL of library and had a final volume of 50µL. Cycling conditions followed Dabney and Meyer (2012), for 20 cycles. The amplified library was eluted in 60µL of EB buffer during the final purification step.

Amplified DNA libraries were quantified using a spectrophotometer and gel electrophoresis. DNA quantity was determined by averaging four Nanodrop (ND-1000) readings. Average sequence length and distribution was estimated using a 2% agarose gel. DNA concentration (in nM) was calculated using the equation:

$$\frac{\text{DNA quantity [ng/µL]} * 10^6}{649 * \text{Average sequence length [bps]}}$$

Amplified libraries were diluted with EB buffer to 10nM and pooled. Pooled libraries were sequenced on the Illumina HiSeq-2000 platform at the Exeter Sequencing Facility (University of Exeter, UK), using a single lane of 2 x 100 cycles on a paired-end flow cell, following the manufacturer's instructions. The paired-end sequencing output consisted of two *FASTQ* files (reads one, two), which sequentially list the first 100bp of the 5' (read one) and 3' (read two) ends of each sequenced DNA fragment. Quality scores were output in Illumina 1.5 format.

2.5.3. Quality Control and Preparation for Alignment

Prior to alignment and contig assembly, files underwent a series of quality control and other preparatory procedures (fig. 2.2). The *FASTQ* files were combined and quality filtered using the standard (Blankenberg *et al*. 2010) and FASTX (http://hannonlab.cshl.edu/fastx_toolkit/index.html) toolkits (all v1.0.0 unless otherwise stated), respectively, on the Galaxy server (Goecks *et al*. 2010). Quality

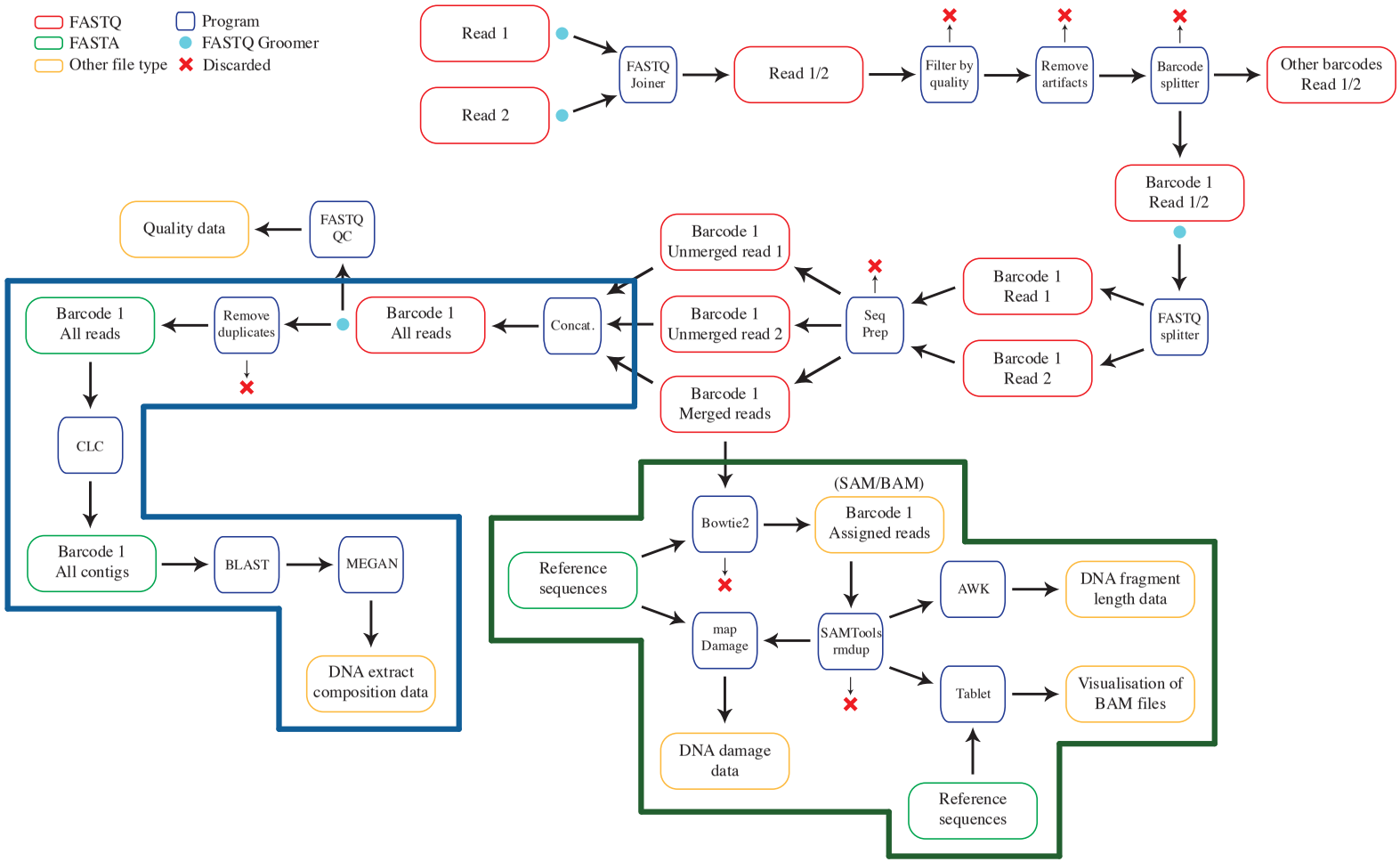


Figure 2.2 Customised workflow for analysis of Illumina paired-end DNA sequence data. Data were quality filtered, binned by barcode, and merged. They were then analysed for DNA fragment length and damage (encircled in green, see section 4.3.1), as well as DNA extract composition (encircled in blue, see section 5.3).

scores were converted to Sanger using the FASTQ Groomer v1.0.4 tool. The files were then concatenated using FASTQ Joiner. The concatenated file was quality filtered using 'filter by quality' with reads consisting of >5% of bases with a quality score of ≤15 being removed [following Kircher *et al*. (2011)]. The file was then further filtered to remove sequencing artifacts, using the 'remove sequencing artifacts' tool.

Based on barcode, the filtered *FASTQ* file was split using a modified version of the FASTX Barcode Splitter tool. This modification allows the tool to search for the barcode at the end of the sequence/lane description, rather than at either the 5' or 3' end of the sequence (J.A. Thomas, pers. comm.). Experimentation showed that allowing a single mismatch when determining barcodes was optimal for this dataset (see section 4.5.1), and so these datasets were used for further downstream analyses. Datasets were split into read one and read two files using FASTQ Splitter. Reads from both files were merged using 'SeqPrep' (https://github.com/jstjohn/SeqPrep), which combined the two reads if overlap was detected, and was also used to remove adapter sequence. Parameters in SeqPrep were set to 15 (quality score cut-off for mismatches), 10 (minimum length of merged read), 10 (minimum overlap to merge reads), AGATCGGAAGAGCACACGTC (read one adapter), and AGATCGGAAGAGCGTCGTGT (read two adapter). SeqPrep output consisted of three *FASTQ* files (merged reads, unmerged read one, unmerged read two) per barcode dataset, with between 87.2 and 99.7% of filtered reads being merged.

2.6. References

Avila-Arcos MC, Cappellini E, Romero-Navarro JA, *et al*. (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports* **1**, 74.

Ball GE, Currie DC (1997) Ground Beetles (Coleoptera: Trachypachidae and Carabidae) of the Yukon: Geographical Distribution, Ecological Aspects, and Origin of the Extant Fauna. In: *Insects of the Yukon* (eds. Danks HV, Downes JA), pp. 445-489. Biological Survey of Canada (Terrestrial Arthropods), Ottawa.

Blankenberg D, Gordon A, Von Kuster G, *et al*. (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-1785.

Castalanelli MA, Severtson DL, Brumley CJ, *et al*. (2010) A rapid non-destructive DNA extraction method for insects and other arthropods. *Journal of Asia-Pacific Entomology* **13**, 243–248.

Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* **289**, 1139-1139.

Dabney J, Meyer M (2012) Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87-94.

Elias SA (1994) *Quaternary Insects and Their Environments* Smithsonian Institute, USA.

Elias SA (2010) *Advances in Quaternary Entomology* Elsevier, Oxford, UK.

Gilbert MTP, Bandelt HJ, Hofreiter M, Barnes I (2005) Assessing ancient DNA studies. *Trends in Ecology and Evolution* **20**, 541-544.

Gilbert MTP, Binladen J, Miller W, *et al*. (2007a) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research* **35**, 1-10.

Gilbert MTP, Moore W, Melchior L, Worobey M (2007b) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One* **2**, e272.

Goecks J, Nekrutenko A, Taylor J, Team G (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**, 1-13.

Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001) Ancient DNA. *Nature Reviews Genetics* **2**, 353-359.

Kircher M, Heyn P, Kelso J (2011) Addressing challenges in the production and analysis of illumina sequencing data. *Bmc Genomics* **12**, 382.

Lindroth CH (1968) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 5. *Opuscula Entomologica, Supplementum* **33**, 649-944.

Luff ML (2007) *Handbooks for the identification of British Insects: The Carabidae (ground beetles) of Britain and Ireland*, 2nd edn. Royal Entomological Society, UK.

Mack L (2008) *Investigating mitochondrial DNA phylogenies of Arctic and European beetle taxa* MSc thesis, Mainz University.

Mandrioli M (2008) Insect collections and DNA analyses: how to manage collections? *Museum Management and Curatorship* **23**, 193 -199.

Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010**, pdb prot5448.

Rambaut A (2002) Se-Al (Sequence Alignment Editor).

Reiss RA (2006) Ancient DNA from ice age insects: proceed with caution. *Quaternary Science Reviews* **25**, 1877-1893.

Rohland N, Hofreiter M (2007) Comparison and optimization of ancient DNA extraction. *Biotechniques* **42**, 343-352.

Rychlik W, Rychlik P (2005) Oligo: Primer Analysis Software. Molecular Biology Insights, Inc., Cascade, Colorado, USA.

Thomsen PF, Elias S, Gilbert MT, *et al*. (2009) Non-destructive sampling of ancient insect DNA. *PLoS One* **4**, e5048.

Wandeler P, Hoeck PE, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution* **22**, 634-642.

Chapter 3. An assessment of DNA preservation in museum and ancient specimens of two ground beetles, *Amara alpina* and *Pterostichus brevicornis* (Coleoptera: Carabidae)

3.1. Abstract

Studies have demonstrated that DNA is preserved in dry-stored museum and ancient permafrost-preserved beetle specimens. Both of these sources represent potentially large untapped genetic resources. An assessment of their potential utility, through investigation into the preservation of DNA in these remains, would provide valuable data for researchers who are keen to utilise these specimens. Here, using the ground beetle species *Amara alpina* and *Pterostichus brevicornis*, a Sanger sequencing based approach was employed to examine the preservation of mitochondrial and multi-copy nuclear DNA in museum and ancient beetle specimens. In addition, this technique allowed the rate of ancient beetle specimen misidentification to be assessed. It was found that DNA could be recovered from >95% of museum specimens, with the age of specimen affecting both the number of successful amplifications and maximum mitochondrial fragment length retrieved. DNA was recovered from ~45% of ancient *A. alpina* specimens, but could not be recovered from ancient *P. brevicornis* specimens. The number of successful amplifications for ancient *A. alpina* specimens was not affected by specimen age or locality. DNA misincorporation rates for both museum and ancient specimens were comparable to previous studies, with nuclear DNA exhibiting a greater misincorporation rate in ancient specimens. These results suggest that there is great genetic potential in dry-stored museum specimens, but ancient permafrost-preserved specimen potential may be more reliant on the species under study. Encouragingly, ancient specimen misidentification was not detected.

3.2. Introduction

It has been shown that DNA is recoverable from degraded beetle remains, both from old (>50yrs) dry-stored museum (Gilbert *et al*. 2007b; Goldstein & Desalle 2003; Thomsen *et al*. 2009) and ancient permafrost-preserved specimens (Mack 2008;

Thomsen *et al*. 2009). Both of these DNA sources represent potentially vast untapped genetic resources that are not only easily accessible, but can also add a temporal aspect to a suite of biological questions. However, the potential of these remains for large-scale genetic analyses has yet to be assessed. In addition, ancient specimens are usually recovered as isolated sclerites (fig. 2.1). Based on exoskeletal morphology, it is often possible to identify these specimens to the species level (Coope 2004; Elias 2010). Genetic testing of these remains would provide an independent measure of whether or how often specimens are misidentified. Such uncertainty could then potentially be incorporated into palaeoclimatic analyses, such as the mutual climatic range method, that rely on accurate species-level identification (Elias 2010). This study provided an assessment of DNA preservation in museum and permafrost-preserved specimens of two closely related beetle species, *Amara alpina* and *Pterostichus brevicornis*, as well as assessment of the rate of ancient specimen misidentification.

Hundreds of museum and ancient DNA studies have utilised simplex PCR methods, whereby one primer set is used per reaction (hereafter referred to as PCR), which has been the technique of choice for the amplification of degraded beetle DNA to a concentration suitable for Sanger sequencing [e.g. Goldstein and Desalle (2003), King *et al*. (2009), Reiss (2006), Thomsen *et al*. (2009)]. Although NGS technologies are rapidly gaining ground, the Sanger sequencing approach is still widely used for analyses, as well as being the method for screening specimens prior to NGS analysis, to ensure correct specimen identification and the presence of endogenous aDNA (Gilbert *et al*. 2008; Green *et al*. 2006). A Sanger sequencing based approach was therefore highly suitable for this type of study.

The overarching aims of this study were to provide background data on mitochondrial and nuclear DNA preservation in beetle remains from dry-stored museum and ancient permafrost-preserved sources, and assessing the ancient specimen misidentification rate, as inferred from Sanger sequencing. Specifically, this included assessing: 1) the proportion of specimens from which DNA could be amplified, 2) the number of DNA fragments that could be amplified (amplification success) and whether this was influenced by specimen age, 3) how age affects the maximum amplifiable mtDNA

fragment length from museum specimens, 4) if locality is related to DNA recovery in ancient specimens, 5) the observed misincorporation rates in both types of remains, and 6) the rate of ancient specimen misidentification.

3.3. Materials and Methods

3.3.1. Specimen and Sequence Data Selection

A total of 420 museum and ancient specimens from the two study species were used in this study. These included a total of 9 modern, 213 museum, and 198 ancient specimens, respectively. However, because of the small sample sizes of the modern datasets (*A. alpina*: n=6, *P. brevicornis*: n=3), data from modern and museum specimens were combined and are referred to as museum in the remainder of this chapter. All collected Sanger sequence data that had passed the quality control of section 2.4.4 were used (table 2.2), unless otherwise stated. All fragment lengths stated include primer sequence.

3.3.2. Proportion of Specimens with Amplifiable DNA

Specimens were considered to yield mtDNA or nuDNA if one or more fragments had been recovered for each data type. All specimens that failed to yield any DNA were tested with a minimum of three or two primer sets for mtDNA and nuDNA, respectively. These primer sets were also tested on specimens from taxa that are congeneric (*A. glacialis* for *A. alpina*) or consubgeneric (*P. ventricosus*, *P. pinguedineus* for *P. brevicornis*) to the study taxa, in order to reduce the likelihood of any amplification failures being due to specimen misidentification. One ancient specimen of *A. alpina*, which did yield mtDNA, was excluded, due to exhaustion of the extract prior to testing for the presence of nuDNA. All mtDNA and 36.4% of nuDNA PCR products were sequenced and verified as authentic. The non-sequenced PCR products were amplified with primer sets that showed no evidence of contaminant co-amplification, as indicated by data from PCR products that were sequenced.

3.3.3. Amplification Success of mtDNA

Amplification success was defined by the number of DNA fragments recovered for each specimen from independent PCR reactions. For *A. alpina*, eight PCR primer sets (COI sets 1, 3, 5, 8, 10, 12, 13, 15; appendix 2b) were employed, which had products of between 124 and 196bps. For *P. brevicornis*, six PCR primer sets (COI sets 17, 20, 25; COII sets 27, 29, 32; appendix 2b) were used, which had products of between 152 and 176bps. For museum specimens, it was often possible to amplify longer fragments based on different primer set combinations. Therefore, if a fragment were attained from a combination that would have amplified four 'fragments', this would have been treated as four fragments retrieved. For museum specimens, age was determined using either direct collection date, or dates inferred from information on specimen labels (appendix 1). Eight specimens could not be dated using these approaches, due to a lack of inferable label data. Dated specimens were binned into 50yr intervals. For ancient specimens, age was based on calibrated dates (table 2.1) and binned into 20ka intervals. 'LP' specimens were classed as undated, and specimens classed as >60ka ranged from >100 to >560ka. To assess if older specimens yielded fewer DNA fragments, two-tailed Kruskal-Wallis tests were employed in SPSS v19.0.0.2. Undated specimens were excluded from these tests, as were museum specimens that did not yield mtDNA.

3.3.4. Maximum Fragment Length Recovered by Age

Museum specimens of *A. alpina* were initially tested with a primer set that amplified a 446bp fragment. Successfully amplified specimens were tested for longer fragments and those that failed to amplify a product were tested for sequentially shorter fragments. Specimens of *P. brevicornis* were initially tested with a primer set that amplified a 302bp fragment and then tested following the aforementioned approach. The three modern specimens of *P. brevicornis* were not included, as these were not tested with the longer primer sets. In order to test for relationships, linear regressions were performed in SPSS. Specimens for which a collection year could not be determined (see section 3.3.3) were excluded from the analysis, as were those which did not yield mtDNA.

3.3.5. aDNA Recovery by Locality

A chi-square goodness-of-fit test was performed to test if specimen locality (listed in table 2.1), and therefore local preservation conditions, markedly affected the aDNA recovery rate in *A. alpina*. MtDNA and nuDNA recovery rates were calculated separately. Localities 7 to 15 were excluded from the calculation, due to small sample size (<10). Expected values were calculated by multiplying the total number of specimens (table 2.1) by the average recovery rate across all six localities (0.419 for mtDNA, 0.407 for nuDNA).

3.3.6. Calculation of Observed Misincorporation Rate

Nucleotide misincorporation rate was broadly investigated in museum and ancient specimens from a subset of PCR product sequences (PPSs). PPSs were used if both DNA strands had been successfully sequenced. Overall misincorporation rate was calculated through the combination of two rates in each PPS. First, the rate of single base ambiguities, which were defined as multiple peaks that prevented base determination during visual inspection of chromatograms. Second, the rate of base mismatches, which was determined from overlapping regions of PPSs from the same individual. Misincorporation rates were calculated from 691 PCR products (table 3.2). Datasets omitted from these calculations included both ITS1 and the *P. brevicornis* COII datasets. These were excluded because of the presence of repeatedly problematic base positions. In the multi-copy ITS1 dataset, this was interpreted as an artifact of intragenomic heterogeneity, which has been observed in other insects (Fairley *et al*. 2005; Parkin & Butlin 2004; Sword *et al*. 2007; Vogler & Desalle 1994). This issue was not detected in the multi-copy 28S dataset, which was included. The *P. brevicornis* COII dataset concerns were attributed to potential mitochondrial heterogeneity or NUMT contamination, and are discussed in more detail in section 7.5.3. Inclusion of these datasets would have overestimated, and therefore biased, the observed misincorporation rate. Standard errors for misincorporation rates were calculated through resampling the data (bootstrapping), using 100,000 permutations, in *Statistics 101* (http://www.statistics101.net/). Significance testing between pairs of misincorporation datasets was performed using two-tailed approximate permutation tests (APTs), using 100,000 permutations. Each test was conducted five times and an

average probability value recorded. Modified false discovery rate corrections (Benjamini & Yekutieli 2001) were applied where appropriate. APTs were performed using a custom built script in *Statistics 101* (appendix 3). For each dataset, the proportions of the six possible nucleotide substitution combinations (NSCs; C-T/T-C, C-A/A-C, etc.) were calculated, through assessment of the two nucleotides that occurred at each mismatch site or contributed to ambiguous chromatogram peaks. The combination was classed as ambiguous if three or more peaks were observed. Due to the nature of the dataset, it was not possible to assess which of the two bases represented the misincorporation in each combination. Chi-square goodness-of-fit tests were applied to test if some NSCs were more prevalent than others. Ambiguous combinations were excluded, as was the museum *P. brevicornis* nuDNA dataset, due to small sample size. Expected values were calculated by assuming equal proportions of NSCs per dataset.

3.4. Results

3.4.1. Proportion of Specimens with Amplifiable DNA

In both species, >95% of museum specimens yielded both mtDNA and nuDNA (table 3.1). The four specimens that did not yield any DNA (*A. alpina*: n=3, *P. brevicornis*: n=1) were all from the CNC and had been collected from the same locality, in the same year, by the same collector, and represent all the tested specimens from this locality. The four specimens that only yielded nuDNA (two from each species) were from different localities and collected between 1875 and 1980. In ancient specimens of *A. alpina*, 25.9% yielded both mtDNA and nuDNA, whereas 54.4% yielded neither. 10.9% yielded only mtDNA, with a similar proportion yielding only nuDNA (8.8%). Misidentification was not detected in the ancient *A. alpina* specimens that yielded beetle DNA (n=68; table 3.1). However, aDNA could not be yielded from ancient specimens of *P. brevicornis*, even though the use of the same primer sets on consubgeneric taxa resulted in successful amplification.

Species	Age class	N	Mt-Nu	Prop.	Mt	Prop.	Nu	Prop.	None	Prop.
Amara alpina	Museum	139	134	96.40	0	0.00	2	1.44	3	2.16
Pterostichus brevicornis	Museum	83	80	96.39	0	0.00	2	2.41	1	1.20
Amara alpina	Ancient	*148	38	25.85	16	10.88	13	8.84	80	54.42
Pterostichus brevicornis	Ancient	50	0	0.00	0	0.00	0	0.00	50	100.00
		420	252	60.00	16	3.81	17	4.05	134	31.90

Table 3.1 Summary of specimens used in this study, their age class, and the proportion of specimens from which mtDNA and nuDNA were recovered. Mt-Nu: mitochondrial and nuclear DNA recovered, Mt: mitochondrial DNA only recovered, Nu: nuclear DNA only recovered, None: no endogenous DNA recovered. Prop: Proportion of specimens (%). *One specimen was excluded for calculations of DNA recovery (see section 3.3.2).

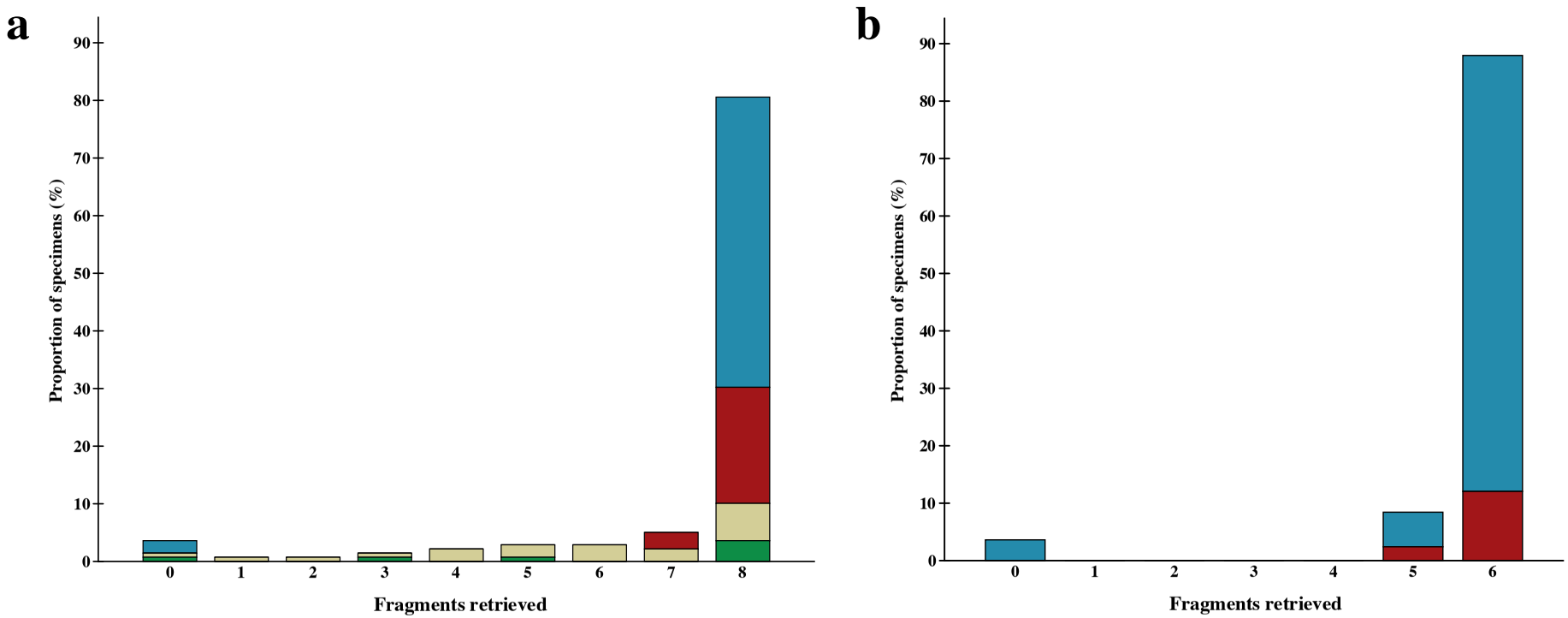


Figure 3.1 Amplification success for museum specimens of *A. alpina* and *P. brevicornis*, based on the number of mtDNA fragments retrieved. Colours indicate specimen age; blue: 1951-2004, red: 1901-1950, off-white: 1851-1900, green: no date. a: *A. alpina*, b: *P. brevicornis*.

3.4.2. Amplification Success of mtDNA

All fragments were successfully amplified from 80.6% and 88.0% of museum specimens of *A. alpina* and *P. brevicornis*, respectively (figs. 3.1a, b). In *A. alpina*, the remaining 15.8% of specimens that amplified only some fragments consisted almost exclusively of 19th Century specimens. The relationship between specimen age and amplification success is highly significant (Two-tailed Kruskal-Wallis test: χ^2=58.995, d.f.=2, n_1=25, n_2=31, n_3=70, p<0.001). In *P. brevicornis*, the remaining 8.4% of specimens amplified five of the six fragments. In this case, there was a non-significant relationship between specimen age and amplification success (Two tailed Kruskal-Wallis test: χ^2=1.094, d.f.=1, n_1=12, n_2=68, p=0.296). In the 37.2% of ancient *A. alpina* specimens that yielded mtDNA, a bimodal distribution was observed, whereby the majority of specimens yielded either one to two fragments, or seven to eight (fig. 3.2). There was no significant relationship between specimen age and the number of fragments that could be retrieved (Two-tailed Kruskal-Wallis test: χ^2=7.042, d.f.=3, n_1=5, n_2=17, n_3=89, n_4=24, p=0.071).

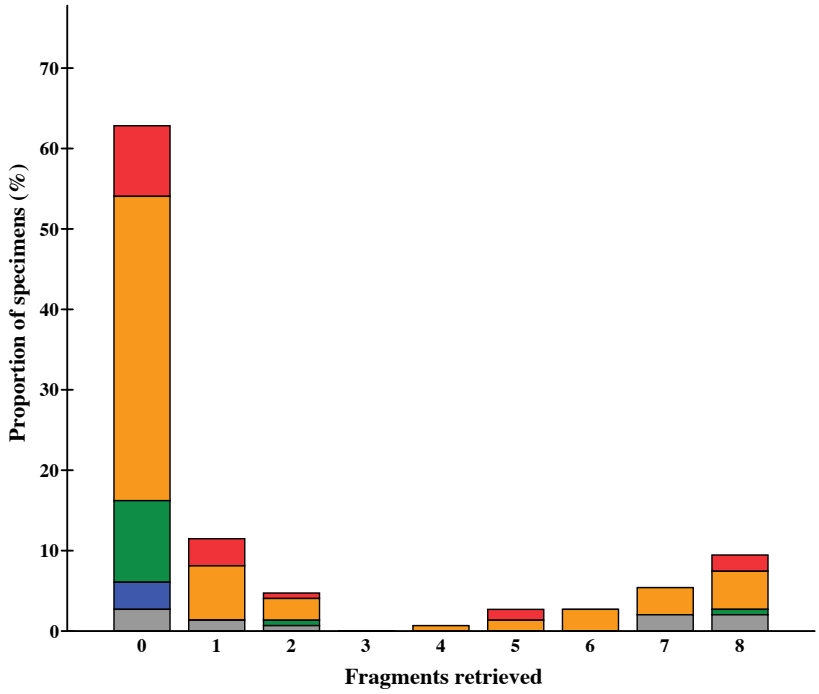


Figure 3.2 Amplification success for ancient permafrost-preserved *A. alpina* specimens, based on the number of mtDNA fragments retrieved. Colours indicate specimen age; red: 1-20ka, orange: 21-40ka, green: 41-60ka, blue: >61ka (includes specimens of >100ka and >560ka), grey: no date [includes LP (Late Pleistocene) specimens, which could not be designated a time bin].

3.4.3. Maximum Fragment Length Recovered by Age

In both species, age predicts a linear relationship with the maximum fragment length that can be retrieved from museum specimens (fig. 3.3a,b). In *A. alpina*, a highly significant relationship is observed (Linear regression: R^2=0.441, F=98.612,

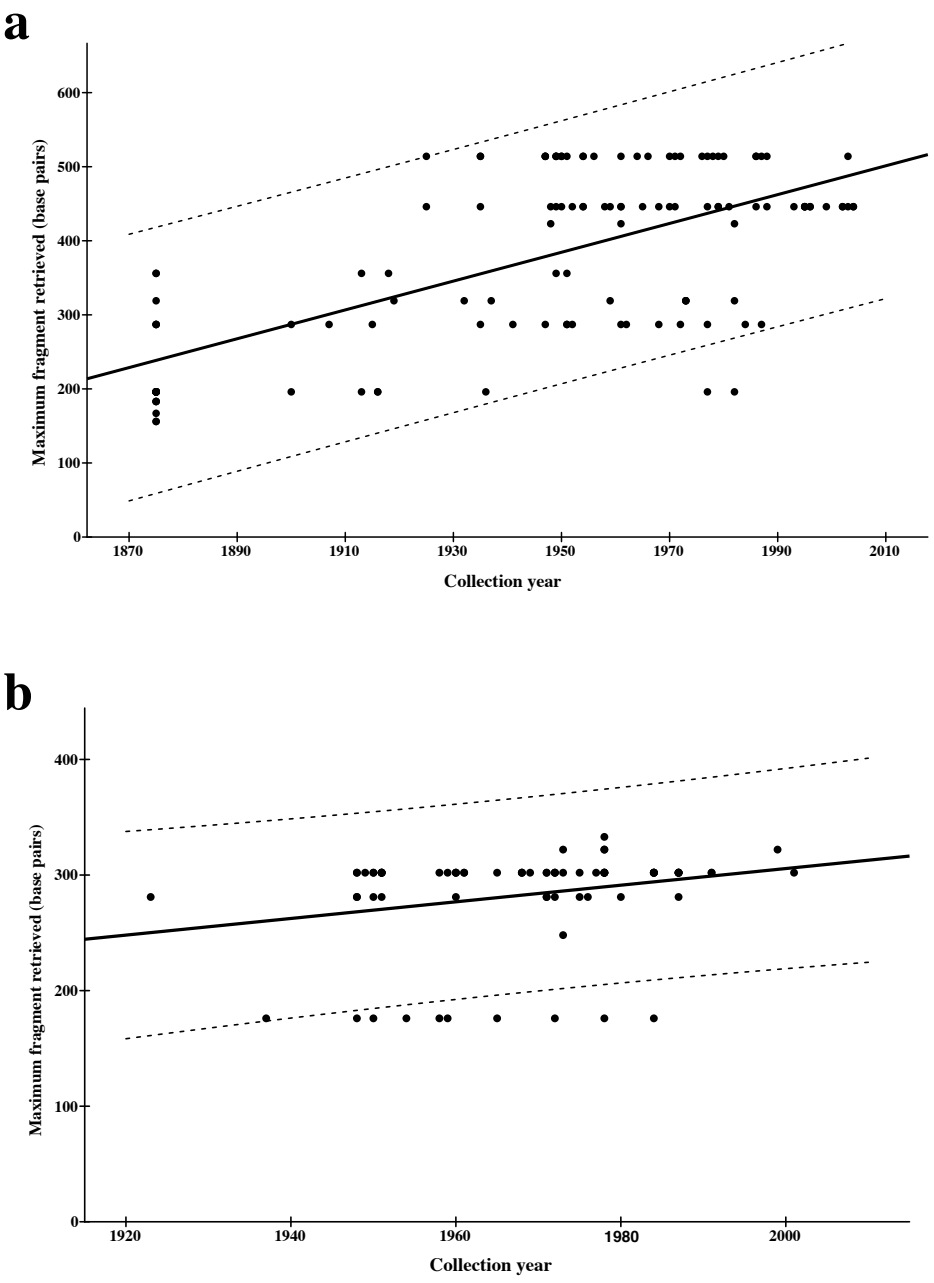
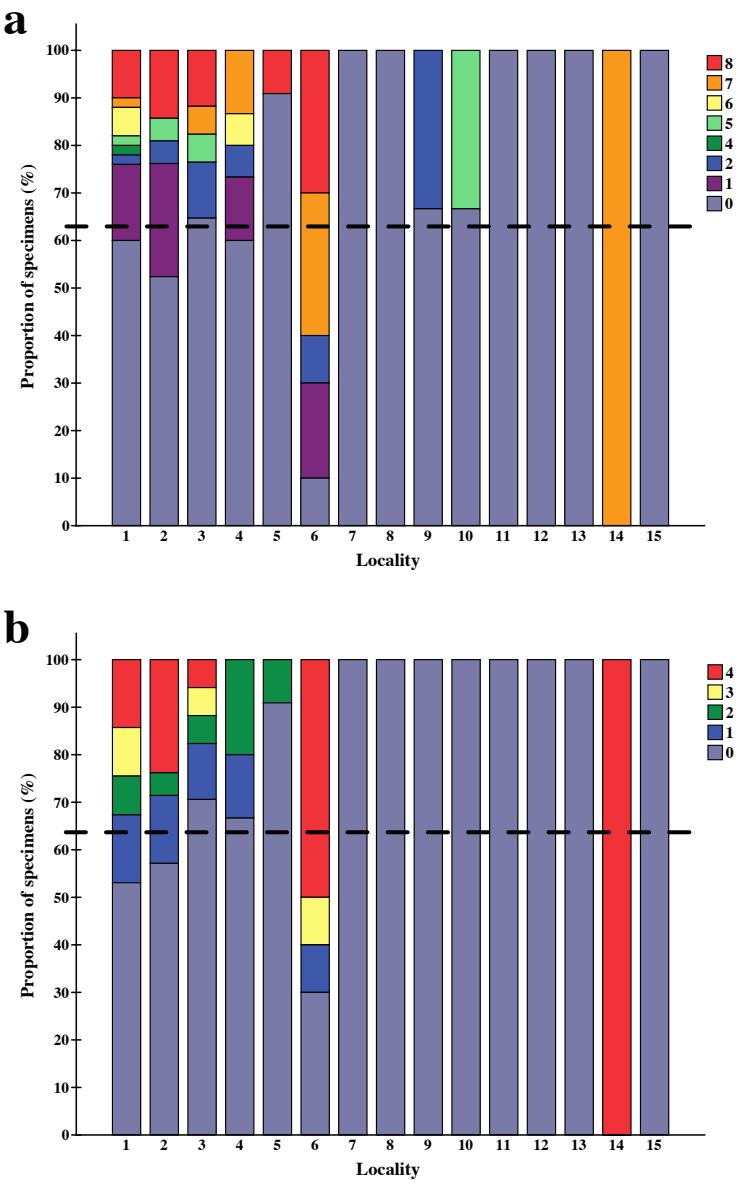


Figure 3.3 Maximum mtDNA fragment length retrieved from museum specimens by age (collection year). a: *A. alpina* (Linear regression: $y = 1.946x - 3410.417$; F = 98.612, d.f. =1,125, $p<0.001$; R^2 = 0.441), b: *P. brevicornis* (Linear regression: $y = 0.72x - 1134.048$; F = 5.155, d.f. = 1,75, p=0.026; R^2 = 0.064). Dotted lines are 95% confidence intervals. Specimens that failed to yield any DNA have been omitted.

d.f.=1,125, p<0.001), which predicts that shorter fragments (100bps) should still be recoverable from specimens of earliest 19th Century (1804) age. In *P. brevicornis*, a significant relationship is observed (Linear regression: R^2=0.064, F=5.155, d.f.=1,75, p=0.026) , which predicts that shorter fragments should still be recoverable from specimens collected in the early 18th Century (1713).

3.4.4. aDNA Recovery by Locality

The aDNA recovery rate for mtDNA and nuDNA was not significantly affected by local preservation conditions, as inferred from the six best-sampled localities (fig. 3.4a,b; table 2.1; Chi-square goodness-of-fit: mtDNA: χ^2=8.737, d.f.=5, p=0.1199,



Previous page: *Figure 3.4* Amplification success for ancient permafrost-preserved *A. alpina* specimens by locality. Localities are arranged by number of specimens tested, in descending order. a: Success of mtDNA recovery, b: Success of nuDNA recovery. Colours indicate the number of fragments retrieved per specimen (a: 0-8, b: 0-4). Dotted lines indicate the overall failure-success rate (a: 63-37%, b: 65-35%). a: n=148, b: n=147. Locality numbers correspond to table 2.1.

nuDNA: χ^2=6.042, d.f.=5, p=0.3022). Ancient mtDNA was recovered from localities aged between 5,900 and 55,000 RC yrs BP, whereas nuDNA was retrieved from localities of between 25,300 and 41,000 RC yrs BP.

3.4.5. Observed Misincorporation Rate

The observed mtDNA base misincorporation rate for museum *A. alpina* and *P. brevicornis* and ancient *A. alpina* specimens was consistent at 0.258 to 0.285% (±0.046 to 0.058% *s.e.m.*; fig. 3.5; table 3.2), with all differences between the three mtDNA misincorporation rates being non-significant (Approximate permutation tests: p=0.741 to 0.873; table 3.3). However, the rate of nuDNA misincorporation was more variable, with ancient *A. alpina* specimens exhibiting the greatest rate (0.670 ±0.166%) and the lowest rate displayed by museum *P. brevicornis* (0.058 ±0.057%),

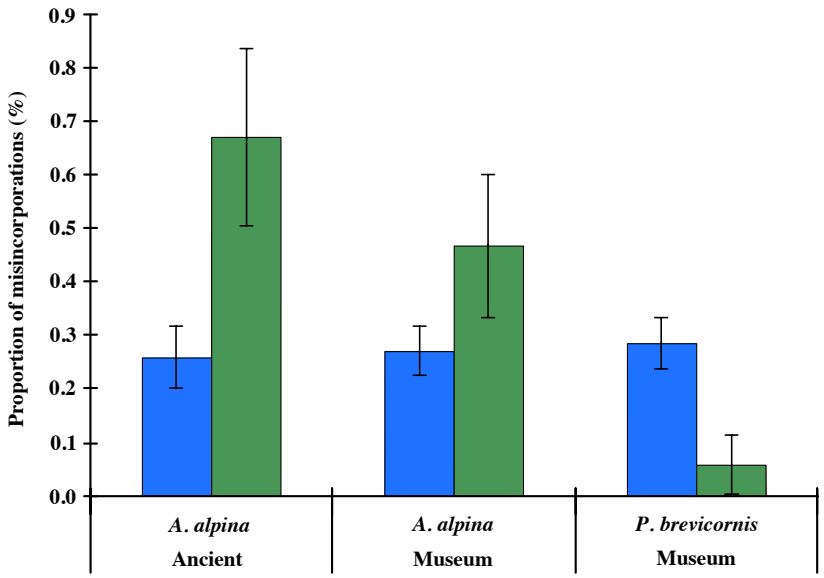


Figure 3.5 Percentage of misincorporated bases in *A. alpina* and *P. brevicornis* specimens of ancient permafrost and dry-stored museum origin. Means are based on OR2 in table 3.2. Error bars represent one standard error, calculated from 100,000 bootstrap replicates. Blue: mtDNA, green: nuDNA.

Species	Age	DNA	S	N	A	R1 (%)	O	M	R2 (%)	OR1 (%)	OR2 (%)
A. alpina	Ancient	mt	184	21902	50	0.228	7856	4	0.051	0.279	0.258
A. alpina	Ancient	nu	43	3597	21	0.584	771	4	0.519	1.103	0.670
A. alpina	Museum	mt	261	52026	78	0.150	35940	17	0.047	0.197	0.270
A. alpina	Museum	nu	39	5937	25	0.421	981	0	0.000	0.421	0.467
P. brevicornis	Museum	mt	138	24558	63	0.257	7795	0	0.000	0.257	0.285
P. brevicornis	Museum	nu	26	3066	2	0.065	105	0	0.000	0.065	0.058
			691	111086	239	0.215	53448	25	0.047	0.262	0.180

Table 3.2 DNA misincorporation rates (OR1, OR2) in ancient and museum specimens of *A. alpina* and museum specimens of *P. brevicornis*. S: PCR product sequences (PPSs) examined, N: nucleotides examined, A: single ambiguities, R1: total rate of single ambiguities [(A/N)*100], O: total length of PPS regions that overlap with other PPSs, allowing for the detection of mismatches, M: mismatches, R2: total rate of mismatches [(M/O)*100], OR1: total overall rate of misincorporation (R1+R2), OR2: average overall rate of misincorporation per PPS.

Var1			Var2			*P*	FDR	Sign.
Taxon	Age	DNA	Taxon	Age	DNA			
A. alpina	Ancient	mt	*A. alpina*	Ancient	nu	0.0059	-	+
A. alpina	Museum	mt	*A. alpina*	Museum	nu	0.1213	-	-
P. brevicornis	Museum	mt	*P. brevicornis*	Museum	nu	0.0414	-	+
A. alpina	Ancient	mt	*A. alpina*	Museum	mt	0.8730	+	-
A. alpina	Ancient	mt	*P. brevicornis*	Museum	mt	0.7412	+	-
A. alpina	Museum	mt	*P. brevicornis*	Museum	mt	0.8408	+	-
A. alpina	Ancient	nu	*A. alpina*	Museum	nu	0.3610	+	-
A. alpina	Ancient	nu	*P. brevicornis*	Museum	nu	0.0051	+	+
A. alpina	Museum	nu	*P. brevicornis*	Museum	nu	0.0140	+	+

Table 3.3 Significant differences between pairs of mtDNA and nuDNA misincorporation rates in ancient and museum specimens of *A. alpina* and museum specimens of *P. brevicornis*. Significance testing (at $p<0.05$ unless otherwise stated) was performed on the OR2 data (table 3.2) using an approximate permutation test. Var: variable, FDR: Modified false discovery rate correction applied ($p<0.027$), Sign: significant result.

Species	Age	DNA	C-T/T-C	G-A/A-G	G-C/C-G	A-T/T-A	C-A/A-C	G-T/T-G	Ambiguous	χ^2	d.f.	*P*
A. alpina	Ancient	mt	0.222	0.278	0.037	0.167	0.111	0.093	0.093	14.061	5	**0.015**
A. alpina	Ancient	nu	0.640	0.280	0.000	0.000	0.040	0.040	0.000	48.680	5	**<0.001**
A. alpina	Museum	mt	0.326	0.284	0.074	0.179	0.053	0.042	0.042	45.418	5	**<0.001**
A. alpina	Museum	nu	0.200	0.200	0.120	0.200	0.240	0.000	0.040	6.000	5	0.306
P. brevicornis	Museum	mt	0.175	0.079	0.000	0.413	0.095	0.127	0.111	42.786	5	**<0.001**
P. brevicornis	Museum	nu	0.000	1.000	0.000	0.000	0.000	0.000	0.000	N/A	N/A	N/A

Table 3.4 Proportion of nucleotide substitution combinations (NSCs) for the datasets presented in table 3.2. Proportions were calculated by combining A and M for each combination and dividing by the total of A+M per dataset. Ambiguous combinations were single base ambiguities composed of more than two peaks. Expected values were calculated by assuming an equal distribution of the NSCs. Ambiguous combinations were excluded from the chi square calculation. Significant results are in bold.

the latter of which was significantly lower than the rate for museum and ancient *A. alpina* (Approximate permutation tests: p=0.014, p=0.005, respectively). The nuDNA misincorporation rate was greater than the mtDNA rate in museum and ancient *A. alpina*, with the difference being significant in ancient specimens (Approximate permutation tests; ancient: p=0.006, museum: p=0.121). However, the mtDNA misincorporation rate was significantly greater than the nuDNA rate in museum *P. brevicornis* (p=0.041). The prevalence of the different NSCs varied greatly between datasets (table 3.4). The C-T/T-C and G-A/A-G NSCs were more prevalent in all *A. alpina* datasets (Chi-square goodness-of-fit: p<0.001 to 0.015; table 3.4), except the museum nuDNA dataset (p=0.306). However, the A-T/T-A NSC was the most prevalent in the museum *P. brevicornis* mtDNA dataset (p<0.001).

3.5. Discussion

3.5.1. Proportion of Specimens with Amplifiable DNA

A very high proportion (96.4%) of museum *A. alpina* and *P. brevicornis* specimens yielded both mtDNA and nuDNA, therefore demonstrating the effectiveness of the methods employed in this study for retrieving museum DNA from dry-stored beetle remains. Other published DNA yields from museum ground beetles are variable, but appear to be related to PCR cycle number and length of target. For example, Goldstein and Desalle (2003) targeted a small mtDNA fragment (73bps), but only used 28 cycles, resulting in 45.7% of specimens yielding DNA. However, Gilbert and colleagues (2007b) targeted longer fragments (220 to 345bps) of both mtDNA and nuDNA, using 40 cycles, of which 71.4 and 78.5% of specimens yielded DNA, respectively. Lastly, Thomsen and colleagues (2009) targeted small to long mtDNA fragments (73 to 204bps), but used 60 cycles, which resulted in all specimens yielding DNA. A possible explanation for why four museum specimens did not yield any DNA, and were from the same locality and collection trip, is that DNA-degrading substances were used to kill these specimens (Dillon *et al*. 1996; Gilbert *et al*. 2007b; Reiss *et al*. 1995).

This is the first reported recovery of nuDNA from a permafrost-preserved invertebrate and the second report of mtDNA recovered from permafrost-preserved invertebrate

macrofossils, after Thomsen and colleagues (2009). In that study, 21.4% of specimens yielded ancient mtDNA, compared to 36.8% in this study. A potential explanation for this disparity is that species-specific primer sets were employed here. Both of these aDNA recovery rates are fairly low in comparison to studies of permafrost-preserved bone, which have rates of between 22.1 and 79.6% (Barnes *et al*. 2002; Barnes *et al*. 2007; Campos *et al*. 2010a; Campos *et al*. 2010b; Shapiro *et al*. 2004). This may be due to the very small size of the material extracted here (<0.5mg), which is two to three orders of magnitude smaller than the 10 to 1000mg of material used in bone-based studies (Barnes *et al*. 2002; Barnes *et al*. 2007; Campos *et al*. 2010a; Campos *et al*. 2010b; Shapiro *et al*. 2004). This would result in fewer template molecules being available for amplification. Alternatively, thermal age analysis indicates that DNA preservation may be an order of magnitude poorer in ancient beetle remains in comparison to bone (King *et al*. 2009). This may be due to apatite, which is absent in insect sclerites, reducing the fragmentation rate of DNA in bone (Lindahl 1993).

The discrepancy in the DNA recovery rate between ancient specimens of the study species is speculated to have been due to differing ecological preferences. *A. alpina* is found in xeric (dry) and mesic (moist) habitats, whereas *P. brevicornis* favours more mesic conditions (Elias & Crocker 2008). Therefore, *P. brevicornis* would be more exposed to a wetter environment, which would result in the rapid degradation of DNA immediately *post mortem* (Hofreiter *et al*. 2001b; Lindahl 1993), before incorporation into the permafrost. Alternatively, the recovery rate disparity may be due to specimen storage prior to analysis. 44% of ancient *P. brevicornis* specimens were stored at room temperature for 26 to 28yrs prior to DNA extraction, which may have contributed to further degradation (Reiss 2006), whereas all *A. alpina* specimens were stored for <10yrs (table 2.1).

It is intriguing that some museum and ancient specimens only yielded nuDNA. This is unexpected under the assumption that mtDNA is preferentially amplified due to greater copy number and therefore greater chance of preservation. However, there are three potential explanations for this observation. First, the targeted nuDNA fragments were shorter (111 to 127bps) than those of mtDNA, which may have been critical for very degraded specimens. Second, there could potentially be very high copy numbers

of the targeted multi-copy nuDNA markers. High copy numbers of these genes have been reported from other insects [>2000 to 8000 per diploid genome (Kumar & Rai 1990; Oishi *et al*. 1985)]. This would also reduce the ratio of mtDNA to nuDNA. Lastly, there may be a DNA preservation bias between mtDNA and nuDNA (Andersen & Mills 2012). Gilbert and colleagues (2007b) also noted that only multi-copy nuDNA amplification was possible in some ground beetle specimens (same marker as this study). Given that these researchers targeted a much larger nuDNA fragment (ranging from 250 to 345bps) compared to this study, it is unlikely that the first explanation is adequate for addressing this issue. In addition, a study of parasitic wasps (Hymenoptera), also targeting the same genetic marker (although a different expansion segment of 28S), found greater amplification success when targeting multi-copy nuDNA (Andersen & Mills 2012). This suggests that the observation of multi-copy nuDNA being more easily amplifiable than mtDNA may not be ground beetle, or even beetle, specific.

3.5.2. Amplification Success of mtDNA

In both *A. alpina* and *P. brevicornis*, the majority of museum specimens amplified all fragments successfully. In *A. alpina*, specimen age was significantly correlated with amplification success, with less successful specimens aged >100yrs old. This may indicate that the concentration of amplifiable DNA decreases with age, which could be due to the continuing occurrence of strand breaks or other forms of damage in these remains (Lindahl 1993). Alternatively, the accumulation of inter-strand crosslinks through time may have reduced amplification opportunity (Hansen *et al*. 2006). Decreasing amplification success with specimen age has also been observed in a number of other museum beetle DNA studies (Gibson *et al*. 2012; Gilbert *et al*. 2007b; Thomsen *et al*. 2009), as well as studies characterising the DNA in old museum specimens of other insect orders (Andersen & Mills 2012; Strange *et al*. 2009; Ugelvig *et al*. 2011; van Houdt *et al*. 2010; Watts *et al*. 2007). The relationship between amplification success and specimen age was not observed in *P. brevicornis*, which is probably due to the lack of specimens aged >100yrs old in the dataset.

The bimodal distribution of ancient specimens that yielded either few or nearly all mtDNA fragments indicates that targeting shorter fragments or increasing the number

of PCR cycles may increase amplification success. Specimen age was not found to affect the likelihood of successful amplification. This suggests that the amount of amplifiable aDNA present in a specimen may not be correlated with age, which would be consistent with the finding that the median DNA fragment length in ancient specimens remains fairly constant regardless of age (Sawyer *et al*. 2012). Conversely, this may also suggest that inter-strand crosslinks are not as prevalent in ancient beetle specimens as other permafrost-preserved organisms (Hansen *et al*. 2006). However, it should be noted that specimens of >60ka (>100 to >560ka) all failed to yield aDNA, but sample size of these specimens was small (n=5). In addition, King and colleagues (2009) noted reduced amplification success with increasing specimen age in non-frozen sediment-preserved beetles.

3.5.3. Maximum Fragment Length Recovered by Age

The linear relationship between maximum fragment length and age in museum specimens suggests that strand breaks are still occurring for decades after the specimen has been desiccated and stored. This result is in agreement with the amplification success of museum beetle DNA (section 3.5.2), and suggests that longer fragments, which may occur at very low but still amplifiable concentrations, undergo strand breakages, or perhaps become unamplifiable due to inter-strand crosslinks (Hansen *et al*. 2006), for decades *post mortem* in museum specimens. The large confidence interval bounds indicate that there is variation in template fragment length between specimens, even though their storage conditions in collections are likely to have been very similar. The disparity between specimens could be due to storage condition prior to deposition in the collection (Gilbert *et al*. 2007b) or stochastic between specimen variation, which has been observed in studies of ancient DNA (Allentoft *et al*. 2012; Briggs *et al*. 2007; Gilbert *et al*. 2007c). A comparable linear relationship between age and maximum retrievable fragment length has also been shown in museum Lepidoptera (butterfly) specimens (Ugelvig *et al*. 2011). Intriguingly, this study targeted nuclear microsatellites, suggesting that this observation may be applicable to mtDNA and nuDNA, as well as different insect orders.

Both the *A. alpina* and *P. brevicornis* datasets predict that mtDNA, of length viable for both PCR and NGS-based approaches (70bps), should be retrievable from museum beetle specimens of late 18th Century age. This is earlier than the current record for the oldest museum beetle DNA recovered [1820 (Thomsen *et al*. 2009)], and encompasses the vast majority of beetle specimens housed in museum collections, thereby demonstrating the great value and potential utility of these specimens as suitable sources for genetic studies. It is suggested that the slope of the trend line shown in the *A. alpina* dataset (fig. 3.3a) is the more realistic. This is because a combination of not testing 19th Century *P. brevicornis* specimens and specimens with long fragments (>350bps) may have biased the result of this dataset through reducing the slope of the trend line (fig. 3.3b). The *A. alpina* dataset does not suffer from either of these biases. A study of museum DNA from 50yr old Orthoptera (grasshopper) specimens found that >800bp fragment could be amplified of both mtDNA and multi-copy nuDNA (Tagliavia *et al*. 2011). Although fragments of this length were not tested here, the results of Tagliavia and colleagues (2011) should be interpreted with caution, as they used universal insect primers and did not state whether the amplified fragments made phylogenetic sense or belonged to the species under study (Cooper & Poinar 2000).

3.5.4. aDNA Recovery by Locality

Local preservation conditions, as measured by locality, have been suggested as an explanation for variation in aDNA preservation (Allentoft *et al*. 2012), especially if factors such as permafrost incorporation time and the proportion of liquid water vary between localities (Willerslev *et al*. 2004). However, a significant association between local preservation conditions and aDNA recovery was not found here. It should be noted that this result could be due to small sample size, as specimens from both Titaluk River localities (two and six; table 2.1) exhibited above average aDNA recovery indicating that this site may warrant further investigation. The recovery of ancient mtDNA and nuDNA from localities 9 (Rock River: 55,000 RC yrs BP) and 5 (Old Crow, loc. 11: 41,000 RC yrs BP), respectively, represents the oldest ancient invertebrate DNA ever recovered from macrofossils for both of these DNA types (King *et al*. 2009; Thomsen *et al*. 2009).

3.5.5. Observed Misincorporation Rate

Observed nucleotide misincorporations are likely to have arisen from three primary sources: damaged bases caused by miscoding lesions (Gilbert *et al*. 2007a; Gilbert *et al*. 2003), polymerase error (Greenwood *et al*. 1999; Hansen *et al*. 2001), and sequencing error. Given the nature of the datasets, it was not possible to determine the relative contribution of each of these error sources and it is therefore recommended that these results be interpreted as an overall misincorporation rate, which may be of practical, but limited biological, use. Additionally, misincorporation rates may have been underestimated, due to only 48.1% of the examined nucleotides being available for mismatch detection (across all datasets, based on O/N in table 3.2).

The C-T/T-C and G-A/A-G NSCs include both type 1 and 2 transitions [*sensu* Hansen *et al*. (2001)], which are the most prevalent types of misincorporation in aDNA studies of vertebrate remains [e.g. (Binladen *et al*. 2006; Gilbert *et al*. 2007a; Gilbert *et al*. 2003; Hofreiter *et al*. 2001a; Stiller *et al*. 2006)] and plant museum DNA (Staats *et al*. 2011). These studies, on specimens from a range of taxonomic groups, ages, and preservation environments, also report a very low incidence of transversions, which includes A-T/T-A, resulting from miscoding lesions. Upon further investigation, the majority of the museum *P. brevicornis* PPSs, which have a high A-T/T-A transversion rate, were found to have been amplified at low annealing temperatures (≤52C). As this can reduce primer specificity, there is a strong possibility of non-target region co-amplification. Given that other problems have arisen from *P. brevicornis* mtDNA (see section 7.5.3), this particular dataset should be interpreted with caution. The small sample size of *P. brevicornis* nuDNA may have confounded the result of a low misincorporation rate in this dataset.

The general prevalence of C-T/T-C and G-A/A-G NSCs in *A. alpina* is indicative of typical DNA misincorporations. The misincorporation rate in mtDNA is consistent between museum and ancient specimens of *A. alpina*, which may indicate that misincorporations are not accumulating through time, as suggested by Sawyer and colleagues (2012). Gilbert and colleagues (2007a) found a misincorporation rate of 0.514% in permafrost-preserved mtDNA of mammoth. Although not directly comparable, this misincorporation rate is similar to the rate found here for the ancient

A. alpina specimens. NuDNA exhibited a higher rate of misincorporation than mtDNA in ancient *A. alpina*, which is in contrast to another study of permafrost-preserved aDNA misincorporation rates that found no significant difference between the rates of mtDNA and nuDNA misincorporation (Binladen *et al*. 2006).

3.5.6. Ancient Specimen Misidentification Rate

Given the large sample size of ancient *A. alpina* specimens that yielded DNA, the complete lack of apparent misidentification is very encouraging. It is unlikely that species-specific primer sets would have prevented any potentially misidentified specimens from yielding DNA, as primer sets were also successfully tested on a congeneric species (*A. glacialis*). Therefore, the assertion made by Quaternary entomologists that even single broken sclerites can be reliably identified (Coope 2004; Elias 2010) is supported by the data for *A. alpina*. Unfortunately, the potential misidentification rate could not be tested in *P. brevicornis* due to all ancient specimens failing to yield aDNA. The possibility that the misidentification of specimens prevented species-specific primer sets from yielding DNA is considered unlikely as these primer sets were also successfully tested on two consubgeneric species.

3.6. Conclusions

This study provides background data related to the preservation of DNA from two sources of degraded beetle remains, dry-stored museum and ancient permafrost, which should be useful to any researchers who are considering utilising these large potential genetic resources for biological investigation. The results indicate that DNA is well preserved in museum specimens with nearly all specimens that were up to 137yrs old yielding DNA. However, amplification success and the longest possible amplifiable fragment both decreased with age. In ancient specimens of *P. brevicornis*, DNA could not be recovered. Conversely, nearly half of ancient *A. alpina* specimens yielded DNA, with amplification success not affected by age or local preservation conditions. Both museum and ancient specimens exhibited DNA misincorporation rates similar to previous studies, although nuclear DNA had a greater rate than mitochondrial DNA in ancient specimens. Overall, these results suggest that there is great potential for

utilising dry-stored museum specimens for genetic analysis, although the potential of ancient permafrost-preserved specimens may be more dependent on the species under study. Lastly, no evidence is found here to support ancient permafrost-preserved specimen misidentification, which reinforces declarations by Quaternary entomologists that broken sclerites can be identified accurately and precisely.

3.7. References

Allentoft ME, Collins M, Harker D, *et al*. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society of London Series B-Biological Sciences* **279**, 4724-4733.

Andersen JC, Mills NJ (2012) DNA extraction from museum specimens of parasitic hymenoptera. *PLoS One* **7**, e45549.

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **295**, 2267-2270.

Barnes I, Shapiro B, Lister A, *et al*. (2007) Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Current Biology* **17**, 1072-1075.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple tesing under dependency. *The Annals of Statistics* **29**, 1165-1188.

Binladen J, Wiuf C, Gilbert MTP, *et al*. (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* **172**, 733-741.

Briggs AW, Stenzel U, Johnson PL, *et al*. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14616-14621.

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

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

Coope GR (2004) Several million years of stability among insect species because of, or in spite of, Ice Age climatic instability? *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 209-214.

Cooper A, Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* **289**, 1139-1139.

Dillon N, Austin AD, Bartowsky E (1996) Comparison of preservation techniques for DNA extraction from hymenopterous insects. *Insect Molecular Biology* **5**, 21-24.

Elias SA (2010) *Advances in Quaternary Entomology* Elsevier, Oxford, UK.

Elias SA, Crocker B (2008) The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota? *Quaternary Science Reviews* **27**, 2473-2483.

Fairley TL, Kilpatrick CW, Conn JE (2005) Intragenomic heterogeneity of internal transcribed spacer rDNA in neotropical malaria vector *Anopheles aquasalis* (Diptera : Culicidae). *Journal of Medical Entomology* **42**, 795-800.

Gibson CM, Kao RH, Blevins KK, Travers PD (2012) Integrative taxonomy for continental-scale terrestrial insect observations. *PLoS One* **7**, e37528.

Gilbert MTP, Binladen J, Miller W, *et al*. (2007a) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research* **35**, 1-10.

Gilbert MTP, Hansen AJ, Willerslev E, *et al*. (2003) Characterization of genetic miscoding lesions caused by postmortem damage. *American Journal of Human Genetics* **72**, 48-61.

Gilbert MTP, Kivisild T, Gronnow B, *et al*. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787-1789.

Gilbert MTP, Moore W, Melchior L, Worobey M (2007b) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One* **2**, e272.

Gilbert MTP, Tomsho LP, Rendulic S, *et al*. (2007c) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**, 1927-1930.

Goldstein PZ, Desalle R (2003) Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. *Molecular Ecology* **12**, 1993-1998.

Green RE, Krause J, Ptak SE, *et al*. (2006) Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330-336.

Greenwood AD, Capelli C, Possnert G, Paabo S (1999) Nuclear DNA sequences from late Pleistocene megafauna. *Molecular Biology and Evolution* **16**, 1466-1473.

Hansen AJ, Mitchell DL, Wiuf C, *et al*. (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* **173**, 1175-1179.

Hansen AJ, Willerslev E, Wiuf C, Mourier T, Arctander P (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular Biology and Evolution* **18**, 262-265.

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

Hofreiter M, Serre D, Poinar HN, Kuch M, Paabo S (2001b) Ancient DNA. *Nature Reviews Genetics* **2**, 353-359.

King GA, Gilbert MTP, Willerslev E, Collins MJ, Kenward H (2009) Recovery of DNA from archaeological insect remains: first results, problems and potential. *Journal of Archaeological Science* **36**, 1179-1183.

Kumar A, Rai KS (1990) Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera, Culicidae). *Hereditas* **113**, 277-289.

Lindahl T (1993) Instability and Decay of the Primary Structure of DNA. *Nature* **362**, 709-715.

Mack L (2008) *Investigating mitochondrial DNA phylogenies of Arctic and European beetle taxa* MSc thesis, Mainz University.

Oishi M, Locke J, Wyatt GR (1985) The ribosomal RNA genes of *Locusta migratoria*: copy number and evidence for underreplication in a polyploid tissue. *Canadian Journal of Biochemistry and Cell Biology* **63**, 1064-1070.

Parkin EJ, Butlin RK (2004) Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Molecular Biology and Evolution* **21**, 1595-1601.

Reiss RA (2006) Ancient DNA from ice age insects: proceed with caution. *Quaternary Science Reviews* **25**, 1877-1893.

Reiss RA, Schwert DP, Ashworth AC (1995) Field preservation of Coleoptera for molecular genetic analyses. *Environmental Entomology* **24**, 716-719.

Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131.

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

Staats M, Cuenca A, Richardson JE, *et al*. (2011) DNA Damage in Plant Herbarium Tissue. *PLoS One* **6**, e28448.

Stiller M, Green RE, Ronan M, *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 of the United States of America* **103**, 13578-13584.

Strange JP, Knoblett J, Griswold T (2009) DNA amplification from pin-mounted bumble bees (*Bombus*) in a museum collection: effects of fragment size and specimen age on successful PCR. *Apidologie* **40**, 134–139.

Sword GA, Senior LB, Gaskin JF, Joern A (2007) Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera : Acrididae). *Systematic Entomology* **32**, 420-428.

Tagliavia M, Massa B, Albanese I, La Farina M (2011) DNA extraction from Orthoptera museum specimens. *Analytical Letters* **44**, 1058–1062.

Thomsen PF, Elias S, Gilbert MT, *et al*. (2009) Non-destructive sampling of ancient insect DNA. *PLoS One* **4**, e5048.

Ugelvig LV, Nielsen PS, Boomsma JJ, Nash DR (2011) Reconstructing eight decades of genetic variation in an isolated Danish population of the large blue butterfly *Maculinea arion*. *BMC Evolutionary Biology* **11**, 201.

van Houdt JK, Breman FC, Virgilio M, de Meyer M (2010) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources* **10**, 459-465.

Vogler AP, Desalle R (1994) Evolution and Phylogenetic Information-Content of the Its-1 Region in the Tiger Beetle *Cicindela Dorsalis*. *Molecular Biology and Evolution* **11**, 393-405.

Watts PC, Thompson DJ, Allen KA, Kemp SJ (2007) How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses? *Journal of Insect Conservation* **11**, 195–198.

Willerslev E, Hansen AJ, Poinar HN (2004) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology and Evolution* **19**, 141-147.

Chapter 4. Bioinformatic considerations for using short, multi-copy reference sequences in Next-Generation Sequencing alignment, and an assessment of DNA preservation in museum and ancient *Amara alpina* (Coleoptera: Carabidae)

4.1. Abstract

Due to the immense quantity of data produced, NGS has revolutionised ancient DNA research by allowing researchers to address previously unattainable biological questions in unprecedented detail. Work has so far focused on plant, pathogen, and vertebrate taxa, but has neglected the ecologically critical and far more speciose invertebrate fauna. This is partly due to the lack of suitable reference genomes for retrieving endogenous aDNA from many invertebrates, including beetles. Here, using Illumina-generated sequences from dry-stored museum and permafrost-preserved specimens of *Amara alpina*, reads were assigned to several short, multi-copy reference sequences. Some bioinformatic considerations of such an approach were explored, through the comparison of two popular aligners: BWA and Bowtie2. Furthermore, the effect of this approach on the proportion of duplicates removed was investigated. It is shown that Bowtie2 may be a better alternative to BWA for assigning reads to short references and that standard duplicate removal procedures may remove natural as well as artificial duplicates, due to what is termed here as reference sequence saturation. In addition, the preservation of DNA in these samples was investigated through an assessment of fragment length distributions and base misincorporation profiles. It is shown that mtDNA tends to have a longer mean fragment length than nuDNA. The proportion of base misincorporation correlates with age, but contrary to other studies, misincorporations are spread throughout the read and not concentrated at termini. This study further demonstrates the potential of DNA from museum and permafrost-preserved specimens for more wide scale and in depth investigation.

4.2. Introduction

NGS has proven to be a powerful tool for the assessment of a suite of questions using aDNA, due to the extensive breadth and depth of data produced, and has even allowed the reconstruction of entire ancient nuclear genomes (Green *et al*. 2010; Miller *et al*. 2008; Rasmussen *et al*. 2011; Rasmussen *et al*. 2010). The shotgun approach (any DNA present in a sample is sequenced) means that endogenous DNA from the organism, as well as contaminant and other exogenous DNA sources, are sequenced. An effective method for retrieving the endogenous DNA sequences is to keep reads that successfully align to a reference genome. However, the suitability of a reference genome is dependent on the evolutionary distance between the taxon under study and the reference taxon. If the taxa are too divergent, then the reads of any particular locus may not be correctly assigned to the reference. This limitation has meant that the vast majority of museum and ancient DNA studies, that utilise NGS, have investigated taxa for which an appropriate nuclear or mitochondrial reference genome is available, such as elephant/mammoth (Gilbert *et al*. 2007; Miller *et al*. 2008; Poinar *et al*. 2006), human (Gilbert *et al*. 2008; Rasmussen *et al*. 2011; Rasmussen *et al*. 2010), horse (Orlando *et al*. 2011; Schubert *et al*. 2012), rat (Rowe *et al*. 2011), and rhinoceros (Willerslev *et al*. 2009). However, the taxonomic distribution of reference genomes is skewed and there may not always be a suitable reference available. This is especially true of beetles, for which there is only a single reference nuclear genome (as of 14/09/2012), even though beetles consist of >350,000 species (Gullan & Cranston 2010) and originated 285Ma (McKenna & Farrell 2009). The available reference genome was for *Tribolium castaneum* (Richards *et al*. 2008). This is problematic for the study taxon (*Amara alpina*), because *T. castaneum* and *A. alpina* belong to different suborders (Polyphaga and Adephaga, respectively), which diverged around 266Ma (McKenna & Farrell 2009). This genome was therefore not deemed suitable as a reference for *A. alpina*, so in order to identify endogenous beetle DNA, shorter suitable reference sequences must be used. During the alignment of reads to reference sequences, it is useful to remove PCR, or artificial, duplicates that were introduced during the index PCR (section 2.5.2). This is because duplicate reads can potentially bias downstream analyses by artificially over-representing a proportion of the reads. The effect of utilising short multi-copy reference sequences on the removal of duplicates was assessed.

Some of the bioinformatic aspects of NGS data analysis of museum and ancient DNA from a taxon without an appropriate reference nuclear genome were therefore explored. This included a comparison of two widely used aligners [the Burrows Wheeler Aligner (BWA) and Bowtie2], as well as exploring the effect of PCR duplicate removal on the number of assigned reads. Suggestions are provided for analysing taxa of this nature and for future areas of investigation.

Additionally, the data from chapter three were built upon, through further examination of the preservation of DNA in dry-stored museum and ancient permafrost-preserved specimens. This enabled the potential of these DNA sources to be further assessed for future study. Recent aDNA studies have shown differences in the level of DNA preservation between mtDNA and nuDNA, with mtDNA exhibiting fewer strand breaks, and therefore a longer mean fragment length, than nuDNA (Allentoft *et al*. 2012; Schwarz *et al*. 2009) Additionally, recent studies have also observed that DNA misincorporations are mainly clustered around read termini (Briggs *et al*. 2007; Orlando *et al*. 2011). However, the conclusions of these studies are based exclusively on vertebrate bone or keratin (hair, nail), and may or may not be applicable to insect remains. The fragment length distributions of mtDNA and nuDNA, as well as the prevalence of misincorporations, and their location within DNA fragments, were assessed in museum and ancient beetle specimens of *A. alpina*.

Six samples of *A. alpina* were used in this study (table 2.3): two dry-stored museum (Mu) and two ancient permafrost-preserved (An) samples, as well as two modern (Mo; <10yrs) samples for comparison purposes. These samples were pooled and shotgun-sequenced on the Illumina HiSeq-2000 platform (section 2.5.2). Sample-specific reads were recovered based on the 6bp barcode sequence (section 2.5.3). However, it is possible to take a less conservative approach and allow mismatches in the barcode, which can recover more reads. The optimal number of barcode mismatches (between zero and two) for the successful recovery of sample-specific reads was assessed.

The aims of this study were therefore three-fold. First, was to explore the optimal number of barcode mismatches to successfully recover sample-specific reads. Second,

was to investigate some of the bioinformatic aspects for the retrieval of endogenous beetle DNA from museum and ancient samples (table 2.3), using several short, multi-copy reference sequences (nuclear: 183 to 1,043bp, mitochondrial: 658 to 16,823bp). This included a comparison of the BWA and Bowtie2 aligners and addressing issues associated with duplicate removal. Third, the preservation of DNA was characterised through comparison of the fragment length distributions of mtDNA and nuDNA. In addition, the distribution of DNA base misincorporations throughout the DNA fragments in these samples, and if a greater rate of misincorporation was observed in older samples, were both assessed.

4.3. Materials and Methods

4.3.1. Comparison of Aligners

Two freely available read mapping aligners were compared for performance with the study dataset: BWA v0.6.0 (Li & Durbin 2009) and Bowtie2 v2.0.0-beta7 (Langmead & Salzberg 2012). Parameters for BWA were set to the recommendations of Schubert and colleagues (2012), with those for Bowtie2 outlined in appendix 4, following parameter optimisation. The two aligners were tested with the merged read datasets only (from section 2.5.3), in both *FASTQ* and *FASTA* format. Reads were aligned to 15 reference sequences, which were downloaded from Genbank (table 4.1, fig. 2.2: section encircled in green). At least one representative sequence was chosen for each locus available for *Amara*, as well as the three available carabid mitogenomes (as of 14/09/2012). A 16S rDNA sequence for *A. alpina* was not used as a reference, due to very short length [44bps (Thomsen *et al*. 2009)]. The same reference sequences were used for all six samples, with the exceptions of the two references originating from this thesis. In these cases, sample specific sequences were used.

Resultant *BAM* files from both BWA and Bowtie2 were sorted and filtered to remove PCR duplicate sequences, using 'rmdup' in the 'SAMTools' v0.1.18 suite (Li *et al*. 2009) and 'MarkDuplicates' in the 'Picard' v1.56 suite (http://picard.sourceforge.net). The optimal combination of aligner and file format (Bowtie2/*FASTQ*) was determined based on maximum number of assigned reads in the least amount of CPU time across all samples. These datasets were used for all downstream analyses. To ensure the

Taxon	Marker(s)	Mt/Nu	Length (bp.)	Genbank Accession	Reference
Amara alpina	COI	Mt	759	N/A	This thesis
Amara alpina	COI	Mt	658	JX259745	Gibson *et al*. (2012)
Amara alpina	COI	Mt	658	JX259747	Gibson *et al*. (2012)
Damaster mirabilissimus	Mitogenome	Mt	16823	NC016469	Wan *et al*. (2012)
Trechini sp.	Mitogenome	Mt	11460	HQ232802	Timmermans *et al*. (2010)
Calosoma sp.	Mitogenome	Mt	16462	NC018339	Song *et al*. (2010)
Amara aulica	COI-tRNA-COII	Mt	862	AY551824	Sanchez-Gea *et al*. (2004)
Amara aenea	COI-tRNA-COII	Mt	1641	FJ173213	Ruiz *et al*. (2009)
Amara alpina	28S	Nu	183	N/A	This thesis
Amara alpina	ITS1	Nu	325	AY004249	Unpublished
Amara aenea	Ef1a	Nu	708	FJ173150	Ruiz *et al*. (2009)
Amara apricaria	wingless	Nu	507	AF398565	Ober (2002)
Amara chalcites	wingless	Nu	420	AB243550	Sasakawa and Kubota (2007)
Amara aulica	18S	Nu	402	GU348075	Raupach *et al*. (2010)
Amara sp.	28S	Nu	1043	AF398694	Ober (2002)

Table 4.1 Details of the 15 reference sequences used to retrieve *A. alpina* sequences from the merged read files. Mt: mitochondrial DNA, Nu: nuclear DNA. Ef1a: Elongation factor-1α.

reasonable assignment of reads to reference sequences, *BAM* files were indexed and visually inspected using 'Tablet' v1.12.09.03 (Milne *et al*. 2010). Due to bias with regard to the proportion of mtDNA and nuDNA reads that were excluded during duplicate removal, all subsequent analyses incorporated both the complete ('all data') and duplicates removed ('no duplicates') datasets. This bias was determined through a two-tailed chi-squared goodness of fit test, with expected values calculated by assuming an equal likelihood of duplicate removal for mtDNA and nuDNA.

4.3.2. Fragment Length Distribution and DNA Damage Estimation

Fragment length information was extracted from aligned reads in the *BAM* files, both before and after duplicate removal, and pooled into two categories for each sample based on mtDNA or nuDNA. Distributions were plotted using a five-point centred moving average in order to smooth the length distributions. Descriptive statistics were calculated and comparisons of fragment length distributions between DNA categories were performed using t-tests in SPSS. Appropriate T-statistics were selected using the results of Levene's test for equality of variances.

DNA damage, in the form of base misincorporations at the first (5') and last (3') 25 nucleotides of the aligned reads in the *BAM* files, was determined using 'mapDamage' v0.3.6 (Ginolhac *et al*. 2011). MapDamage assesses all reads in an alignment and displays the rate of C→T and G→A mismatches and indels at each nucleotide position, but cannot distinguish between damage-induced misincorporations and other sources of mismatch, such as divergence from reference sequences and sequencing errors (Briggs *et al*. 2007; Schubert *et al*. 2012), as well as SNPs, and heterogeneous multi-copy sequences (A. Ginolhac, pers. comm.).

4.4. Results

4.4.1. Sequence Data

A total of 43.3 million paired-end reads were obtained, of which between 7.8 and 9.0 million [7.4 to 8.7 million after filtering (a.f.)] were assigned to each of the modern (Mo) and museum (Mu) samples, based on the barcode incorporated into the adapter

sequence (section 2.5.2). Additionally, 3.5 to 4.3 million (3.1 to 3.6 million a.f.) were assigned to each of the ancient (An) samples (fig. 4.1). Across all six samples, allowing one mismatch in the barcode increased the number of assigned reads by 1.48% [1.31% a.f.], whereas allowing two mismatches further increased the number of assigned reads by 0.23% (0.16% a.f.). Based on allowing a single mismatch in the barcode, 4.84% (4.48% a.f.) of reads were not assigned to a used barcode. Of these reads, 16% [18% a.f., or 0.77% (0.81% a.f.) of all reads] could be assigned to four barcodes used at RHUL (BC2, 4, 9, 10), that were not used during DNA library preparation, with 71% of these reads assigned to BC4 (fig. 4.1).

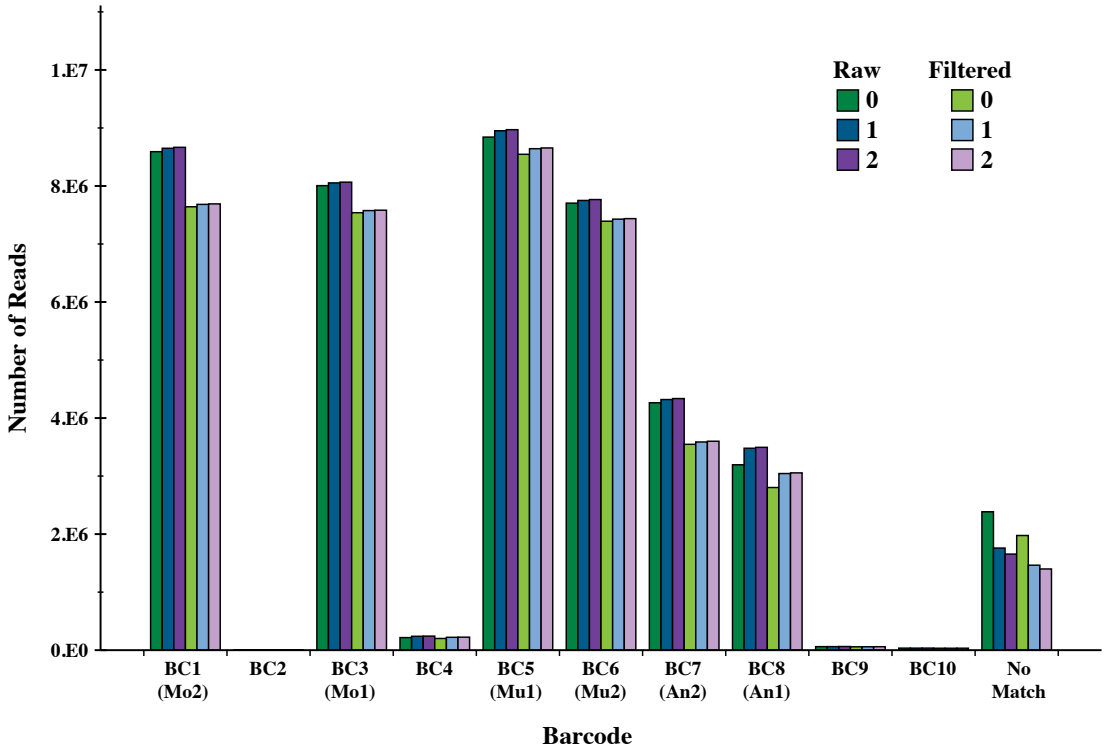


Figure 4.1 Proportion of reads assigned to each barcode. Raw reads are unprocessed. Filtered reads have undergone quality filtering and sequencing artefact removal. Colours indicate how many mismatches were allowed in determining the barcode. Mo: modern, Mu: museum, An: ancient.

4.4.2. Comparison of Aligners and Duplicate Removal Bias

Both read mapping programs, BWA and Bowtie2, assigned comparable numbers of reads to the reference sequences across all six samples (table 4.2). BWA assigned the same number of reads regardless of whether the read file format was in *FASTQ* or *FASTA*, although analyses of *FASTA* files used less CPU time. Bowtie2 assigned more

		Mo1				Mo2			
Aligner	**File format**	**Reads mapped**	**CPU time**	**Unique reads**	**Prop. retained**	**Reads mapped**	**CPU time**	**Unique reads**	**Prop. retained**
BWA	FASTQ	14044	93.78	**2834**	0.202	10975	101.65	3659	0.333
BWA	FASTA	14044	87.63	**2834**	0.202	10975	91.15	3659	0.333
Bowtie2	FASTQ	**16506**	388.39	2658	0.161	**13275**	415.85	**4293**	0.323
Bowtie2	FASTA	16469	367.17	2655	0.161	13170	373.14	4241	0.322

		Mu1				Mu2			
Aligner	**File format**	**Reads mapped**	**CPU time**	**Unique reads**	**Prop. retained**	**Reads mapped**	**CPU time**	**Unique reads**	**Prop. retained**
BWA	FASTQ	17309	106	3468	0.200	11984	82.83	1943	0.162
BWA	FASTA	17309	95.16	3468	0.200	11984	70.24	1943	0.162
Bowtie2	FASTQ	**19470**	417.74	**3910**	0.201	**14273**	311.4	1989	0.139
Bowtie2	FASTA	19426	398.5	3888	0.200	14269	296.39	**1991**	0.140

		An1				An2			
Aligner	**File format**	**Reads mapped**	**CPU time**	**Unique reads**	**Prop. retained**	**Reads mapped**	**CPU time**	**Unique reads**	**Prop. retained**
BWA	FASTQ	488	35.12	364	0.746	264	41.6	216	0.818
BWA	FASTA	488	32.12	364	0.746	264	34.57	216	0.818
Bowtie2	FASTQ	**575**	127.49	**420**	0.730	**325**	145.46	**259**	0.797
Bowtie2	FASTA	574	122.22	**420**	0.732	324	135.39	258	0.796

Table 4.2 Performance comparison of the BWA and Bowtie2 aligners, using input sequence data in the *FASTQ* and *FASTA* file formats. BWA parameters were set to -l 1024 [following Schubert *et al*. (2012)]. Bowtie2 parameters were set to -D 20 -R 3 -N 1 -L 20 -i S,1,0.50 (following optimisation, which is outlined in appendix 4). All other parameters were set to default. BWA CPU time includes both aln and samse steps. Unique reads are those remaining after duplicate removal. Proportion (prop.) retained is the number of unique reads divided by the total number of reads mapped. Bold represents greatest number of reads in each column per dataset.

reads when the read file format was in *FASTQ*, although these analyses used more CPU time. Bowtie2 was found to assign more unique reads than BWA in all cases except for Mo1. However, Bowtie2 analyses took roughly four times as much CPU time as BWA, and a slightly higher proportion of assigned reads were identified as duplicates in Bowtie2 datasets (table 4.2).

The duplicate removal programs, rmdups and MarkDuplicates, each removed the same number of duplicates across all datasets. The mtDNA and nuDNA datasets show a general trend of a larger proportion of reads removed as duplicates in datasets with a larger number of 'all data' reads. For the modern and museum samples, between 67.7 and 86.1% of reads were removed as putative duplicates. However, for the ancient samples, only 20.3 to 27.0% of reads were removed (table 4.2). Further investigation revealed that a significantly disproportionate number of reads removed from the modern and museum samples were those assigned to multi-copy nuDNA reference sequences (Chi-square goodness of fit tests: χ^2=518 to 1825, d.f.=1, all p<0.0001; table 4.3). For the ancient samples, the proportions of mtDNA and nuDNA reads that were removed as duplicates were similar and statistically non-significant (χ^2=0.04 to 0.33, d.f.=1, p=0.567 to 0.845). The number of reads assigned to each of the modern and museum samples was around double the number assigned to each of the ancient samples (fig. 4.1). However, the number of reads assigned to the references was an order of magnitude lower in ancient samples, even after duplicate removal (table 4.2).

4.4.3. Fragment Length Distribution and DNA Damage

Fragment length distributions of mtDNA and nuDNA were plotted for each sample before (all data) and after duplicate removal (duplicates removed; fig. 4.2). The distributions broadly exhibit a normal or lognormal distribution, except for the Mo2 and An2 nuDNA distributions, which have low peaks and long trailing tails. The mean fragment length of the 'all data' datasets range from 79 to 109 and 61 to 101 nucleotides for mtDNA and nuDNA, respectively (table 4.4). Prior to duplicate removal, the nuDNA mean fragment length for all samples is significantly shorter than mtDNA (t-tests: p<0.001). However, for the modern and museum samples, the mean fragment length of the 'duplicates removed' datasets are not as differentiated, ranging from 79 to 113 and 80 to 124 nucleotides for mtDNA and nuDNA

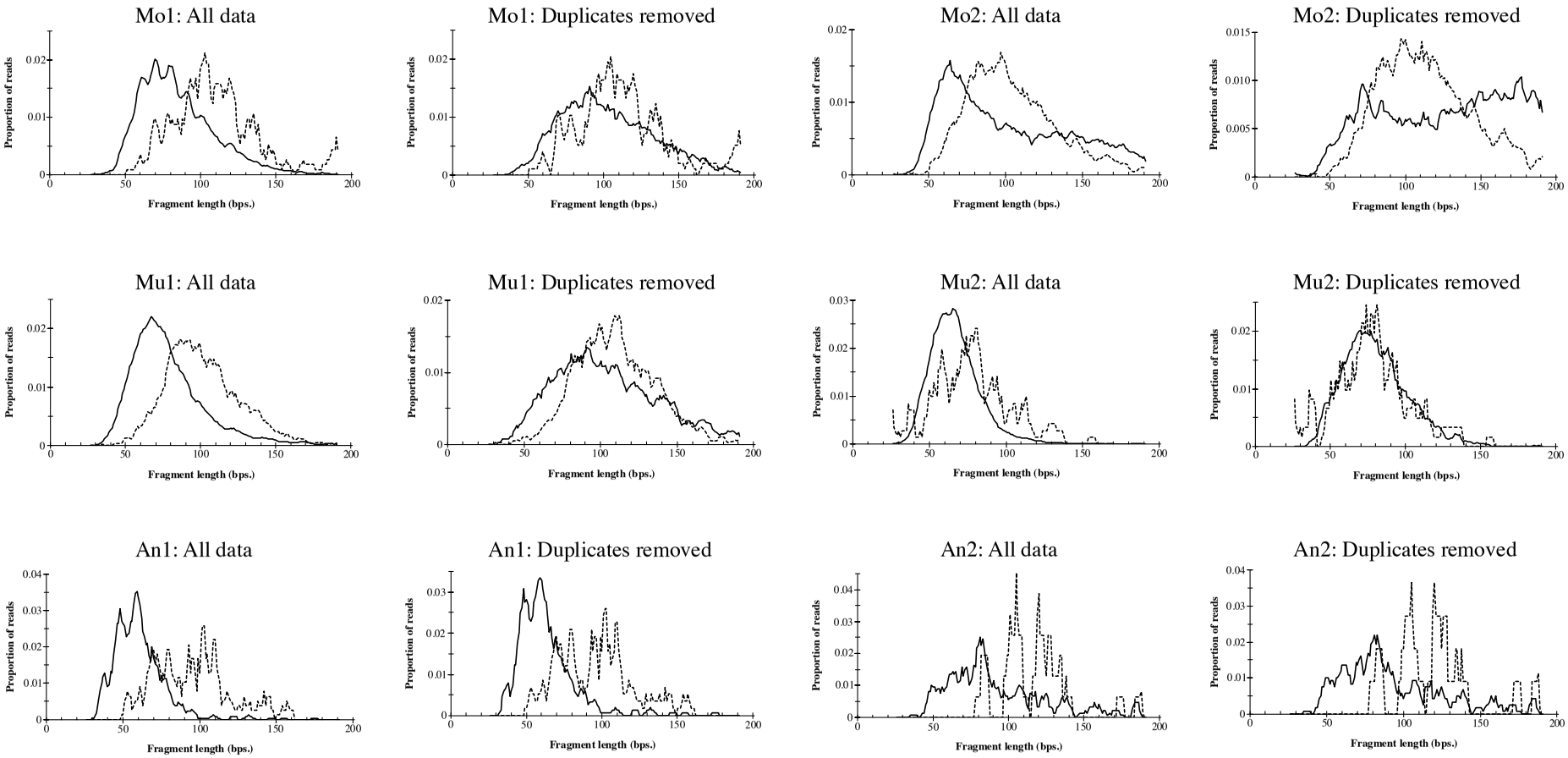


Figure 4.2 Fragment length distributions of *A. alpina* DNA from three age classes. Distributions are shown before (all data) and after (duplicates removed) duplicate removal. Proportion of reads is the number of reads for a given fragment length divided by the total number of reads in that sample. Fitted lines are based on a five-point centred moving average. Solid line: nuclear DNA, dotted line: mitochondrial DNA.

	mtDNA			nuDNA					
Sample	**All data**	**Duplicates removed**	**Proportion retained**	**All data**	**Duplicates removed**	**Proportion retained**	**χ^2**	**d.f.**	***P***
Mo1	227	195	0.859	16279	2463	0.151	696.373	1	**<0.0001**
Mo2	3792	1900	0.501	9483	2393	0.252	518.126	1	**<0.0001**
Mu1	2538	1409	0.555	16932	2501	0.148	1824.642	1	**<0.0001**
Mu2	141	122	0.865	14132	1867	0.132	538.462	1	**<0.0001**
An1	155	115	0.742	420	305	0.726	0.038	1	0.8446
An2	31	22	0.710	294	237	0.806	0.327	1	0.5672

Table 4.3 Bias toward nuDNA during duplicate removal in modern and museum samples of *A. alpina*. A chi-square goodness of fit test was used to determine the significance of the bias. Expected values were calculating by assuming equal proportions of mtDNA and nuDNA were excluded during duplicate removal. Significant results are highlighted in bold.

	All data		All data: mt-nu			Duplicates removed		Duplicates removed: mt-nu		
Sample	**mt**	**nu**	**T**	**d.f.**	***P***	**mt**	**nu**	**T**	**d.f.**	***P***
Mo1	109.0 (27.28; 1.81)	84.8 (24.64; 0.19)	14.689	16504.0	**<0.001**	111.3 (27.96; 2.00)	103.0 (30.61; 0.62)	3.967	232.4	**<0.001**
Mo2	103.7 (28.14; 0.46)	100.8 (39.67; 0.41)	4.800	9746.1	**<0.001**	112.9 (29.74; 0.68)	123.7 (41.63; 0.85)	*-9.846*	*4245.6*	***<0.001***
Mu1	101.6 (24.61; 0.49)	78.3 (24.53; 0.19)	44.547	3337.9	**<0.001**	109.5 (25.74; 0.69)	103.4 (33.14; 0.66)	6.359	3531.3	**<0.001**
Mu2	78.7 (24.00; 2.02)	66.7 (16.14; 0.14)	5.898	141.3	**<0.001**	79.0 (24.06; 2.18)	80.2 (21.86; 0.51)	*-0.580*	*1987.0*	*0.562*
An1	95.3 (25.09; 2.02)	60.9 (17.82; 0.87)	15.703	213.9	**<0.001**	95.8 (24.77; 2.31)	62.8 (19.12; 1.10)	12.917	167.8	**<0.001**
An2	116.5 (23.36; 4.20)	92.1 (32.73; 1.91)	5.286	43.5	**<0.001**	118.7 (25.93; 5.53)	92.5 (33.15; 2.15)	3.607	257.0	**<0.001**

Table 4.4 Descriptive statistics and significance testing results of DNA fragment length distributions from three age classes of *A. alpina*. Descriptive statistics are the mean fragment length followed by the standard deviation and standard error in parentheses. Significance testing results are from a two-tailed t-test. Appropriate t-statistics were chosen based on Levene's test of equality of variances. Significant results are in bold. Italics represent cases where nuDNA has the longer mean fragment length. T: T-statistic, mt: mitochondrial DNA, nu: nuclear DNA.

respectively. In half of cases, mean fragment length for nuDNA was now longer than mtDNA, significantly so in one case (Mo2). Removing duplicates increased the mean fragment length in all datasets, indicating that shorter reads had been preferentially removed.

There is variation in DNA misincorporation patterns between the samples, with the modern and museum samples generally exhibiting a lower rate of misincorporation (fig. 4.3). However, Mo1 exhibits higher rates of C→T and G→A misincorporation at the 5' and 3' ends, respectively. Mu2 (137yrs) exhibits a higher rate of misincorporation (~5%) than Mu1 (62yrs; ~2%). In the ancient samples and Mu2, misincorporations are spread throughout the DNA fragment, at a rate of 5 to 10% in ancient samples (fig. 4.3). Duplicate removal minimally impacted on the DNA

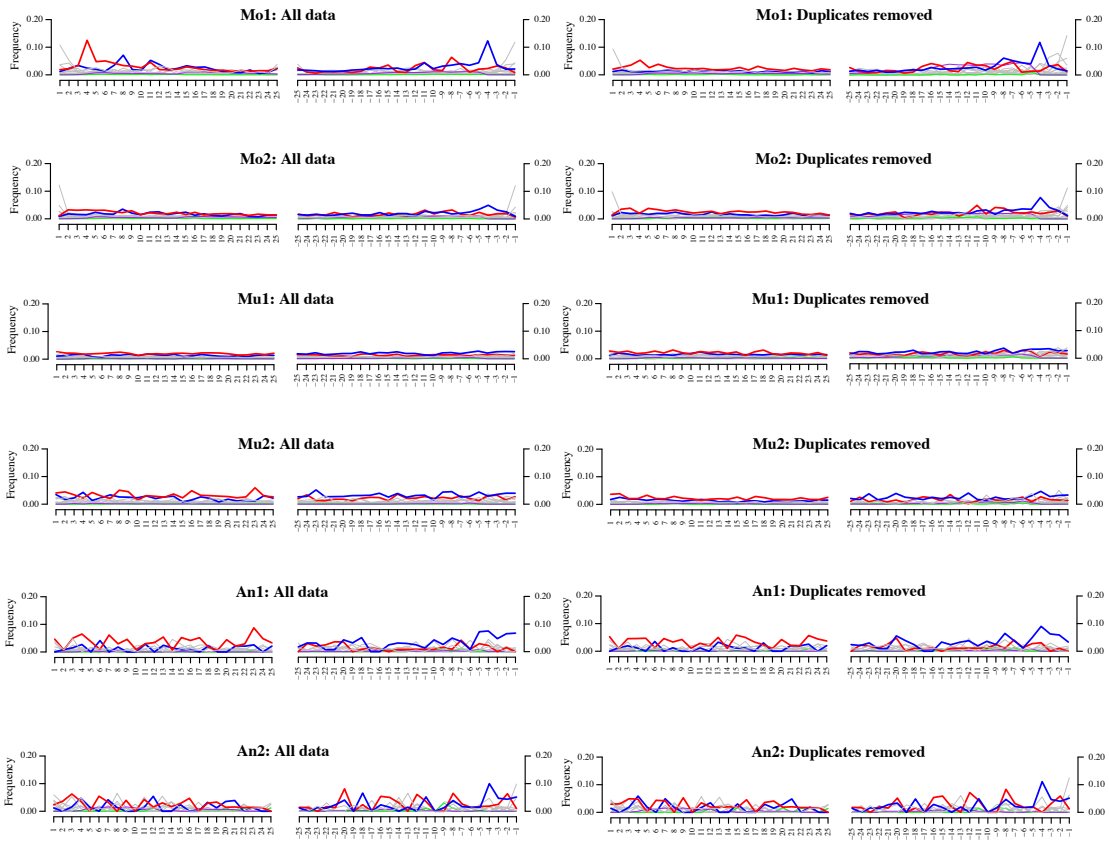


Figure 4.3 Nucleotide misincorporation patterns for *A. alpina* from three different age classes, with all filtered and merged data (all data), and data with duplicates removed. Frequencies of mismatches and indels between read and reference sequences are plotted for the first (5' to 3') and last (3' to 5') 25 nucleotides of reads. Red: C→T, blue: G→A, grey: other mismatch type, pink: insertion, green: deletion.

misincorporation pattern, with the only notable change being a decrease in the rates of misincorporation at the 5' end of Mo1. Visual inspection of individual *BAM* files showed that in the whole mitochondrion-aligned datasets there was large sequence divergence between the reads and the reference, and that in the ITS1-aligned datasets there was an excessive amount of single mismatches between reads and the reference sequence.

4.5. Discussion

4.5.1. Sequence Data

43.3 million paired end reads were obtained, which is 21.7% of the maximum output (200 million reads) of the Illumina HiSeq-2000 platform at Exeter. This low output is likely due to concentration overestimation of the DNA libraries during quantification. Spectrophotometers measure all nucleic acids, regardless of the presence of ligated adapters, and are therefore prone to overestimating DNA concentration (Buehler *et al*. 2010). Interestingly, all four modern and museum samples had similar numbers of reads, indicating that quantification was consistently overestimated. Reducing the calculated concentration by four-fold in samples of this type would have produced a number of reads compatible with the maximum output of the HiSeq-2000. Additionally, the ancient samples had a similar numbers of reads to each other, which were roughly half those attained with the modern and museum samples. It is therefore recommended that reducing the calculated concentration by eight-fold in ancient insect samples would enable maximal output using the outlined method.

Allowing a single mismatch in the barcode increased the number of assigned reads by 1.48% [1.31% after filtering (a.f.)] compared to if only exact barcode matches were considered, whereas allowing two mismatches only increased the number of assigned reads by a further 0.23% (0.16% a.f.). Allowing for two mismatches during barcode assignment considerably increased the proportion of additional reads which were removed by quality filtering (11% for one mismatch, 30% for two mismatches). This, together with the fact that allowing two mismatches increases the chance of a barcode misidentification error, meant only barcodes that were either an exact match or had a single mismatch were considered for further analysis.

0.77% (0.81% a.f.) of reads were assigned to four RHUL barcodes that were not used in the present study. Therefore, assuming that an average of 0.2% of the total reads represent cross contamination per barcode in this experiment, it is possible that ~1.2% of reads binned by barcode have been incorrectly assigned. This cross contamination could have originated during oligonucleotide synthesis or handling in the laboratory (Kircher *et al*. 2011; Kircher *et al*. 2012). A powerful method for reducing the problems of barcode cross contamination is through double barcoding (a barcode at each end of the DNA fragment) (Kircher *et al*. 2012). It is recommended that this method be implemented in similar future studies.

4.5.2. Aligner Comparison and Duplicate Removal Bias

In the majority of cases in this study, Bowtie2 outperformed BWA by assigning more reads to the reference sequences, especially if the input file was in *FASTQ* format, but required more CPU time. The assignment of more reads by Bowtie2 is in agreement with the findings of Langmead and Salzberg (2012), although the finding here of BWA requiring less CPU time differs from their result. Further work on a larger variety of datasets would be needed to verify that these findings are not an artefact of this dataset or the result of short genomic reference sequences. This may be especially worthwhile for large whole nuclear genome wide alignments for which, in ancient and museum DNA analyses, both BWA [e.g. Schubert *et al*. (2012); Orlando *et al*. (2011); Menzies *et al*. (2012); Allentoft *et al*. (2012); Rasmussen *et al*. (2011)] and Bowtie [e.g. Rowe *et al*. (2011)] have recently been used.

In published museum and ancient DNA NGS datasets, the proportion of reads removed as duplicates can be high [63.8 and 37.9% respectively (Rasmussen *et al*. 2011; Rasmussen *et al*. 2010)]. In this study however, duplicate removal was greater than these previous studies, and was biased toward modern and museum nuDNA. Furthermore, a higher proportion of reads were designated duplicates in datasets with a greater number of assigned reads. It is therefore considered unlikely that the duplicates removed in this study were purely the result of PCR amplification (artificial duplicates); a proportion of the removed duplicates may have been natural duplicates that originated from unique template molecules. It is hypothesised that the undesirable removal of natural duplicates may have been due to reference sequence saturation. If

there were a large enough number of unique template molecules that originated from a multi-copy marker, then there would be an increased chance of these molecules appearing as duplicates in an alignment that is only a single copy of the marker. This would occur when reads are aligned to short reference sequences of multi-copy markers (e.g. 183 to 1043bp for 28S, ITS1, 18S), and would also explain why a greater proportion of reads were removed as duplicates in larger datasets. The bias toward nuDNA read removal is not present in the ancient samples, where there are fewer reads and therefore the reference sequence has not been saturated. The likelihood of natural duplicates being removed in these cases is therefore very unlikely (Kircher 2012), and so the 19.4 to 29.0% of reads removed for the ancient samples are likely to represent artificial duplicates. The fact that multi-copy 28S, ITS1, and 18S were preferentially removed compared to the multi-copy mtDNA may suggest a very high copy number of ribosomal DNA in *A. alpina*. Other insects can have thousands of copies per diploid genomes [>2000 to 8000 (Kumar & Rai 1990; Oishi *et al*. 1985)]. Through the excessive removal of duplicates, reference sequence saturation may therefore be a potential problem for the alignment of reads to multi-copy sequences in the absence of an appropriate reference nuclear genome. Other studies [e.g. Schubert *et al*. (2012)], which utilise a reference genome, avoid this by discarding reads that align to multiple parts of the genome, but to do so here would have rejected nearly all of the nuDNA data. A solution to this problem could be to build a degenerate base region into the adapter molecule, in order to distinguish between duplicates of natural and artificial origin (Casbon *et al*. 2011). Similar unique molecular identifiers (UMIs) have been used to count absolute numbers of template molecules (Kivioja *et al*. 2012) and for the detection of ultra-rare alleles (Schmitt *et al*. 2012), where distinguishing between natural and artificial duplicates is essential. Given that the retention of as many non-artefact endogenous reads as possible is highly important for NGS-based aDNA studies, especially in the absence of an appropriate whole nuclear genome reference, investigatory studies using a UMI-based approach are warranted.

The fact that a much lower proportion of reads were assigned to reference sequences for the ancient samples, even with duplicate removal, is consistent with the general observation that ancient specimens yield a far smaller amount of endogenous DNA

compared to those of museum age (Sawyer *et al*. 2012). The possibility that these reads were the result of barcode cross contamination is refuted as visual inspection of sample-specific COI reference sequences matched the assigned reads and the number of assigned reads for both ancient samples was greater than 1.2% of the number assigned to the modern and museum samples. However, it is probable that some reads were the result of barcode cross contamination.

4.5.3. Fragment Length Distribution and DNA Damage

Considering the large disparity in sample age, the fact that the mean fragment lengths of the modern, museum, and ancient samples are similar (60.9 to 123.7 bps) and do not display an obvious temporal signal is explained by the idea that strand breaks are thought to occur rapidly *post mortem* (Sawyer *et al*. 2012) and there can be large variation in fragment length between samples of similar ages (Briggs *et al*. 2007; Gilbert *et al*. 2007). The fragment length distribution patterns are comparable to other aDNA datasets (Kircher 2012), as are published mean fragment lengths from both museum [69 to 87.5 bps (Miller *et al*. 2009; Rasmussen *et al*. 2011)] and permafrost-preserved [60.5 to 128.1 bps (Gilbert *et al*. 2008; Gilbert *et al*. 2007)] specimens. The 'all data' mtDNA fragments are significantly longer than nuDNA. This is in agreement with recent studies of vertebrates from permafrost and sediment deposits, which suggest that nuDNA degrades at a faster rate than mtDNA (Allentoft *et al*. 2012; Schwarz *et al*. 2009). It has been speculated that this may be due to the circular configuration of mtDNA, which makes it is less accessible to exonucleases (Allentoft *et al*. 2012), the double membrane of the mitochondrion offering additional protection (Schwarz *et al*. 2009), or the interaction of nuDNA and histones facilitating strand breaks (Binladen *et al*. 2006). Based on thermal age analysis however, a study of insect DNA from Roman and medieval aged sediments suggested that nuDNA was better preserved (King *et al*. 2009). The removal of 'duplicates' reduced this pattern in the modern and museum datasets, with the pattern reversed in Mo2 and removed in Mu2. This was due to shorter reads being preferentially removed, of which there were more in the nuDNA datasets. This may be an artefact of the consensus or representative sequence determination that occurs during duplicate removal (Kircher 2012). It is unlikely that nuDNA was preferentially amplified, and therefore has more artificial duplicates, as the amplification enzyme (*Pfu* Turbo Cx) used during library

preparation in this study does not have a strong preference for shorter sequences (Dabney & Meyer 2012).

The very low C→T and G→A misincorporation rates observed in Mo2 and Mu1 are consistent with their young age (<100yrs) because, unlike strand breaks, these misincorporation rates positively correlate with age (Sawyer *et al*. 2012). However, the higher misincorporation rate exhibited by Mo1 demonstrated variation between samples. It is not known why the removal of duplicates resulted in a reduced misincorporation rate at the 5' end in this sample only. The older (>100yrs) Mu2 had a higher rate of base misincorporation than the other modern and museum samples, as expected through age. This rate was higher than the 1% misincorporation rate found in 100yr old museum-stored human hair (Rasmussen *et al*. 2011), even though DNA degradation in hair is likely to have begun prior to the individual's death, whilst exposed to natural environmental conditions. Additionally, misincorporations are located throughout the read length and are not clustered at termini, unlike the misincorporation pattern seen in the human hair example (Rasmussen *et al*. 2011). In the ancient samples, misincorporation rates at read termini were also far lower than those predicted by age (28,100 and >11,700 cal. yrs BP). Base misincorporation rates at DNA molecule termini from permafrost-preserved bone are typically around 20 to 30% [e.g. Briggs *et al*. (2007); Orlando *et al*. (2011)]. As with Mu2, misincorporations, at a rate of 5 to 10%, were instead spread throughout the read. The possibility that this was an artefact of small sample size was considered unlikely, as both ancient samples had different sample sizes yet displayed very similar base misincorporation profiles. Additionally, as all samples are conspecific, the rates of mismatch from other sources (sequence error, divergence from reference sequences, SNPs, heterogeneous multi-copy sequences) should be constant between samples regardless of age. Further investigation into the mechanisms of DNA misincorporations in older museum and permafrost-preserved insect specimens is therefore merited. The large sequence divergence between the reads and the reference in the whole mitochondrion-aligned datasets was likely due to evolutionary divergence, as these reference sequences belong to taxa from different genera to *A. alpina*. The excessive amounts of mismatch in the ITS1-aligned datasets could be due to intragenomic heterogeneity or repetitive element variability within the multi-copy

ITS1, two features that have been noted in other insects (Fairley *et al*. 2005; Parkin & Butlin 2004; Schulenburg *et al*. 2001; Sword *et al*. 2007; Vogler & Desalle 1994). However, further investigation would be needed to confirm this inference.

4.6. Conclusions

In this study, some bioinformatic approaches for analysing NGS data from the degraded DNA of a taxon without an appropriate reference genome have been examined. In addition, the state of DNA preservation in museum and ancient specimens of *A. alpina* was explored through the assessment of fragment length distributions and the prevalence of damage, in the form of base misincorporations. Allowing a single mismatch in the barcode allowed for the optimal recovery of sample-specific reads. The BWA and Bowtie2 aligners were compared, and it was found that Bowtie2 aligns more reads. Using rmdup of the SAMtools suite, duplicate removal resulted in a greater proportion of nuDNA reads, as well as reads from larger datasets, being removed. This was explained by reference sequence saturation, whereby short reference sequences of multi-copy markers become saturated with reads, resulting in the removal of both artificial and natural duplicates. Investigation into the use of UMIs to prevent this problem in future studies is recommended. Fragment length distributions are comparable to previous aDNA studies, with nuDNA exhibiting a shorter mean fragment length than mtDNA. Shorter fragments were more likely to be removed as duplicates, which resulted in an increase in the mean fragment length for all datasets after duplicate removal. A general trend of increased base misincorporation with increasing sample age was observed. Unlike previous studies however, misincorporations did not cluster at read termini, but were instead spread throughout the read. These results further demonstrate the potential of DNA recovered from both museum and permafrost-preserved beetle specimens.

4.7. References

Allentoft ME, Collins M, Harker D, *et al*. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society of London Series B-Biological Sciences* **279**, 4724-4733.

Binladen J, Wiuf C, Gilbert MTP, *et al*. (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics* **172**, 733-741.

Briggs AW, Stenzel U, Johnson PL, *et al*. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14616-14621.

Buehler B, Hogrefe HH, Scott G, *et al*. (2010) Rapid quantification of DNA libraries for next-generation sequencing. *Methods* **50**, S15-S18.

Casbon JA, Osborne RJ, Brenner S, Lichtenstein CP (2011) A method for counting PCR template molecules with application to next-generation sequencing. *Nucleic Acids Research* **39**, e81.

Dabney J, Meyer M (2012) Length and GC-biases during sequencing library amplification: A comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87-94.

Fairley TL, Kilpatrick CW, Conn JE (2005) Intragenomic heterogeneity of internal transcribed spacer rDNA in neotropical malaria vector *Anopheles aquasalis* (Diptera : Culicidae). *Journal of Medical Entomology* **42**, 795-800.

Gibson CM, Kao RH, Blevins KK, Travers PD (2012) Integrative taxonomy for continental-scale terrestrial insect observations. *PLoS One* **7**, e37528.

Gilbert MTP, Kivisild T, Gronnow B, *et al*. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787-1789.

Gilbert MTP, Tomsho LP, Rendulic S, *et al*. (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**, 1927-1930.

Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, Orlando L (2011) mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics* **27**, 2153-2155.

Green RE, Krause J, Briggs AW, *et al*. (2010) A draft sequence of the Neandertal genome. *Science* **328**, 710-722.

Gullan PJ, Cranston PS (2010) *The Insects: An Outline of Entomology*, 4th edn. Wiley-Blackwell, UK.

King GA, Gilbert MTP, Willerslev E, Collins MJ, Kenward H (2009) Recovery of DNA from archaeological insect remains: first results, problems and potential. *Journal of Archaeological Science* **36**, 1179-1183.

Kircher M (2012) Analysis of high-throughput ancient DNA sequencing data. In: *Ancient DNA: Methods and Protocols, Methods in Molecular Biology* (eds. Shapiro B, Hofreiter M), pp. 197-228. Springer.

Kircher M, Heyn P, Kelso J (2011) Addressing challenges in the production and analysis of illumina sequencing data. *Bmc Genomics* **12**, 382.

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research* **40**, e3.

Kivioja T, Vaharautio A, Karlsson K, *et al*. (2012) Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods* **9**, 72-74.

Kumar A, Rai KS (1990) Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera, Culicidae). *Hereditas* **113**, 277-289.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.

Li H, Handsaker B, Wysoker A, *et al*. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.

McKenna DD, Farrell BD (2009) Beetles (Coleoptera). In: *The Timetree of Life* (eds. Hedges SB, Kumar S), pp. 278-289. Oxford University Press, Oxford, UK.

Menzies BR, Renfree MB, Heider T, *et al*. (2012) Limited genetic diversity preceded extinction of the Tasmanian tiger. *PLoS One* **7**, e35433.

Miller W, Drautz DI, Janecka JE, *et al*. (2009) The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research* **19**, 213-220.

Miller W, Drautz DI, Ratan A, *et al*. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387-390.

Milne I, Bayer M, Cardle L, *et al*. (2010) Tablet--next generation sequence assembly visualization. *Bioinformatics* **26**, 401-402.

Ober KA (2002) Phylogenetic relationships of the carabid subfamily Harpalinae (Coleoptera) based on molecular sequence data. *Molecular Phylogenetics and Evolution* **24**, 228-248.

Oishi M, Locke J, Wyatt GR (1985) The ribosomal RNA genes of *Locusta migratoria*: copy number and evidence for underreplication in a polyploid tissue. *Canadian Journal of Biochemistry and Cell Biology* **63**, 1064-1070.

Orlando L, Ginolhac A, Raghavan M, *et al*. (2011) True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Research* **21**, 1705-1719.

Parkin EJ, Butlin RK (2004) Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. *Molecular Biology and Evolution* **21**, 1595-1601.

Poinar HN, Schwarz C, Qi J, *et al*. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392-394.

Rasmussen M, Guo X, Wang Y, *et al*. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98.

Rasmussen M, Li YR, Lindgreen S, *et al*. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762.

Raupach MJ, Astrin JJ, Hannig K, *et al*. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* **7**, 1-15.

Richards S, Gibbs RA, Weinstock GM, *et al*. (2008) The genome of the model beetle and pest *Tribolium castaneum*. *Nature* **452**, 949-955.

Rowe KC, Singhal S, Macmanes MD, *et al*. (2011) Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* **11**, 1082-1092.

Ruiz C, Jordal B, Serrano J (2009) Molecular phylogeny of the tribe Sphodrini (Coleoptera: Carabidae) based on mitochondrial and nuclear markers. *Molecular Phylogenetics and Evolution* **50**, 44-58.

Sanchez-Gea J-F, Galian J, Serrano J (2004) Phylogeny of Iberian *Zabrus* (Coleoptera: Carabidae: Zabrini) based on mitochondrial DNA sequence. *European Journal of Entomology* **101**, 503–511.

Sasakawa K, Kubota K (2007) Phylogeny and genital evolution of carabid beetles in the genus *Pterostichus* and its allied genera (Coleoptera: Carabidae) inferred from two nuclear gene sequences. *Annals of the Entomological Society of America* **100**, 100-109.

Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS One* **7**, e34131.

Schmitt MW, Kennedy SR, Salk JJ, *et al*. (2012) Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 14508-14513.

Schubert M, Ginolhac A, Lindgreen S, *et al*. (2012) Improving ancient DNA read mapping against modern reference genomes. *Bmc Genomics* **13**, 178.

Schulenburg JHGvd, Hancock JM, Pagnamenta A, *et al*. (2001) Extreme length and length variation in the first ribosomal internal transcribed spacer of ladybird beetles (Coleoptera : Coccinellidae). *Molecular Biology and Evolution* **18**, 648-660.

Schwarz C, Debruyne R, Kuch M, *et al*. (2009) New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Research* **37**, 3215-3229.

Song H, Sheffield NC, Cameron SL, Miller KB, Whiting MF (2010) When phylogenetic assumptions are violated: base compositional heterogeneity and among-site rate variation in beetle mitochondrial phylogenomics. *Systematic Entomology* **35**, 429–448.

Sword GA, Senior LB, Gaskin JF, Joern A (2007) Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera : Acrididae). *Systematic Entomology* **32**, 420-428.

Thomsen PF, Elias S, Gilbert MT, *et al*. (2009) Non-destructive sampling of ancient insect DNA. *PLoS One* **4**, e5048.

Timmermans MJ, Dodsworth S, Culverwell CL, *et al*. (2010) Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics. *Nucleic Acids Research* **38**, e197.

Vogler AP, Desalle R (1994) Evolution and Phylogenetic Information-Content of the Its-1 Region in the Tiger Beetle *Cicindela Dorsalis*. *Molecular Biology and Evolution* **11**, 393-405.

Wan X, Hong MY, Liao A, *et al*. (2012) Complete mitochondrial genome of a carabid beetle, *Damaster mirabilissimus mirabilissim* (Coleoptera: Carabidae). *Entomological Research* **42**, 44-54.

Willerslev E, Gilbert MTP, Binladen J, *et al*. (2009) Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. *BMC Evolutionary Biology* **9**, 95.

Chapter 5. Taxonomic characterisation of the museum and ancient metagenomes in *Amara alpina* (Coleoptera: Carabidae)

5.1. Abstract

The proportions of endogenous to contaminant DNA are not known for both dry-stored museum and ancient permafrost-preserved insect remains. However, such data would characterise the metagenome of these remains, and may also be of use for future museomic or palaeogenomic study. Using a shotgun NGS-based approach, this study attempted to taxonomically characterise the DNA present in museum and ancient specimens of the beetle *Amara alpina*. The results demonstrate that museum specimens have a metagenome consisting of at least 25 to 40% endogenous insect DNA, as well as a proportion of bacterial parasite and commensal DNA. However, the metagenome of ancient specimens contained only <0.5% endogenous insect DNA, with the vast majority of DNA characteristic of bacteria from the preservational environment. This study demonstrates the vast potential for museomic studies of dry-stored museum specimens, but perhaps limited potential for shotgun-based palaeogenomic studies of ancient permafrost specimens.

5.2. Introduction

Previous attempts to characterise the DNA present in dry-stored museum specimens of insects have often used the concentration of extracted DNA as a measure of DNA preservation (Andersen & Mills 2012; Tagliavia *et al*. 2011; Watts *et al*. 2007; Zimmermann *et al*. 2008). However, this assumes that all extracted DNA is endogenous. Based on DNA from mammalian museum specimens, this assumption may be problematic as contaminants are often present (Menzies *et al*. 2012; Rasmussen *et al*. 2011; Rowe *et al*. 2011). A taxonomic assessment of the DNA extracted from museum specimens would therefore allow for characterisation of the museum metagenome [*sensu* Miller *et al*. (2009)] in these remains. In addition, the proportion of endogenous DNA in ancient permafrost-preserved beetles is unknown.

Shotgun sequencing, using NGS technologies, has proven to be a powerful tool for retrieving large amounts of genetic data from degraded DNA in both museum [e.g. Miller *et al*. (2009); Rowe *et al*. (2011); Miller *et al*. (2011); Rasmussen *et al*. (2011)] and ancient permafrost-preserved [e.g. Gilbert *et al*. (2007); Rasmussen *et al*. (2010)] specimens of a variety of non-insect organisms, and has allowed for highly accurate, in-depth study and DNA characterisation of these specimens [see Paijmans *et al*. (2012) for a review]. An NGS-based approach is therefore highly suited for taxonomic characterisation of the DNA present in degraded insect remains. This approach would assess the potential of museum and ancient specimens, through comparison of the relative proportions of endogenous to exogenous DNA, for future museomic and palaeogenomic study, as well as characterising the museum and ancient metagenomes, respectively.

The overarching aim of this study therefore was to taxonomically characterise the genetic component of samples derived from two dry-stored museum (Mu) and two permafrost-preserved ancient (An) specimens of *Amara alpina*. In order to provide a comparison, two modern (Mo; <10yrs) specimens were also evaluated (table 2.3). Specific aims included the assessment of the amount of endogenous insect DNA sequences in museum and ancient samples, and whether the proportions are sufficient for large-scale museomics and palaeogenomics, respectively. Additionally, the amount of exogenous DNA in the samples, the likely provenance of these sequences, and their potential utility for museum and ancient insect metagenomics was investigated.

5.3. Materials and Methods

Following from section 2.5.3, the three *FASTQ* files (merged reads, unmerged read one, unmerged read two) were concatenated, and the 'Collapse' tool in Galaxy was used to remove duplicates introduced during the index PCR (fig. 2.2: section encircled in blue). Reads were assembled into contigs using *de novo* assembly in the CLC assembly cell v4.0.6-beta (http://www.clccell.com/), with the minimum contig length set to 40 nucleotides. Contigs were produced in order to increase the likelihood of robust taxonomic assignment (Prufer *et al*. 2010) and to further collapse PCR

duplicates. Contigs were compared to the nucleotide collection of the BLAST database, using BLAST v2.2.25 (database downloaded on 23/08/2012). Output from BLAST was analysed using the MEtaGenome ANalyzer (MEGAN) v4.70.4 tool (Huson *et al*. 2011), with the *gi_taxid_nucl.bin* downloaded on 03/09/2012. Taxonomic information for each of the samples was combined, normalised to 100,000 contigs, and collapsed to the rank of Class. The twelve most abundant classes across all six samples were scrutinised further by assessing the major composite genera, which each comprised >2% of the identifiable contigs across all samples.

5.4. Results

5.4.1. Modern and Museum Samples

Contigs from the modern (Mo) and museum (Mu) samples range in length from 40 to between 5971 and 7935bps, with N50 values of between 125 and 184bps (table 5.1). The taxonomic profiles (fig. 5.1) of these samples are very similar. The class with the dominant number of contigs is the Insecta (24.4 to 39.2%), with the Mammalia having the next greatest number of contigs (6.5 to 12.8%). Other abundant classes comprise <4% of assigned contigs, with the exception of the Bacterial class Mollicutes, which has 12.3% of contigs in Mo2. The major components of the Insecta contigs belong to a wide variety of orders, including the Lepidoptera, Hymenoptera, Diptera, Hemiptera, and Coleoptera (table 5.2). These components each occur in roughly equal ratios across the four modern and museum samples. The major components of the Mammalia contigs are *Homo* (1.3 to 4.7%) and *Mus* (1.4 to 1.6%), whereas *Spiroplasma* makes up the majority of the Mollicutes contigs. Other major components of the modern and museum samples included *Danio* (Actinopterygii) and

	Contig Summary		
Sample	**Number**	**Longest (bps)**	**N50 (bps)**
Mo1	810816	5971	150
Mo2	1089627	7935	184
Mu1	768811	6541	153
Mu2	491595	6724	125
An1	280185	11617	214
An2	285395	5897	174

Table 5.1 Details of the contigs used for taxonomic assessment. Contigs ranged in length from 40bps to the maximums shown in the table. Mo: modern, Mu: museum, An: ancient.

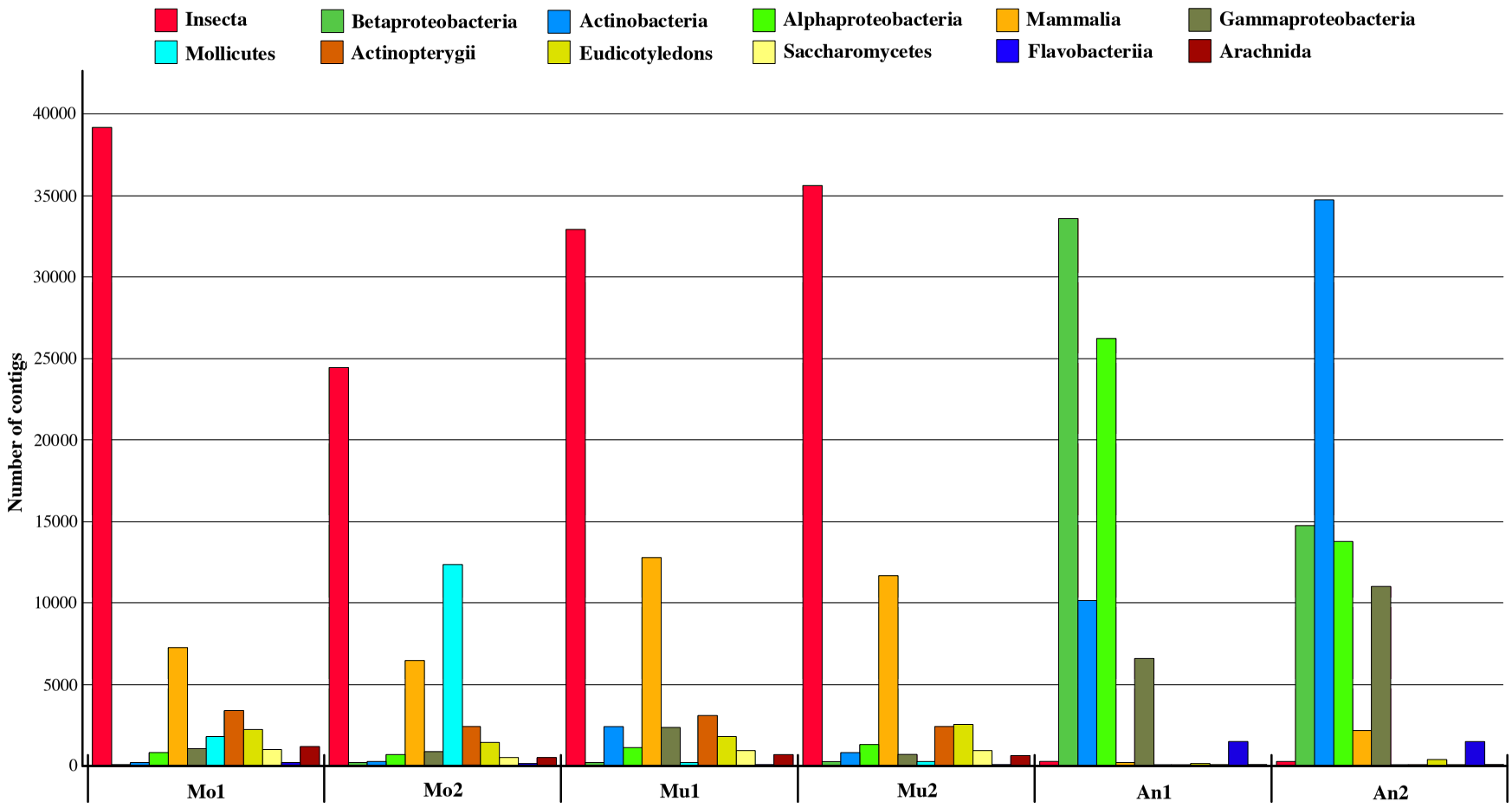


Figure 5.1 DNA extract compositions of the six samples of *A. alpina* from three different age classes, showing the twelve most abundant classes and phyla across all datasets, in descending order. Data are normalised so each dataset has 100,000 contigs. Invertebrates: red, vertebrates: orange, plants and fungi: yellow, Proteobacteria: green, other bacteria: blue.

Kingdom	Phylum/Class	Major component taxa*	Inferred Origin
Animalia	Insecta	*Drosophila, Bombyx, Tribolium, Apis, Nasonia, Acyrthosiphon, Abax, Anopheles, Bombus, Culex, Pediculus, Megachile, Aedes* (all Mo, Mu)	*A. alpina*
Prokaryota	Betaproteobacteria	*Acidovorax* (An1), *Polaromonas* (An), *Dechloromonas* (An1), *Burkholderia* (An), *Delftia* (An)	(Peri)glacial soil near aquatic habitat
Prokaryota	Actinobacteria	*Rhodococcus* (An2), *Streptomyces* (An1)	Soil
Prokaryota	Alphaproteobacteria	*Wolbachia* (Mo, Mu) *Caulobacter*, *Bradyrhizobium*, *Rhodopseudomonas* (all An)	*A. alpina* parasite Soil near aquatic habitat
Animalia	Mammalia	*Homo* (Mo, Mu, An2), *Mus* (Mo, Mu)	Human and stochastic contamination
Prokaryota	Gammaproteobacteria	*Pseudomonas* (An2), *Acinetobacter* (An1)	Soil
Prokaryota	Mollicutes	*Spiroplasma* (Mo)	*A. alpina* commensal
Animalia	Actinopterygii	*Danio* (Mo, Mu)	Stochastic contamination
Plantae	Eudicotyledons	N/A	Stochastic contamination
Fungi	Saccharomycetes	N/A	Stochastic contamination
Prokaryota	Flavobacteriia	*Flavobacterium* (An)	Soil
Animalia	Arachnida	N/A	Stochastic contamination

Table 5.2 Components of the major groups identified in figure 5.1 and their likely origin. Major component taxa includes those with >2000 identifiable contigs across all samples. Higher taxonomy is based on the NCBI database. Mo: modern, Mu: museum, An: ancient, numbers relate to sample. *As identified by BLAST.

Wolbachia (Alphaproteobacteria). The *Wolbachia* contigs for Mo1 are assigned to the *wPip* strain, whereas contigs for the other modern and museum samples are assigned to the *wRi* strain.

5.4.2. Ancient Samples

The two ancient (An) samples comprise contigs that range in length from 40 to between 5897 and 11617bps, with N50 values of between 174 and 214bps (table 5.1). These samples have profoundly different taxonomic profiles to the modern and museum samples, with the vast majority of contigs assigned to the Proteobacteria and Actinobacteria (fig. 5.1). However, the ratios of the groups differ between the samples, with the Alpha- and Betaproteobacteria (26.2 to 33.6% of contigs) dominating in An1 and the Actinobacteria (34.7% of contigs) dominating in An2. Gammaproteobacteria and Flavobacteriia make up the remaining abundant groups in both samples with 6.7 to 10.1% and 1.5% of contigs, respectively. An2 additionally has contigs assigned to Mammalia (2.1%), which are mainly composed of *Homo*. In addition to the ratios of the bacterial groups differing, their components also differ, with the Actinobacteria and Gammaproteobacteria in each sample being dominated by different taxa. Additionally, there is a higher diversity of Betaproteobacteria in An1 (table 5.2). 0.2 and 0.3% of contigs are assigned to the Insecta for An1 and An2, respectively.

5.5. Discussion

5.5.1. DNA Extract Content

There is a large disparity between the taxonomic composition of the modern-museum and ancient samples. Somewhat surprisingly, the compositions of the modern and museum samples are very similar, even considering the large age range of these samples (8 to 137yrs) and their independent storage histories in separate museums (table 2.3). This suggests that the museum metagenome of historical dried insect specimens may be fairly consistent, regardless of specimen age or storage collection. The ancient samples differ more substantially from one another in taxonomic

composition, but are each dominated by five bacterial groups, which can be attributed to their preservational environment.

5.5.2. Insect DNA

In the modern and museum samples, the largest proportion of contigs is assigned to the Insecta. The major component taxa of these contigs include a wide variety of pests (*Tribolium, Acyrthosiphon*), disease vectors (*Anopheles, Aedes*), and insects of economic and/or scientific importance (*Bombyx, Apis, Drosophila*). These insects have available reference nuclear genomes and are therefore biased toward when BLAST determines taxonomic assignments. The exception is *Abax,* which does not have a reference genome, but is taxonomically close to *A. alpina* (both Carabidae: Harpalinae). The number and variety of component taxa demonstrates that there is not a single suitable reference nuclear genome available for *A. alpina*. The assignment of 24.4 to 39.2% of contigs is therefore very encouraging considering a substantial number of *A. alpina* contigs may not have been assigned due to a lack of an appropriate reference sequence (Prufer *et al*. 2010). The museum samples, which were extracted from single legs, contain similar proportions of insect DNA to the modern samples, which were extracted from whole specimens. This suggests that a single leg is sufficient for yielding insect DNA from museum remains. The proportion of endogenous insect DNA in these samples is not as high as some other studies [>52.3 to 89.8% retrieved from 50 to 100yr old bone, hair, and pelt tissue (Menzies *et al*. 2012; Rasmussen *et al*. 2011; Rowe *et al*. 2011)], but is comparable to the 32.1% recovered from 100yr old dried thylacine hair samples (Miller *et al*. 2009) and therefore opens the potential for large-scale museome extraction from these dry-stored insect specimens.

In the ancient samples, the number of contigs assigned to the Insecta is very low (<0.5%), which is probably due to DNA preservation and small specimen size, as well as the biases discussed above. The amount of endogenous insect DNA in these chitinous remains is around two orders of magnitude lower than other permafrost or cold-preserved tissues, such as bone and hair, which typically range from 40 to 90% (Gilbert *et al*. 2008; Lindqvist *et al*. 2010; Miller *et al*. 2008; Poinar *et al*. 2006). This

suggests a limited potential for shotgun-based palaeogenomics from ancient insect specimens.

5.5.3. Bacterial DNA: Parasites and Commensals

Parasites and commensals constitute most of the bacteria in the modern and museum samples. The arthropod parasite *Wolbachia*, is present in all the modern and museum samples, with two distinct strains identified. The *wPip* strain is found in Mo1 only, whereas the *wRi* strain is exclusive to Mo2, Mu1, and Mu2. These strains belong to different *Wolbachia* supergroups [*wRi*: A, *wPip*: B; (Klasson *et al*. 2009)], which would therefore indicate two separate infection events. *Spiroplasma* are arthropod commensals, which can be pathogenic (Regassa & Gasparich 2006). These mollicutes are found in both modern samples and at very low proportions in the museum samples. *Spiroplasma* are at greatest abundance (12.3% of contigs) in Mo2 indicating a heavy load of these bacteria in this sample. As a commensal, *Spiroplasma* are found in the arthropod gut and become pathogenic when they enter the haemolymph (Regassa & Gasparich 2006). Given that the modern samples had far higher proportions of *Spiroplasma* compared to the museum samples, and that DNA was extracted from whole specimens rather than legs in the modern samples, it is inferred that the *Spiroplasma* detected here were commensals. Another major reproductive parasite identified in beetles, *Rickettsia* (Duron *et al*. 2008), was not detected in any of the samples analysed.

5.5.4. Bacterial DNA: Preservational Environment

Although different, the bacterial compositions of the ancient samples are consistent with the preservational environment (permafrost). The major component genera indicate this environment to have been glacial or periglacial soils/sediments, near aquatic sources. This was inferred from *Polaromonas* [found in glacial and periglacial deposits (Darcy *et al*. 2011)], *Caulobacter* [found in aquatic or semi aquatic habitats (Laub *et al*. 2007)], and the remaining genera being typical of soil/sediment environments (Doughari *et al*. 2011; Janssen 2006; Philippot *et al*. 2007). This complements the locality information for these samples (Goldbottom Creek and Titaluk River localities, in permafrozen sediment), and opens up the possibility of

identifying potentially unknown or dubious provenance information based on the bacterial metagenome of ancient specimens. However, the diversity and ubiquity of many environmental bacterial genera would only allow for general inferences on sample provenance to be made.

5.5.5. Other DNA: Contaminants

Contamination of human origin was observed in all samples, although the proportion was very small (<0.1%) in An1. The proportions of human contamination are highest in the modern and museum samples. These proportions are in the region of those isolated from dry-stored hair and pelt tissue [1.7 to 8.9% (Menzies *et al*. 2012; Miller *et al*. 2009; Rasmussen *et al*. 2011)], but are far higher than the <0.15% reported from bone (Rowe *et al*. 2011). Proportions of human contamination in the ancient samples are comparable with other permafrost or cold-preserved taxa, which ranges from negligible to 4.5% (Lindqvist *et al*. 2010; Miller *et al*. 2008; Poinar *et al*. 2006). Samples from the three age classes had their DNA extracted, and libraries constructed, in two separate, isolated laboratories (modern, museum-ancient). Together with stochastic contaminants (*Mus*, *Danio*, Eudicotyledons, Saccharomycetes) occurring in equal ratios across the modern and museum samples, this suggests that the laboratories used were not the source of this contamination. As with the insect taxonomic assignments, the stochastic contaminants are generally assigned to taxa for which there is a reference nuclear genome. These assignments may have resulted from the spurious designation of conserved DNA, contaminant or otherwise, and therefore should be treated with caution.

5.6. Conclusions

In this study, a broad assessment of the DNA content of degraded insect specimens derived from museum and ancient sources was undertaken, in order to assess their taxonomic compositions and potential for future genomic and metagenomic studies. Although only a single species was assessed in this study, a biological or technical reason for *A. alpina* not to be typical of degraded insect remains is not foreseen. Modern and museum samples of varying age and from separate institutions had very similar taxonomic profiles, which were characterised by endogenous insect DNA,

DNA from human sources, and DNA derived from known arthropod parasites. Other bacterial groups, likely to have derived from the museum environment, occur only at low levels. Dried insect museum samples are therefore of great potential for future museomic and museum metagenomic studies. Ancient samples, derived from Late Pleistocene permafrost deposits, had very low proportions of endogenous insect DNA, with most DNA originating from bacteria of the permafrost environment. Therefore, the potential for future palaeogenomic studies on these remains may be limited, but there is potential for metagenomic study, especially if the preservational setting of the specimen is in question.

5.7. References

Andersen JC, Mills NJ (2012) DNA extraction from museum specimens of parasitic hymenoptera. *PLoS One* **7**, e45549.

Darcy JL, Lynch RC, King AJ, Robeson MS, Schmidt SK (2011) Global distribution of *Polaromonas* phylotypes--evidence for a highly successful dispersal capacity. *PLoS One* **6**, e23742.

Doughari HJ, Ndakidemi PA, Human IS, Benade S (2011) The ecology, biology and pathogenesis of *Acinetobacter* spp.: an overview. *Microbes and Environments* **26**, 101-112.

Duron O, Bouchon D, Boutin S, *et al*. (2008) The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology* **6**, 27.

Gilbert MTP, Kivisild T, Gronnow B, *et al*. (2008) Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787-1789.

Gilbert MTP, Tomsho LP, Rendulic S, *et al*. (2007) Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. *Science* **317**, 1927-1930.

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research* **21**, 1552-1560.

Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology* **72**, 1719-1728.

Klasson L, Westberg J, Sapountzis P, *et al*. (2009) The mosaic genome structure of the *Wolbachia wRi* strain infecting *Drosophila simulans*. *Proceedings of the*

National Academy of Sciences of the United States of America **106**, 5725-5730.

Laub MT, Shapiro L, McAdams HH (2007) Systems biology of *Caulobacter*. *Annual Review of Genetics* **41**, 429-441.

Lindqvist C, Schuster SC, Sun Y, *et al*. (2010) Complete mitochondrial genome of a Pleistocene jawbone unveils the origin of polar bear. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5053-5057.

Menzies BR, Renfree MB, Heider T, *et al*. (2012) Limited genetic diversity preceded extinction of the Tasmanian tiger. *PLoS One* **7**, e35433.

Miller W, Drautz DI, Janecka JE, *et al*. (2009) The mitochondrial genome sequence of the Tasmanian tiger (*Thylacinus cynocephalus*). *Genome Research* **19**, 213-220.

Miller W, Drautz DI, Ratan A, *et al*. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* **456**, 387-390.

Miller W, Hayes VM, Ratan A, *et al*. (2011) Genetic diversity and population structure of the endangered marsupial *Sarcophilus harrisii* (Tasmanian devil). *Proceedings of the National Academy of Sciences of the United States of America* **108**, 12348-12353.

Paijmans JLA, Gilbert MTP, Hofreiter M (2012) Mitogenomic analyses from ancient DNA. *Molecular Phylogenetics and Evolution*, http://dx.doi.org/10.1016/j.ympev.2012.1006.1002.

Philippot L, Hallin S, Schloter M (2007) Ecology of denitrifying prokaryotes in agricultural soil. *Advances in Agronomy* **96**, 249-305.

Poinar HN, Schwarz C, Qi J, *et al*. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392-394.

Prufer K, Stenzel U, Hofreiter M, *et al*. (2010) Computational challenges in the analysis of ancient DNA. *Genome Biology* **11**, R47.

Rasmussen M, Guo X, Wang Y, *et al*. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98.

Rasmussen M, Li YR, Lindgreen S, *et al*. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757-762.

Regassa LB, Gasparich GE (2006) Spiroplasmas: evolutionary relationships and biodiversity. *Frontiers in Bioscience* **11**, 2983-3002.

Rowe KC, Singhal S, Macmanes MD, *et al*. (2011) Museum genomics: low-cost and high-accuracy genetic data from historical specimens. *Molecular Ecology Resources* **11**, 1082-1092.

Tagliavia M, Massa B, Albanese I, La Farina M (2011) DNA extraction from Orthoptera museum specimens. *Analytical Letters* **44**, 1058–1062.

Watts PC, Thompson DJ, Allen KA, Kemp SJ (2007) How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses? *Journal of Insect Conservation* **11**, 195–198.

Zimmermann J, Hajibabaei M, Blackburn DC, *et al*. (2008) DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology* **5**, 18.

Chapter 6. Global genetic structuring and the postglacial history of North American *Amara alpina* (Coleoptera: Carabidae), as inferred from museum and ancient DNA

6.1. Abstract

This work is the first to focus on the response of an invertebrate taxon to climatically driven change using a combined museum and ancient DNA based approach. The mechanism by which the arctic ground beetle *Amara alpina* colonised Canada at the end of the last ice age was assessed, utilising information from both modern and ancient representatives. Additionally, the possibility of major population turnover in this species during the last ice age was investigated. It was found that colonising individuals originated from both the western (Beringian) and southern (lower 48 states of the USA) refugia, with the region west of Hudson Bay being colonised from Beringia and the east of Hudson Bay region being colonised from the lower 48. A distinctive group was identified in present day Canada, which is only found in regions that were glaciated during the last ice age. No evidence for major population turnover in ancient Beringian individuals was found, which is in contrast to the major turnover events seen in the mammalian megafauna over this period.

6.2. Introduction

This work focuses on the response of an invertebrate to climatically driven change, based on inferences from museum and ancient DNA. The postglacial colonisation routes into Canada were investigated in the cold-adapted ground beetle *Amara alpina* and compared to the Schwert-Ashworth (1988) model (fig. 1.3). *A. alpina* is ideal for this type of study as it is currently distributed throughout Beringia, arctic Canada, and on peaks in the Rocky and Appalachian mountains (Lindroth 1968). Additionally, it was present in Beringia (Elias *et al*. 2000) and the lower 48 states of the USA (Schwert 1992) during the last ice age. A previous study has investigated the colonisation of Canada at the end of the last ice age by *A. alpina* (Reiss *et al.* 1999). Reiss and colleagues concluded that Canada was mainly colonised from the Beringian

refugium, and that individuals in the southern refugium became isolated on the high mountains of the Rocky and Appalachian mountain ranges; conclusions that were in line with the Schwert-Ashworth (1988) model. Furthermore, Hudson Bay was not found to have been a barrier to eastward dispersal from Beringia. However, Reiss and colleagues' (1999) study utilised restriction fragment length polymorphism (RFLP) analysis, and only employed modern individuals with limited sampling. This study had the advantage of utilising more informative sequence data and individuals from both the past and present to provide a more complete picture.

In addition to glaciation-induced displacement of species, the last ice age was also a time of population turnover and extinction for much of the mammalian megafauna, especially during times of climatic instability, such as Marine Isotope Stage (MIS) 3 (~60,000-28,000 cal. yrs BP; fig. 1.1) (Barnes *et al.* 2002; Barnett *et al.* 2009; Shapiro *et al.* 2004) and the Late Pleistocene to Earliest Holocene (13,900 to 11,700 cal. yrs BP) (Faith & Surovell 2009; Koch & Barnosky 2006). Based on morphological evidence from subfossils, arctic beetles have maintained morphological stasis, and do not seem to have undergone extinction during the entire climatically-turbulent Quaternary, with nearly all species surviving to the present day (Coope 2004; Elias 2010). However, morphological evidence is limited in its resolution for delimiting populations. With the finer resolution offered by aDNA data, it is possible to assess the extent to which *A. alpina* was affected, at the population level, by the aforementioned periods of climatic instability. If extirpated diversity were found within *A. alpina*, then it would be useful to assess if this diversity loss was on a regional or global scale through sampling of the entire modern distribution. This is especially important for arctic beetles as they are thought to have migrated rapidly, and over large distances, during the Quaternary (Coope 2004; Elias 2010).

The aims of this study were therefore threefold. First, genetic data from the entire modern distribution of *A. alpina* were required to make meaningful comparisons with data retrieved from ancient individuals. Second, was to assess the mode of Canadian postglacial colonisation, using both modern (museum) and ancient DNA data. This included assessing whether Hudson Bay acted as a barrier to dispersal, and from which refugium, or refugia, the regions west and east of Hudson Bay were colonised

after the last glaciation. Third was to assess genetic diversity through time, including if there had been any extinction events at the population level, and crucially whether major climatic events during MIS3 and the Late Pleistocene to Earliest Holocene period affected *A. alpina*. Altogether, these aims allowed an assessment of the responses of a key arctic taxon to past climatically driven changes in North America.

6.3. Materials and Methods

6.3.1. Marker and Sample Selection

Only the Sanger sequencing produced COI and 28S datasets were used in this study (table 2.2), due to problems with ITS1 (see section 3.3.6). Phylogenetic and population genetic analyses were conducted using the COI dataset. COI sequence data from Mack (2008) and Thomsen and colleagues (2009) were incorporated into the analysis. Data from Gibson and colleagues (2012) were not included, as their region of COI did not overlap with the target region. Samples were considered if >70% (≥531/759) of bases had been determined. This was to minimise erroneous haplotype designation that can result from missing data. Due to ancient samples being derived from incomplete specimens (sclerites), the minimum number of individuals was calculated for identical sequences from the same locality and age. This resulted in five samples being excluded from further analysis. Altogether, 139 samples were used for downstream analysis (appendix 1).

6.3.2. Phylogenetic Analysis

Sequence alignments were conducted manually in Se-Al v2.0 (Rambaut 2002). Sequences were ordered from complete to least complete in Se-Al, and collapsed into haplotypes using TCS v1.21 (Clement *et al*. 2000). Datasets were analysed for the best substitution model and partition fit, which were selected using the Bayesian Information Criterion, in jModeltest (Posada 2008) and PartitionFinder (Lanfear *et al*. 2012). Two partitions were selected (first-second and third codon positions), using the HKY+G and GTR substitution models, respectively. Phylogenetic analyses were conducted in MrBayes v3.2.1 (Ronquist *et al*. 2012) using two runs of four chains. Analyses were run for 10,000,000 generations, with trees sampled every 1,000, and

the first 25% discarded as burn-in. *A. aulica* (Genbank: AY551824) was initially used to outgroup the phylogeny, but its large genetic distance from the ingroup prevented effective phylogenetic resolution (fig. 6.1a). To rectify this, the outgroup was set as Mt-Hg3, as this is the most divergent clade within *A. alpina* (figs. 6.1a, 6.2). Branch supports were estimated using Bayesian posterior probability values from MrBayes, and bootstrap support from RAxML v7.2.8 (Stamatakis 2006). 10,000 bootstrap replicates were run under the selected partitions using the GTR+G substitution model.

6.3.3. Population Genetic Analysis

Population genetic analyses were conducted in Arlequin v3.5 (Excoffier & Lischer 2010). Eight modern populations were based on geographic region (fig. 6.3), and a further six populations were based on ancient localities within West and East Beringia (regions 3, 4). These six populations were selected to account for temporal heterogeneity in the dataset (Depaulis *et al*. 2009). To visualise mitochondrial haplotype diversity, a minimum spanning haplotype network (MSN) was calculated, using 10,000 permutations and the Tamura-Nei (TrN) +G (α=4) substitution model. The TrN+G model was the best-supported evolutionary model available in Arlequin. The MSN was constructed using HapStar v0.5 (Teacher & Griffiths 2011). Molecular diversity summary statistics, including haplotype and nucleotide diversity indices, and inter-population sequence divergence estimates were calculated. Comparisons between populations were conducted using the exact test of population differentiation (Raymond & Rousset 1995), with the Markov Chain set to 1,000,000 iterations and the number of dememorisation steps set to 100,000.

6.3.4. Comparison with nuDNA Data

The 28S dataset was supplemented with data from Genbank (appendix 5a), using all available 28S sequences of *Amara*. Due to the low genetic variation in this dataset, alignments were constructed in Se-Al to visualise nuclear haplotype diversity.

6.4. Results

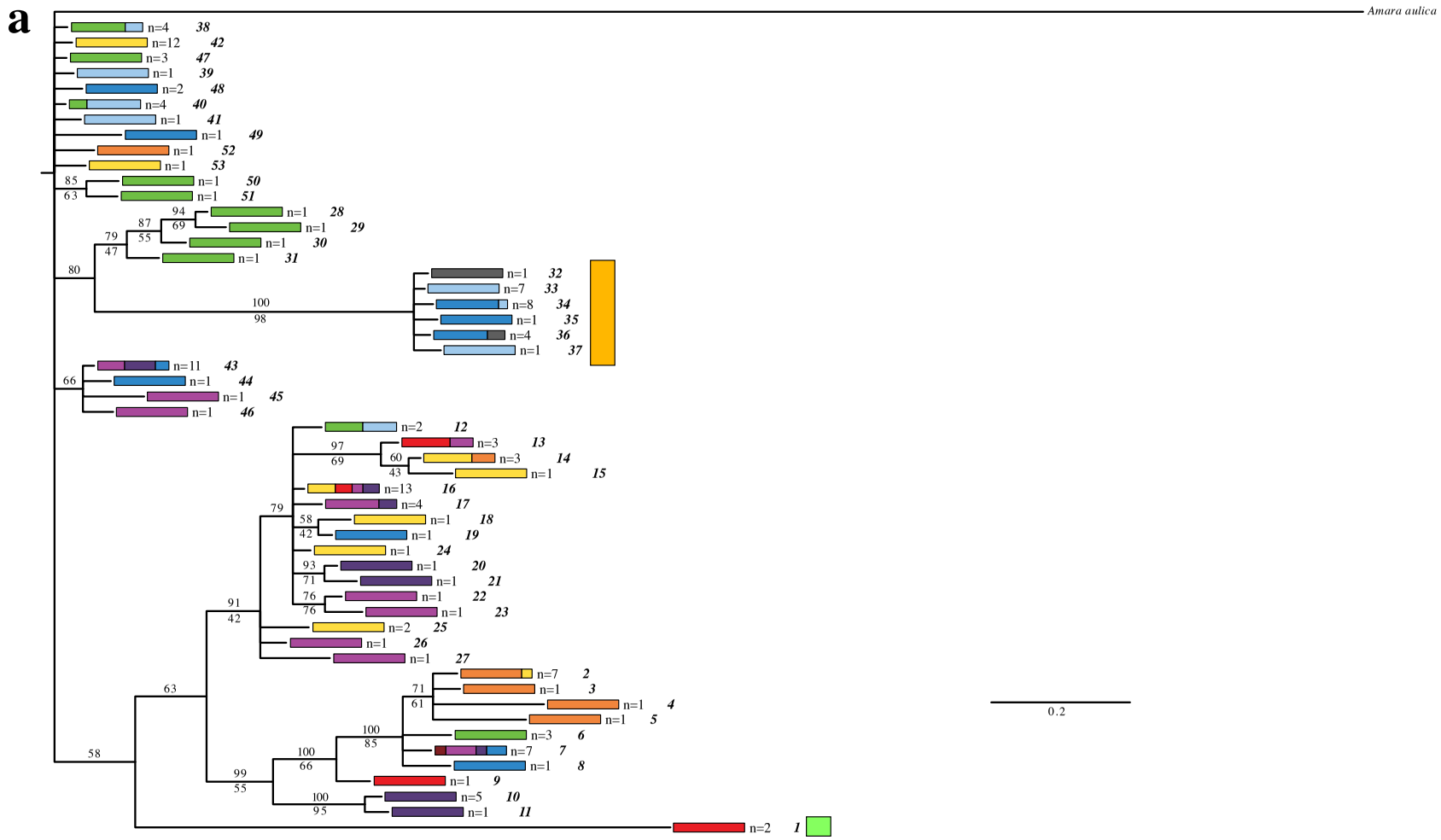
6.4.1. DNA Sequence Data

DNA was recovered in sufficient quantities to be included in the phylogenetic analysis from 144 (50.2%) of the 287 specimens tested, although there was great variation in success rate based on age class (modern/museum: 86.3%, ancient: 16.2%). A total of 53 mitochondrial and seven nuclear haplotypes were identified, with maximum sequence divergence of 5.54% and 1.09% for mt. and nuDNA, respectively.

Based on phylogenetic and network-based analyses (figs. 6.1b, 6.2), *A. alpina* falls into three mitochondrial haplogroups (Mt-Hgs1 to 3), which are defined as diverging from one another by ten mutational steps or more. Mt-Hg1 consists of 46 haplotypes (2 to 31, 38 to 53) and includes all ancient individuals, as well as containing representatives from all study regions, except region 7 ("Hudson Bay"; fig. 6.3). Nucleotide diversity within Mt-Hg1 was 1.20% (table 6.1). Mt-Hg2 consists of six

Region	Population	Age Class	Data	N	*H*	*h*		π	
1	Scandinavia	Modern	All	26	9	76.00	(7.51)	0.81	(0.44)
2	North Russia	Modern	All	11	6	72.73	(14.44)	0.94	(0.54)
3	West Beringia	Modern	All	8	4	82.14	(10.07)	1.88	(1.08)
4	East Beringia	Modern	All	19	11	92.40	(3.75)	1.09	(0.59)
5	W Hudson Bay	Modern	All	21	10	87.14	(5.69)	1.92	(1.00)
6	E Hudson Bay	Modern	All	16	8	80.00	(9.16)	1.78	(0.95)
7	"Hudson Bay"	Modern	All	2	2	100.00	(50.00)	0.15	(0.20)
8	Lower 48	Modern	All	17	11	93.38	(3.93)	1.05	(0.57)
4	TR Pleistocene	Ancient	All	6	3	60.00	(21.52)	0.39	(0.27)
4	TR Holocene	Ancient	All	3	3	100.00	(27.22)	0.79	(0.64)
4	GB Creek	Ancient	All	5	1	0.00	(0.00)	0.00	(0.00)
4	Old Crow 11	Ancient	All	1	1	100.00	(0.00)	0.00	(0.00)
4	Old Crow 106	Ancient	All	3	3	100.00	(27.22)	1.13	(0.90)
3	Ledovy Obryv	Ancient	All	1	1	100.00	(0.00)	0.00	(0.00)
3	West Beringia	Modern	Mt-Hg1	6	3	73.33	(15.52)	0.51	(0.34)
5	W Hudson Bay	Modern	Mt-Hg1	10	7	93.33	(6.20)	1.09	(0.62)
6	E Hudson Bay	Modern	Mt-Hg1	7	5	85.71	(13.71)	0.45	(0.30)
5	W Hudson Bay	Modern	Mt-Hg2	11	3	56.36	(13.40)	0.09	(0.08)
6	E Hudson Bay	Modern	Mt-Hg2	9	3	41.67	(19.07)	0.12	(0.11)
1-6, 8	All Mt-Hg1	Both	Mt-Hg1	115	46	95.67	(0.78)	1.20	(0.61)
5-7	All Mt-Hg2	Modern	Mt-Hg2	22	6	76.19	(5.45)	0.22	(0.15)
3	All Mt-Hg3	Modern	Mt-Hg3	2	1	0.00	(0.00)	0.00	(0.00)

Table 6.1 Molecular diversity summary statistics of *A. alpina*, based on 759bps of COI. Regions correspond to fig. 6.3. Standard deviations are in parentheses. N: individuals, *H*: haplotypes, *h*: haplotype diversity (%), π: nucleotide diversity/within population sequence divergence (%).W: west of, E: east of, GB: Goldbottom.


a
Amara aulica
n=4 38
n=12 42
n=3 47
n=1 39
n=2 48
n=4 40
n=1 41
n=1 49
n=1 52
n=1 53
85
63
n=1 50
n=1 51
94
87
69
n=1 28
n=1 29
79
55
n=1 30
47
n=1 31
80
100
98
n=1 32
n=7 33
n=8 34
n=1 35
n=4 36
n=1 37
66
n=11 43
n=1 44
n=1 45
n=1 46
n=2 12
97
69
n=3 13
60
n=3 14
43
n=1 15
n=13 16
79
n=4 17
58
42
n=1 18
n=1 19
n=1 24
93
71
n=1 20
n=1 21
76
76
n=1 22
n=1 23
91
42
n=2 25
n=1 26
n=1 27
63
n=7 2
71
61
n=1 3
n=1 4
n=1 5
100
85
n=3 6
58
100
66
n=7 7
n=1 8
n=1 9
99
55
100
95
n=5 10
n=1 11
n=2 1
0.2

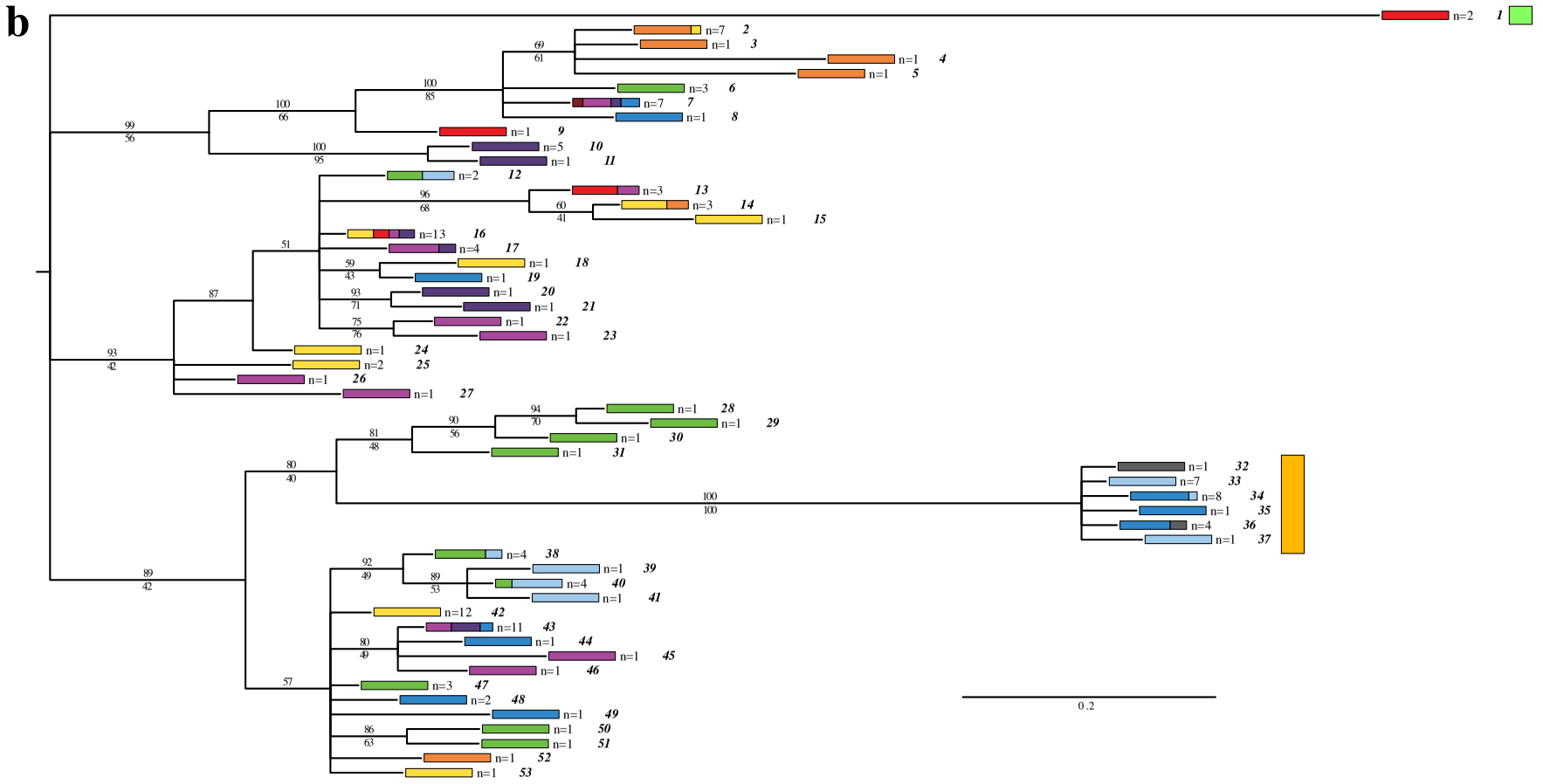


Figure 6.1 Phylogenetic relationships of 53 mitochondrial haplotypes of *Amara alpina*, constructed from 759bps of COI using Bayesian inference. Colours correspond to fig. 6.3: tip bars to geographic locality and clade bars indicate haplogroup designation (Mt-Hg2: orange, 3: green). Mt-Hg1 consists of the remaining haplotypes. Haplotype numbers are in bold-italic. Support values above branches are Bayesian posterior probability values (≥50%) and below branches are maximum likelihood bootstrap values (≥40%). n: individuals. a: outgroup is *A. aulica*. b: outgroup is Mt-Hg3

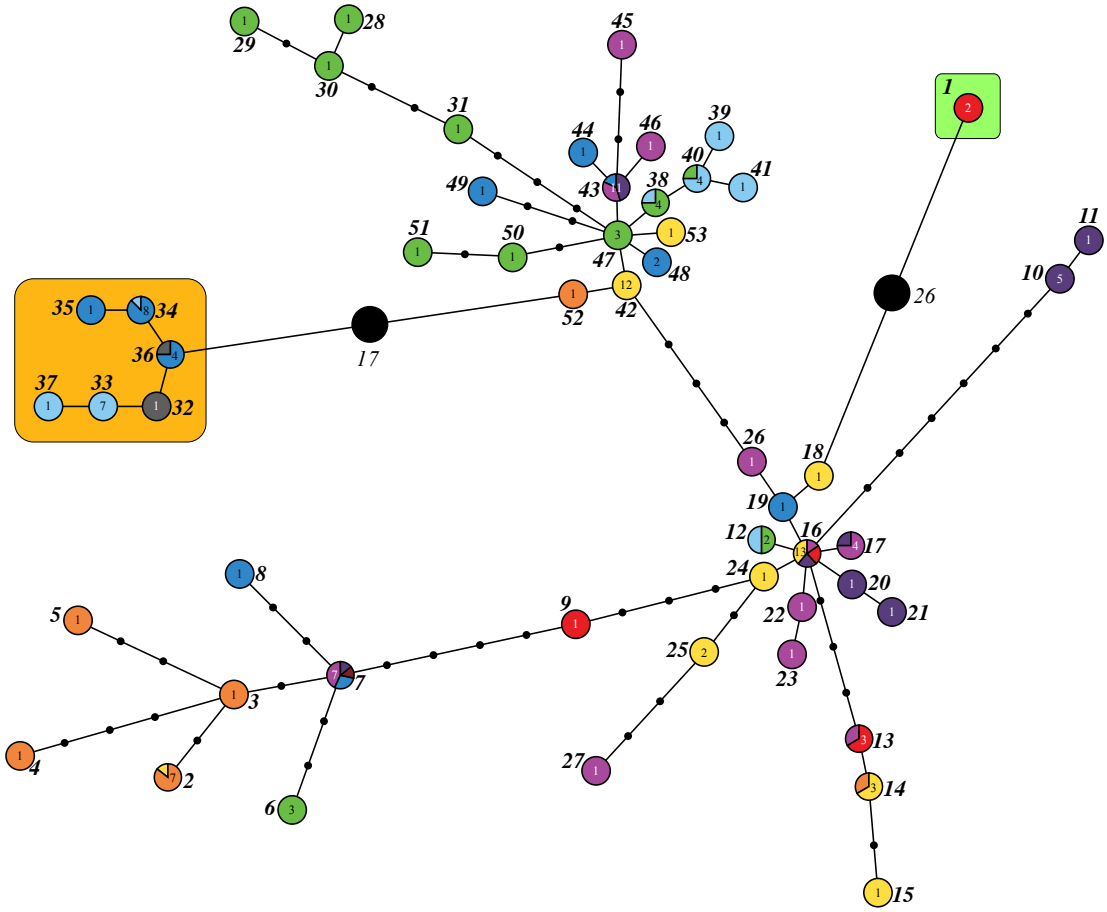


Figure 6.2 Mitochondrial haplotype network of 53 *A. alpina* haplotypes, constructed from 759bps of COI. Colours correspond to fig. 6.3: within-circle to geographic locality and outside-circle indicates haplogroup designation (Mt-Hg2, 3 only). Mt-Hg1 consists of the remaining haplotypes. Coloured circles indicate haplotypes, small black circles indicate missing haplotypes, and large black circles indicate >10 missing haplotypes (exact number in italic next to circle). Numbers within circles indicate individuals. Haplotype numbers are in bold-italic.

haplotypes (32 to 37), which are all found in the non-Beringian regions of Canada (regions 5 to 7), and has 0.22% nucleotide diversity. Mt-Hg1 is paraphyletic with respect to Mt-Hg2, with Mt-Hg2 grouping with haplotypes from the lower 48 (28 to 31; fig. 6.1). Mt-Hg3 consists of a single haplotype (1), represented by two individuals, from Anadyr in West Beringia. Sequence divergence between the three haplogroups ranged from 2.63 to 5.54% (table 6.3).

Excluding populations with a small sample size (≤2), nucleotide diversity within modern Mt-Hg1 populations ranged from 0.45 to 1.09% (table 6.1), with the highest diversity in East Beringia and the west of Hudson Bay (regions 4, 5) and the lowest in

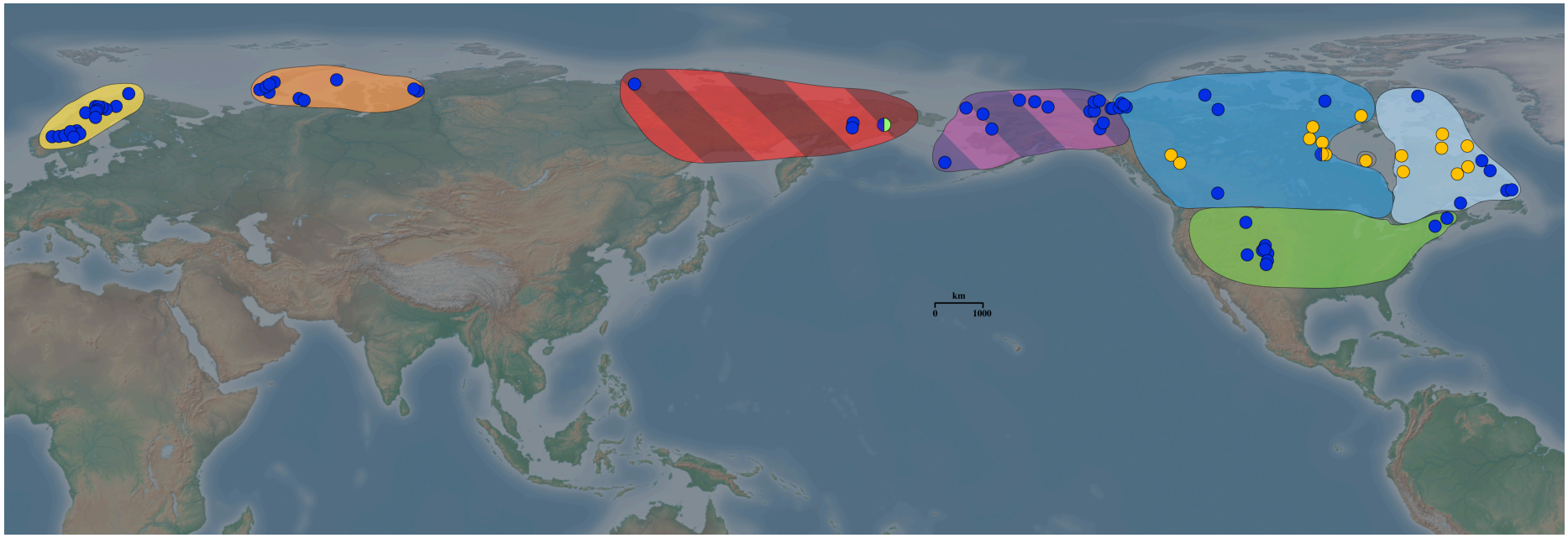


Figure 6.3 Geographic distribution of *A. alpina* individuals investigated. Within-circle colour corresponds to haplogroup designation: Mt-Hg1: Royal Blue, Mt-Hg2: Amber, Mt-Hg3: Light Green. Geographic regions are indicated as: region 1 (Scandinavia; yellow), region 2 (North Russia; orange), region 3 (West Beringia; red), region 4 (East Beringia; purple), region 5 (west of Hudson Bay; blue), region 6 (east of Hudson Bay; light blue), region 7 ("Hudson Bay"; grey), region 8 (Lower 48; green). The darker stripes in regions 3 and 4 represent the ancient localities in West and East Beringia.

the east of Hudson Bay (region 6). Sequence divergence between regions ranged from 0.00 to 1.33% (regions 5/8, 2/6, respectively; tables 6.2, 6.3). Geographically disparate regions, such as Scandinavia, East Beringia, and the lower 48 (regions 1, 4, 8), each contain the bulk of modern genetic diversity found within Mt-Hg1 (figs. 6.1b, 6.2). Nucleotide diversity within ancient Mt-Hg1 populations ranged from 0.00 to 1.13% (Goldbottom Creek, Old Crow 106, respectively), with sequence divergence between populations ranging from 0.12 to 1.80% (TR Pleistocene/Holocene, Old Crow 11/Ledovy Obryv, respectively). In Mt-Hg2, nucleotide diversity ranged from 0.09 to 0.12% (West/east of Hudson Bay), with a sequence divergence between the two regions of 0.24%.

For the 28S dataset (fig. 6.4), 89% of tested samples (n=64) belonged to a single nuclear haplotype (Nu-Ht1). Nu-Ht1 consists of the majority of Mt-Hg1 and all Mt-Hg2 individuals. The two individuals that constitute Mt-Hg3 also compose a distinct nuclear haplotype (Nu-Ht2), which is two mutational steps away from Nu-Ht1. These mutational steps occur in a variable section of the 28S D3 region. Other nuclear haplotypes (n=5) all consist of one sample and are one mutational step away from Nu-Ht1. These nuclear haplotypes are all represented by ancient samples, with the exception of one museum sample (Nu-Ht4), and were all produced from a single PCR reaction. With the exception of Nu-Ht4, these single mutational changes all represent C→T and G→A changes. A further four nuclear 'haplotypes', all one mutational step away from Nu-Ht1 and represented by ancient samples, were collapsed to Nu-Ht1 after further PCR reactions revealed that the single mutational step resulted from a misincorporated base.

6.4.2. Regional Analysis of mtDNA Data

There is significant differentiation between most modern populations, based on the whole dataset (table 6.2). However, "Hudson Bay" (region 7) is not significantly differentiated from any other population, and West and East Beringia (regions 3, 4) are not differentiated. These results persist if subsets of the data based on haplogroup designation (table 6.3) are taken, with exceptions in Mt-Hg1 of East Beringia and west of Hudson Bay (regions 4, 5), and east of Hudson Bay and the lower 48 (regions 6, 8), becoming non-significantly differentiated.

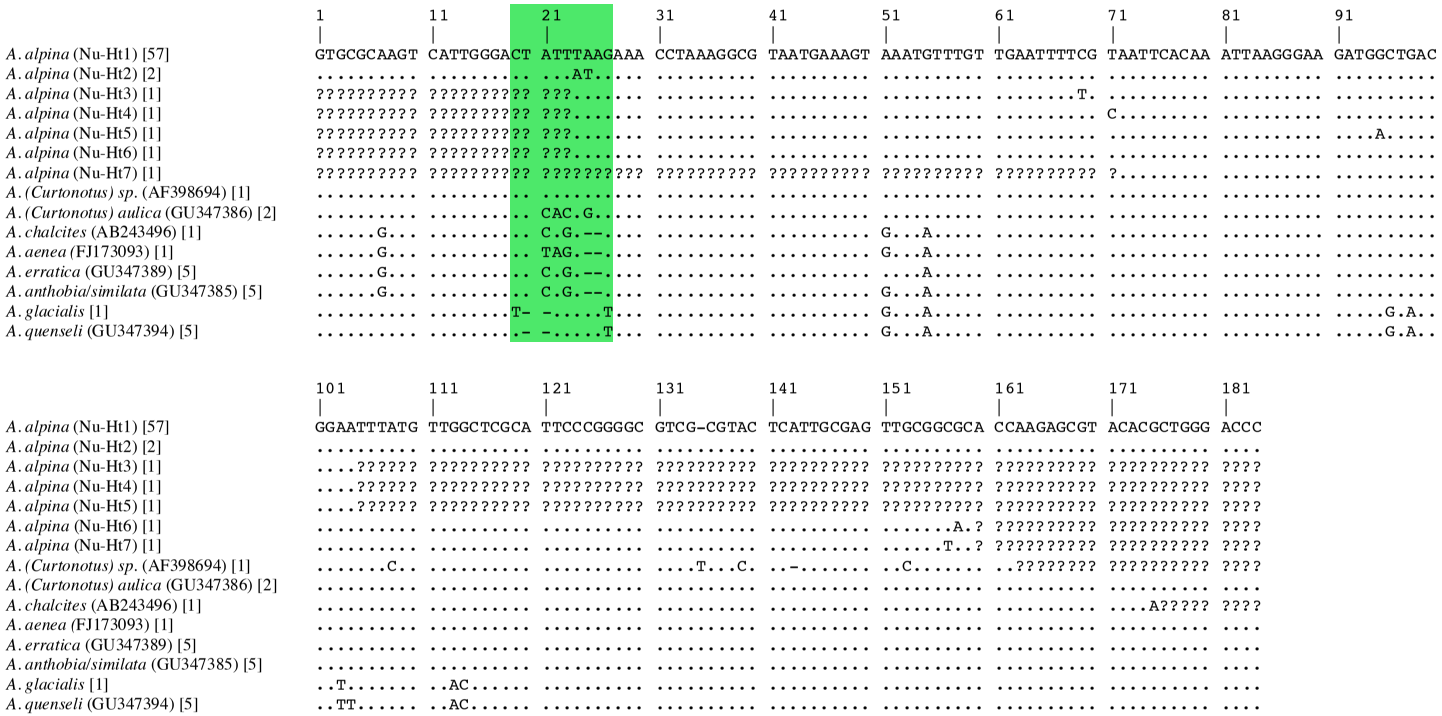


Figure 6.4 Alignment of seven nuDNA haplotypes of *A. alpina*, based on 183bps of the D3 region of 28S rDNA. The alignment is supplemented with seven sequences from Genbank and *A. glacialis* from this thesis. Parentheses indicate either haplotype designation or Genbank accession number. Square brackets indicate number of individuals. Area highlighted in green is a highly variable section of the D3 region identified in *Amara*. ?: missing data, .: identical to reference sequence, -: gap.

Next page: *Table 6.2* Exact test of population differentiation and sequence divergence results for all sequence data. Below diagonal: exact test *p*-values, with standard deviations in parentheses. Significant results ($p<0.05$) are in bold. Above diagonal: percentage sequence divergence between populations. Regions correspond to fig. 6.3. TR: Titaluk River. W: west of, E: east of, GB: Goldbottom.

Age			Modern						
	Region		1	2	3	4	5	6	7
		Population	Scandinavia	North Russia	West Beringia	East Beringia	W Hudson Bay	E Hudson Bay	"Hudson Bay"
Modern	1	Scandinavia	-	0.85	0.27	0.04	0.77	1.04	2.67
	2	North Russia	**<0.001 (<0.001)**	-	0.86	0.62	1.44	1.87	3.50
	3	West Beringia	**0.005 (<0.001)**	**0.001 (<0.001)**	-	0.25	1.26	1.58	3.24
	4	East Beringia	**<0.001 (<0.001)**	**<0.001 (<0.001)**	0.067 (0.001)	-	0.78	1.06	2.70
	5	W Hudson Bay	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**0.002 (<0.001)**	-	0.04	0.57
	6	E Hudson Bay	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	-	0.49
	7	"Hudson Bay"	0.061 (0.001)	0.268 (0.003)	0.112 (0.001)	0.177 (0.002)	0.353 (0.004)	0.182 (0.003)	-
	8	Lower 48	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**0.002 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	**0.002 (<0.001)**	0.262 (0.003)
Ancient	4	TR Pleistocene	**0.002 (<0.001)**	**0.003 (<0.001)**	**0.032 (<0.001)**	**0.012 (<0.001)**	**0.001 (<0.001)**	**0.002 (<0.001)**	0.214 (0.002)
		TR Holocene	0.084 (0.002)	0.155 (0.003)	0.359 (0.002)	0.732 (0.003)	**0.038 (0.001)**	0.090 (0.002)	1.000 (<0.001)
		GB Creek	**<0.001 (<0.001)**	**0.001 (<0.001)**	**0.003 (<0.001)**	0.609 (0.003)	0.067 (0.001)	**0.001 (<0.001)**	**0.048 (<0.001)**
		Old Crow 11	0.226 (0.003)	0.495 (0.003)	0.223 (0.002)	0.404 (0.003)	0.270 (0.003)	0.416 (0.003)	1.000 (<0.001)
		Old Crow 106	0.079 (0.002)	0.155 (0.002)	0.356 (0.002)	0.731 (0.003)	0.156 (0.002)	0.087 (0.002)	1.000 (<0.001)
	3	Ledovy Obryv	0.225 (0.003)	0.499 (0.003)	0.222 (0.001)	1.000 (<0.001)	0.685 (0.003)	0.411 (0.003)	1.000 (<0.001)

Age			Modern	Ancient					
	Region		8	4					3
		Population	Lower 48	TR Pleistocene	TR Holocene	GB Creek	Old Crow 11	Old Crow 106	Ledovy Obryv
Modern	1	Scandinavia	0.21	0.59	0.12	0.45	0.56	0.07	1.23
	2	North Russia	0.70	1.00	0.81	1.42	1.41	0.35	0.30
	3	West Beringia	0.67	0.82	0.18	1.10	0.65	0.12	1.37
	4	East Beringia	0.19	0.52	0.05	0.44	0.47	0.08	0.89
	5	W Hudson Bay	0.61	1.23	1.10	0.84	1.61	0.94	1.63
	6	E Hudson Bay	0.76	1.65	1.46	0.96	1.98	1.32	2.07
	7	"Hudson Bay"	2.41	3.04	3.05	2.67	3.50	2.97	3.80
	8	Lower 48	-	0.85	0.55	0.26	1.10	0.31	0.87
Ancient	4	TR Pleistocene	**0.007 (<0.001)**	-	0.12	1.28	0.88	0.32	1.22
		TR Holocene	0.154 (0.002)	0.761 (0.001)	-	0.96	0.25	0.14	1.17
		GB Creek	**0.005 (<0.001)**	**0.007 (<0.001)**	**0.018 (<0.001)**	-	1.50	0.82	1.61
		Old Crow 11	0.497 (0.003)	0.428 (0.002)	1.000 (<0.001)	0.166 (0.001)	-	0.17	1.80
		Old Crow 106	0.145 (0.002)	0.118 (0.001)	1.000 (<0.001)	**0.018 (<0.001)**	1.000 (<0.001)	-	0.52
	3	Ledovy Obryv	0.505 (0.004)	0.426 (0.002)	1.000 (<0.001)	0.167 (0.001)	1.000 (<0.001)	1.000 (<0.001)	-

Age				Modern					
	Hg.			1					
		Region		1	2	3	4	5	6
			Population	Scandinavia	North Russia	West Beringia	East Beringia	W Hudson Bay	E Hudson Bay
Modern	1	1	Scandinavia	-	-	0.23	-	0.18	0.36
		2	North Russia	-	-	0.89	-	0.54	1.33
		3	West Beringia	**0.025 (0.001)**	**0.003 (<0.001)**	-	0.19	0.66	0.97
		4	East Beringia	-	-	0.184 (0.002)	-	0.10	0.42
		5	W Hudson Bay	**<0.001 (<0.001)**	**0.002 (<0.001)**	**0.014 (<0.001)**	0.277 (0.003)	-	0.28
		6	E Hudson Bay	**<0.001 (<0.001)**	**0.005 (<0.001)**	**0.021 (0.001)**	**0.003 (<0.001)**	**0.026 (0.001)**	-
		8	Lower 48	-	-	**0.006 (<0.001)**	-	**0.002 (<0.001)**	0.245 (0.002)
Ancient	1	4	TR Pleistocene	-	-	**0.047 (0.001)**	-	**0.013 (<0.001)**	**0.020 (0.001)**
			TR Holocene	-	-	0.356 (0.002)	-	0.197 (0.002)	0.300 (0.002)
			GB Creek	-	-	**0.004 (<0.001)**	-	0.214 (0.002)	**0.008 (<0.001)**
			Old Crow 11	-	-	0.284 (0.001)	-	0.454 (0.002)	0.627 (0.002)
			Old Crow 106	-	-	0.357 (0.001)	-	0.580 (0.002)	0.301 (0.002)
		3	Ledovy Obryv	-	-	0.287 (0.001)	-	1.000 (<0.001)	0.624 (0.002)

Age				Modern	Ancient				
	Hg.			1	1				
		Region		8	4				3
			Population	Lower 48	TR Pleistocene	TR Holocene	GB Creek	Old Crow 11	Old Crow 106	Ledovy Obryv
Mod.	1	3	West Beringia	0.76	0.66	0.00	1.20	0.37	0.02	1.38
		5	W Hudson Bay	0.00	0.75	0.46	0.21	1.01	0.17	0.62
		6	E Hudson Bay	0.18	1.15	0.76	0.33	1.28	0.65	1.55

Age				Modern			Both		
	Hg.			2			1	2	3
		Region		5	6	7	1-6, 8	5-7	3
			Population	W Hudson Bay	E Hudson Bay	"Hudson Bay"	All Mt-Hg1	All Mt-Hg2	All Mt-Hg3
Mod.	2	5	W Hudson Bay	-	0.24	0.06	-	-	-
		6	E Hudson Bay	**<0.001 (<0.001)**	-	0.09	-	-	-
		7	"Hudson Bay"	0.115 (0.001)	0.109 (0.001)	-	-	-	-
Both	1	1-6, 8	All Mt-Hg1	-	-	-	-	2.63	3.35
	2	5-7	All Mt-Hg2	-	-	-	**<0.001 (<0.001)**	-	5.54
	3	3	All Mt-Hg3	-	-	-	0.063 (0.003)	**0.014 (<0.001)**	-

Previous page: *Table 6.3* Exact test of population differentiation and sequence divergence results for sequence data subsets based on haplogroup designation. Below diagonal: exact test *p*-values, with standard deviations in parentheses. Significant results ($p<0.05$) are in bold. Above diagonal: percentage sequence divergence between populations. Only sequence divergence results shown in the middle section of the table. Gaps in the table for Mt-Hg1 are to avoid duplication of data presented in table 6.2. Regions correspond to fig. 6.3. TR: Titaluk River, Hg.: Mitochondrial haplogroup (Mt-Hg). W: west of, E: east of, GB: Goldbottom.

6.4.3. Ancient Individuals

The majority of ancient individuals (75%) either belong, or are very close (≤2 mutational steps), to modern mitochondrial haplotypes (7, 16, 20, 21, 43). Two haplotypes (10, 11) however, which are solely composed of ancient individuals from the Titaluk River (TR Pleistocene/Holocene), are at least six mutational steps away from the closest modern haplotype (16; fig. 6.2). These two haplotypes comprise individuals of Late Pleistocene and ?Holocene age. All individuals from the Goldbottom Creek population (28,100 cal. yrs BP) belong to the same haplotype (43), which does not consist of individuals from any other ancient population. The TR Pleistocene population is significantly differentiated from all modern populations, except "Hudson Bay" (region 7), but is not differentiated from the ancient populations, with the exception of the Goldbottom Creek population. The Goldbottom Creek population is significantly differentiated from all populations, except those from modern populations in East Beringia and west of Hudson Bay (regions 4, 5), and ancient populations at Old Crow 11 and Ledovy Obryv (tables 6.2, 6.3).

6.5. Discussion

6.5.1. Genetic Diversity of *Amara alpina*

Compared to other intraspecific sequence divergence estimates of ground beetle COI, which typically vary between 0 and 3.8% (Gibson *et al*. 2012; Raupach *et al*. 2010), *A. alpina* is very genetically diverse, with 5.54% sequence divergence at this locus. The high level of genetic diversity in this species is unsurprising, as *A. alpina* is also very variable morphologically and, under older classifications, has been subdivided into separate species or subspecies (Lindroth 1968). Analysis of the whole

mitogenome, using RFLP analysis, revealed an intraspecific sequence divergence estimate of 0.32% in *A. alpina* (Reiss *et al*. 1999); more than an order of magnitude lower than the estimate of the present study. The disparity between these two estimates could be due to the choice of restriction enzymes used by Reiss and colleagues, or that COI evolves at a faster rate than the average entire mitogenome in *A. alpina*. A recent study found that six individuals of *A. alpina* from two localities in Colorado had a COI sequence divergence of 2.02% (Gibson *et al*. 2012). This is comparable to the 1.05% COI divergence within the whole lower 48 found in this study.

Within Mt-Hg1, the bulk of modern genetic diversity can be found either in Scandinavia, East Beringia, or the lower 48 (regions 1, 4, 8). This is indicative of significant migration between regions, a phenomenon suggested by the beetle subfossil record (Elias 2010), and therefore a complex demographic history of this species as a whole.

Seven 28S haplotypes were identified in *A. alpina*. Given that Nu-Ht2 and Mt-Hg3 consist of the same individuals, and differ from Nu-Ht1 in a variable region of the D3, it is likely that Nu-Ht2 represents a valid haplotype. Nu-Hts3-7 however, are likely to have resulted from miscoding lesion induced misincorporations. This is because Nu-Hts3-7 are all represented by one individual, which are nearly all ancient, and differ from the common reference (Nu-Ht1) by single base changes that are characteristic of miscoding lesions (Gilbert *et al*. 2007).

6.5.2. Postglacial Colonisation Mode of Canada

After deglaciation had initiated at the end of the last ice age, *A. alpina* colonised Canada from both the East Beringian (region 4) and lower 48 (region 8) refugial areas, based on Mt-Hg1. The west of Hudson Bay (region 5) was primarily colonised from East Beringia, whereas the east of Hudson Bay (region 6) was colonised from the lower 48. These inferences are based on the close association of haplotypes between these region pairs, and the observation that the populations between these regions are not significantly differentiated. Hudson Bay is interpreted as a postglacial longitudinal barrier to dispersal. The areas adjacent to Hudson Bay have only been available for

colonisation for the last 8,000 years (Carlson *et al*. 2008), but these populations (east and west of Hudson Bay) are significantly differentiated, both within Mt-Hg1 and Mt-Hg2, therefore indicating a barrier to dispersal. However, "Hudson Bay" (region 7) displays non-significant differentiation from other populations, which is likely due to small sample size (n=2). Non-significant differentiation was also observed between West and East Beringia (regions 3, 4); this is unsurprising given that these regions were biogeographically connected until around 13,300 cal. yrs BP (England & Furze 2008).

The interpretations of this study are partly in line with the Schwert-Ashworth model of postglacial colonisation of Canada (Schwert & Ashworth 1988), which was expanded upon by Reiss and colleagues (1999) for *A. alpina*. This model states that Beringia was the principal source of colonisers, with colonisation from the lower 48 confined to the northernmost Appalachian Mountains. Based on Mt-Hg1, the results here support the notion that the west of Hudson Bay was colonised from Beringia, which is consistent with the Schwert-Ashworth model. However, the east of Hudson Bay, which included sites as far north as Baffin Island, was found to only have been colonised from the lower 48, contrary to the model. Additionally, as previously discussed and further contrary to the Schwert-Ashworth model, the evidence found here supports the idea that Hudson Bay acted as a barrier to eastward dispersal from Beringia for a species presently found both west and east of Hudson Bay.

6.5.3. Cryptic Diversity in Canada and West Beringia

The presence of Mt-Hg2 in only Canada is intriguing because, despite extensive sampling, individuals from this haplogroup were not found in either of the two classical refugial regions, both at present and, in the case of Beringia, in the past, contrary to the assumption of Reiss and colleagues (1999). This raises the question as to where this haplogroup took refuge during the last North American glaciation. A possible postglacial colonisation event from Europe must be refuted, because all sampled Palearctic individuals fall within Mt-Hg1. Furthermore, the possibility that Mt-Hg2 originated from introgression or hybridisation with a closely related species, resulting in *A. alpina* morphologically, is unlikely as Mt-Hg2 falls within Mt-Hg1 in the mitochondrial phylogeny. Additionally, both haplogroups have identical nuDNA

that lacks any apparent heterozygosity. The possibility that these specimens have been misidentified is considered extremely unlikely, as Mt-Hg2 individuals were sampled from two independent institutions (CNC, NRM), both well known for their taxonomic expertise. Additionally, Reiss and colleagues (1999) found a divergent haplotype, considered in that study to be a putative ancestor of North American *A. alpina*, which has an identical distribution to Mt-Hg2 in this study.

Given the available evidence, four potential scenarios are proposed for the spatial distribution of Mt-Hg2 during the last glaciation. First, is that Mt-Hg2 survived in small nunatak (mountain top) refugia, both near the Hudson Bay region and in the mountains of British Columbia. Low nucleotide diversity in Mt-Hg2, which may indicate a past population bottleneck induced by small refugia, would support this scenario. Second, is that Mt-Hg2 survived in a cryptic northern refugium [*sensu* Stewart *et al*. (2009)] in the High Canadian Arctic. This refugium has been invoked to explain phylogeographic results from studies of several small mammal species (Fedorov & Stenseth 2002; Waltari & Cook 2005). Third, is a combination of the aforementioned scenarios, whereby individuals from the Hudson Bay region survived in the High Canadian Arctic refugium and individuals from the mountains of British Columbia survived on nunataks. A similar combined scenario has been inferred to explain the distribution of ground beetles in the European Alps (Lohse *et al*. 2011). The ecological feasibility of these scenarios is demonstrated by the ability of *A. alpina* to tolerate very cold temperatures (Bennike *et al*. 2000; Sømme 1974), as well as having a varied generalist diet (Braten *et al*. 2012; Chernov 1988; Ottesen 1996). The final scenario is that Mt-Hg2 colonised Canada from the lower 48 refugium and subsequently became extinct in this region. Although subfossil specimens of *A. alpina* are known from lower 48 ice age lowland deposits (Schwert 1992), it is doubtful that aDNA would be preserved due to the assumed high thermal age [*sensu* Smith *et al*. (2003)] of these specimens. The phylogenetic analysis is suggestive of a potential lower 48 origin for Mt-Hg2. However, further analysis of additional markers would be needed to robustly assess this observation.

Two individuals from Anadyr in West Beringia represent Mt-Hg3. This haplogroup is even more distinct than Mt-Hg2, differing substantially (3.35% divergence) from Mt-

Hg1, and also differing in the short section of nuDNA analysed. Morphological examination of these specimens by S.A. Elias revealed that they are not misidentified. Given that Mt-Hg3 is very different from, but also co-occurs with, Mt-Hg1, it is suggested that Mt-Hg3 may represent a cryptic species or subspecies of *A. alpina*. Further work would be required to test this hypothesis robustly.

6.5.4. Genetic Diversity through Time

All ancient individuals of *A. alpina* in this study fall within the crown group. Thus, since the Late Pleistocene in East Beringia there has not been a major genetic diversity loss comparable to that observed in some Beringian megafaunal species, such as bison (*Bison bison*) (Shapiro *et al*. 2004) and musk ox (*Ovibos moschatus*) (Campos *et al*. 2010). The only evidence for diversity loss is two haplotypes composed solely of ancient individuals from the Titaluk River. Assuming that this loss is real, and not the result of non-sampling suitable modern individuals, the existence of individuals of presumed Holocene age (<11,700 cal. yrs BP) in this haplogroup would seem to indicate that this diversity loss was not the result of classical periods of population and species level upheaval: MIS3 (Barnes *et al*. 2002; Barnett *et al*. 2009; Brace *et al*. 2012; Shapiro *et al*. 2004) and the Late Pleistocene to Earliest Holocene (Faith & Surovell 2009; Koch & Barnosky 2006). However, it should be noted that the Titaluk River dates are not well constrained and this inference rests on the TR Holocene assemblage being of Holocene age. Nevertheless, the lack of significant diversity loss in *A. alpina*, would seem to suggest resilience in this species to climatic drivers that devastated the megafauna during MIS3 and the Late Pleistocene to Earliest Holocene. Sampling of other regions from which *A. alpina* subfossils of suitable age and preservation are known, such as Greenland (Böcher 2012; Böcher *et al*. 2012), would be needed to further confirm and extend these inferences, as would more extensive sampling of localities in West Beringia (Kuzmina *et al*. 2011). Investigation of Greenland specimens could be particularly fruitful as *A. alpina* is absent from this region at present (Böcher 2012).

6.6. Conclusions

This study has showcased the utility of a combined modern and ancient DNA approach to assess questions related to the response of an invertebrate taxon to a period of past climatic change. Through this approach, it was found that *A. alpina* is very genetically diverse, with this diversity spread throughout its modern geographic range. Two disparate haplogroups (Mt-Hg2, Mt-Hg3) suggest the possible existence of cryptic clades within *A. alpina* that warrant further investigation. At the end of the last ice age, Mt-Hg1 colonised Canada from both refugial regions (Beringia, lower 48). Individuals from Beringia colonised the region west of Hudson Bay, whereas those from the lower 48 colonised the region east of Hudson Bay. Hudson Bay is thought to have been a barrier to eastward dispersal for *A. alpina*, contrary to previous hypotheses. Mt-Hg2 is hypothesised to have either survived on small nunatak refugia, in a cryptic northern refugium, a combination of these two, or in the lower 48 refugium with subsequent extinction in this region. Without aDNA data, it would not have been possible to determine the absence of Mt-Hg2 in Beringia during the last glaciation. Lastly, the combined approach indicated that there has not been major population turnover in *A. alpina* within Beringia during the past 50,000 years, which is in line with palaeontological data from the beetle subfossil record, but in contrast to the population histories of much of the mammalian megafauna during this time.

6.7. References

Barnes I, Matheus P, Shapiro B, Jensen D, Cooper A (2002) Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science* **295**, 2267-2270.

Barnett R, Shapiro B, Barnes I, *et al*. (2009) Phylogeography of lions (*Panthera leo* ssp.) reveals three distinct taxa and a late Pleistocene reduction in genetic diversity. *Molecular Ecology* **18**, 1668-1677.

Bennike O, Björck S, Böcher J, Walker IR (2000) The Quaternary arthropod fauna of Greenland: a review with new data. *Bulletin of the Geological Society of Denmark* **47**, 111–134.

Böcher J (2012) Interglacial insects and their possible survival in Greenland during the last glacial stage. *Boreas* **41**, 644-659.

Böcher J, Bennike O, Wagner B (2012) Holocene insect remains from south-western Greenland. *Polar Research* **31**, 18367.

Brace S, Palkopoulou E, Dalén L, *et al*. (2012) Serial population extinctions in a small mammal indicate Late Pleistocene ecosystem instability. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 20532–20536.

Braten AT, Flo D, Hagvar S, *et al*. (2012) Primary Succession of Surface Active Beetles and Spiders in an Alpine Glacier Foreland, Central South Norway. *Arctic, Antarctic, and Alpine Research* **44**, 2-15.

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

Carlson AE, Legrande AN, Oppo DW, *et al*. (2008) Rapid early Holocene deglaciation of the Laurentide ice sheet. *Nature Geoscience* **1**, 620-624.

Chernov YI (1988) *The Living Tundra* Cambridge University Press, Cambridge.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Coope GR (2004) Several million years of stability among insect species because of, or in spite of, Ice Age climatic instability? *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **359**, 209-214.

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

Elias SA (2010) *Advances in Quaternary Entomology* Elsevier, Oxford, UK.

Elias SA, Berman D, Alfimov A (2000) Late Pleistocene beetle faunas of Beringia: where east met west. *Journal of Biogeography* **27**, 1349–1363.

England JH, Furze MFA (2008) New evidence from the western Canadian Arctic Archipelago for the resubmergence of Bering Strait. *Quaternary Research* **70**, 60–67.

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

Faith JT, Surovell TA (2009) Synchronous extinction of North America's Pleistocene mammals. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 20641-20645.

Fedorov VB, Stenseth NC (2002) Multiple glacial refugia in the North American Arctic: inference from phylogeography of the collared lemming (*Dicrostonyx groenlandicus*). *Proceedings of the Royal Society of London Series B-Biological Sciences* **269**, 2071-2077.

Gibson CM, Kao RH, Blevins KK, Travers PD (2012) Integrative taxonomy for continental-scale terrestrial insect observations. *PLoS One* **7**, e37528.

Gilbert MTP, Binladen J, Miller W, *et al*. (2007) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Research* **35**, 1-10.

Koch PL, Barnosky AD (2006) Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution and Systematics* **37**, 215-250.

Kuzmina SA, Sher AV, Edwards ME, *et al*. (2011) The late Pleistocene environment of the Eastern West Beringia based on the principal section at the Main River, Chukotka. *Quaternary Science Reviews* **30**, 2091-2106.

Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and Evolution* **29**, 1695-1701.

Lindroth CH (1968) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 5. *Opuscula Entomologica, Supplementum* **33**, 649-944.

Lohse K, Nicholls JA, Stone GN (2011) Inferring the colonization of a mountain range--refugia vs. nunatak survival in high alpine ground beetles. *Molecular Ecology* **20**, 394-408.

Mack L (2008) *Investigating mitochondrial DNA phylogenies of Arctic and European beetle taxa* MSc thesis, Mainz University.

Ottesen PS (1996) Niche segregation of terrestrial Alpine beetles (Coleoptera) in relation to environmental gradients and phenology. *Journal of Biogeography* **23**, 353-369.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Rambaut A (2002) Se-Al (Sequence Alignment Editor).

Raupach MJ, Astrin JJ, Hannig K, *et al*. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* **7**, 1-15.

Raymond M, Rousset F (1995) An exact test for population differentiation. *Evolution* **49**, 1280-1283.

Reiss RA, Ashworth AC, Schwert DP (1999) Molecular genetic evidence for the post-Pleistocene divergence of populations of the arctic-alpine ground beetle *Amara alpina* (Paykull) (Coleoptera: Carabidae). *Journal of Biogeography* **26**, 785-794.

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

Schwert DP (1992) Faunal transitions in response to an Ice Age: The Late Wisconsinan record of Coleoptera in the north-central United States. *The Coleopterists Bulletin* **46**, 68-94.

Schwert DP, Ashworth AC (1988) Late Quaternary History of the Northern Beetle Fauna of North-America - a Synthesis of Fossil and Distributional Evidence. *Memoirs of the Entomological Society of Canada*, 93-107.

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

Smith CI, Chamberlain AT, Riley MS, Stringer C, Collins MJ (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution* **45**, 203-217.

Sømme L (1974) Anaerobiosis in some alpine Coleoptera. *Norsk Entomologisk Tidsskrift* **21**, 155-158.

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Stewart JR, Lister AM, Barnes I, Dalen L (2009) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society of London Series B-Biological Sciences* **277**, 661-671.

Teacher AGF, Griffiths DJ (2011) HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources* **11**, 151-153.

Thomsen PF, Elias S, Gilbert MT, *et al*. (2009) Non-destructive sampling of ancient insect DNA. *PLoS One* **4**, e5048.

Waltari E, Cook JA (2005) Hares on ice: phylogeography and historical demographics of *Lepus arcticus*, *L. othus*, and *L. timidus* (Mammalia: Lagomorpha). *Molecular Ecology* **14**, 3005-3016.

Chapter 7. Postglacial colonisation of Canada by, and taxonomic considerations on, the arctic ground beetle *Pterostichus brevicornis* (Coleoptera: Carabidae)

7.1. Abstract

As a complement to the study in chapter six, the mechanism by which an arctic ground beetle, *Pterostichus brevicornis*, colonised Canada at the end of the last ice age was investigated, utilising genetic information from museum specimens. Over the course of this study, sequence heterogeneity was detected in some individuals, which was used to further examine colonisation patterns. Additionally, the dataset was used to taxonomically assess the two subspecies of *P. brevicornis*: *P. b. brevicornis* and *P. b. delicatus*. Based on mitochondrial DNA data, it is shown that colonising individuals originated from both the western (Beringian) and southern (lower 48 states of the USA) refugia, with the latter being the likely source of colonists for the region to the east of Hudson Bay. However, the sequence heterogeneity data suggest that the true story may be more complex and that further work is required. Mitochondrial DNA evidence does not support the current subspecies designations, although a highly divergent clade was identified in *P. brevicornis*.

7.2. Introduction

In order to examine if closely related beetle taxa have an individualistic or generalised response to past climatic changes, the mode of postglacial colonisation of Canada was investigated in a second ground beetle species. Using genetic evidence from museum specimens, the postglacial colonisation routes of Canada were assessed in the cold-adapted ground beetle *Pterostichus brevicornis* and compared to the Schwert-Ashworth (1988) model (fig. 1.3). Genetic data from ancient individuals were not used here due to the failure of aDNA retrieval from permafrost-preserved specimens (chapter three). This species was suitable for comparable investigation as it is taxonomically close to *A. alpina* (same subfamily: Harpalinae). Furthermore, *P. brevicornis* is currently distributed throughout Beringia and northern Canada, and is

found on mountaintops in the Appalachians (Ball & Currie 1997). During the last ice age, it was present in both the western (Elias *et al*. 2000) and southern (Warner *et al*. 1988) refugia. Although *A. alpina* and *P. brevicornis* are both adapted to tundra habitats, they differ in their ecological requirements (section 1.6).

In addition to assessing the postglacial colonisation of Canada, this study presented an opportunity to assess the taxonomy of *P. brevicornis*. This species is divided into two subspecies, *P. b. brevicornis* and *P. b. delicatus*, based on exoskeletal morphology and modern range (Ball 1966). *P. b. delicatus* is confined to the Bering Islands, whereas the polymorphic *P. b. brevicornis* is found throughout the remainder of the species' distribution [circumpolar from far west Russia to far east Canada (Ball 1966; Ball & Currie 1997; Lindroth 1966)]. Considering that the Bering Islands have only been isolated from mainland Asia and North America for around 13,300 cal. yrs BP (England & Furze 2008), and that morphological 'intergrades' between the two subspecies are known from Alaska (Ball 1966; Lindroth 1966), taxonomic investigation was warranted. The genetic data produced by this study could therefore be used to substantiate the concept of a Bering Island-confined *P. b. delicatus*. Additionally, given the polymorphic nature of *P. b. brevicornis*, this study also provided an opportunity to investigate the existence of cryptic clades in this species.

The primary aims of this study were therefore twofold. First was to use mitochondrial DNA-based regional analysis to assess the postglacial colonisation mode of *P. brevicornis* in Canada, through investigation of whether colonisation of the east of Hudson Bay was primarily from the western (Beringia) or the southern (lower 48 states of the USA) refugia. This included investigating whether Hudson Bay was a barrier to eastward dispersal from Beringia. Second was to assess whether there is genetic support for Bering Island individuals of *P. brevicornis* to be considered a distinct subspecies, and to investigate if there are cryptic clades within *P. brevicornis*. During the course of this study, sequence heterogeneity was noted in part of the mitochondrial dataset. An attempt was made to identify the source of this heterogeneity and to use its presence or absence in samples to further the regional analysis.

7.3. Materials and Methods

7.3.1. Marker and Sample Selection

Only the Sanger sequencing generated COI, COII and 28S datasets were used in this study (table 2.2), due to problems with ITS1 (see section 3.3.6). COI data from Mack (2008) were also incorporated. Phylogenetic and population genetic analyses were employed using the COI and concatenated COI-COII datasets. Samples were considered if >70% (COI: ≥235/336, COI-COII: ≥457/653) of bases had been determined. This was to minimise the chance of missing data resulting in erroneous haplotype designation. Five repeatedly problematic base positions in the COII dataset (see section 7.3.3) were designated missing data for all samples. Altogether, 80 samples were used for downstream analysis (appendix 1).

7.3.2. Phylogenetic and Population Genetic Analyses

Phylogenetic and population genetic analyses followed the methods outlined in sections 6.3.2 and 6.3.3, respectively, with the following modifications. For phylogenetic analysis of the COI dataset, two partitions were selected (first-second and third codon positions) with the HKY substitution model used for both partitions. For the COI-COII dataset, four partitions were selected (first-second and third codon positions for each marker) with the HKY substitution model used the first, second, and fourth partitions, and the F81 model used for the third partition. *P. riparius* (Genbank: EU142584) was used to outgroup the phylogeny. In an attempt to improve resolution, mitochondrial haplogroup (Mt-Hg) 2, the most divergent clade within *P. brevicornis* (figs. 7.1, 7.2), was also utilised as outgroup. However, this did not affect tree topology and was therefore not used.

For the regional analysis, six populations were selected based on five major geographic regions and the Bering Islands (fig. 7.3). The best-supported evolutionary model available in Arlequin was the TrN and TrN+G model for the COI and COI-COII datasets, respectively. Minimum spanning haplotype networks were produced for both datasets, but all subsequent analyses were conducted using the COI-COII dataset only. Mt-Hg2 was excluded from regional analysis, due to its large divergence from the remainder of the samples (Mt-Hg1). Summary statistics were produced, both

within and between populations, and exact tests of population differentiation were used to compare populations based on haplotype composition.

7.3.3. COII Sequence Heterogeneity Characterisation

In the COII dataset, five repeatedly problematic base positions, in the form of sequence mismatches and chromatogram ambiguities, were identified causing COII sequence heterogeneity (CSH). The presence of CSH was determined if one or more of these base positions was found to be problematic for a sample. The occurrence of CSH was characterised genetically and geographically, through comparison with the COI-COII dataset and comparison to longitude, respectively. The COI-COII dataset was split into two groups based on the presence or absence of CSH within samples. An exact test of group differentiation was conducted in Arlequin, to assess if the mitochondrial haplotype compositions differed between CSH and non-CSH groups. Longitude was also scrutinised as a potential factor to explain the geographical distributions of the two groups. A two-tailed Shapiro-Wilk test was used to assess if the data were normally distributed, and significance testing between the populations was conducted using a two-tailed Mann-Whitney U test. These statistical tests were performed in SPSS.

7.3.4. Comparison with nuDNA Data

All available 28S sequences from this thesis for *P. (Cryobius)*, including *P. brevicornis*, were supplemented with data from Genbank (appendix 5b). In order to visualise nuclear haplotype diversity, alignments were constructed in Se-Al, due to the small number of haplotypes in this dataset.

7.4. Results

7.4.1. DNA Sequence Data

Based on the COI dataset, a total of 10 mitochondrial haplotypes were identified (figs. 7.1a, 7.2a), with a maximum sequence divergence of 4.11%. The COI-COII dataset yielded 18 haplotypes (figs. 7.1b, 7.2b), with 3.61% maximum divergence. Two

nuclear haplotypes were identified in the 28S dataset (figs. 7.4), which diverge by 7.51%.

Based on phylogenetic and network-based analyses of the COI-COII dataset (figs. 7.1b, 7.2b), *P. brevicornis* falls into two mitochondrial haplogroups (Mt-Hgs1, 2; defined by criteria in section 6.4.1). Mt-Hg1 consists of 17 haplotypes (2 to 18) from all study regions. Nucleotide diversity within Mt-Hg1 was 0.48% (table 7.1; 0.50% based on the COI dataset). Mt-Hg2 consists of a single haplotype (1), which is found in East Beringia (region 3) only. Nucleotide diversity within Mt-Hg1 populations ranged from 0.00 to 0.64% (table 7.1), with the highest diversity in East Beringia and the lowest in the Bering Islands, east of Hudson Bay, and the Appalachian Mountains (regions 2, 5, 6, respectively). Sequence divergence between regions ranged from 0.00 to 0.94% (regions 5/6, 1/2, respectively; table 7.2). East Beringia and west of Hudson Bay (regions 3, 4) contain nearly all mitochondrial genetic diversity found within *P. brevicornis* (figs. 7.1, 7.2). Haplotypes from the most widely separated populations (West Beringia, Appalachian Mountains) have 0.15% sequence divergence. The Bering Island individuals all belong to the same haplotype (9), which also consists of individuals from East Beringia and the Northwest Territories (NWT).

Region	**Population**	**Data**	**N**	***H***	***h***		**π**	
1	West Beringia	Mt-Hg1	2	2	100.00	(50.00)	0.15	(0.22)
2	Bering Islands	Mt-Hg1	3	1	0.00	(0.00)	0.00	(0.00)
3	East Beringia	Mt-Hg1	29	11	86.70	(3.58)	0.64	(0.36)
4	W Hudson Bay	Mt-Hg1	29	9	74.38	(7.57)	0.34	(0.22)
5	E Hudson Bay	Mt-Hg1	8	1	0.00	(0.00)	0.00	(0.00)
6	Appalachians	Mt-Hg1	7	1	0.00	(0.00)	0.00	(0.00)
3	All Mt-Hg2	Mt-Hg2	2	1	0.00	(0.00)	0.00	(0.00)
1-6	All Mt-Hg1	Mt-Hg1	78	17	85.51	(1.92)	0.48	(0.28)
1, 3-4, 6	CSH	Mt-Hg1	38	13	84.92	(3.55)	0.43	(0.26)
2-6	No CSH	Mt-Hg1	40	9	76.28	(4.46)	0.45	(0.27)

Table 7.1 Molecular diversity summary statistics of *P. brevicornis*, based on 336 and 317bps of COI and COII, respectively. Regions correspond to fig. 7.3. Standard deviations are in parentheses. N: individuals, *H*: haplotypes, *h*: haplotype diversity (%), π: nucleotide diversity/within population sequence divergence (%), W: west of, E: east of, CSH: COII sequence heterogeneity.

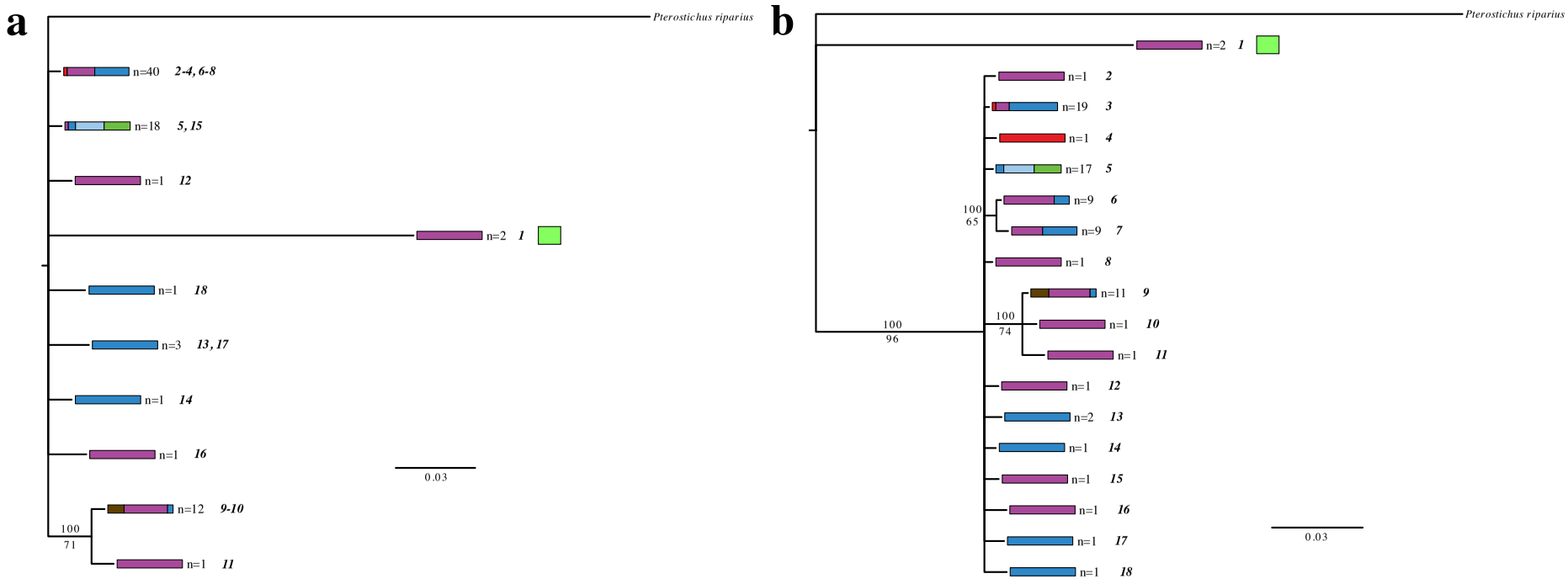


Figure 7.1 Bayesian mitochondrial phylogenies of *P. brevicornis* based on, a: 336bps of COI, and b: 336 and 317bps of COI and COII, respectively. Terminals are haplotypes. Coloured bars denote the regional population(s) from which haplotypes originated and are based on fig. 7.3. Branch supports are Bayesian posterior probabilities (above) and maximum likelihood bootstraps (below). n: individuals. Haplotype numbers are in bold-italic.

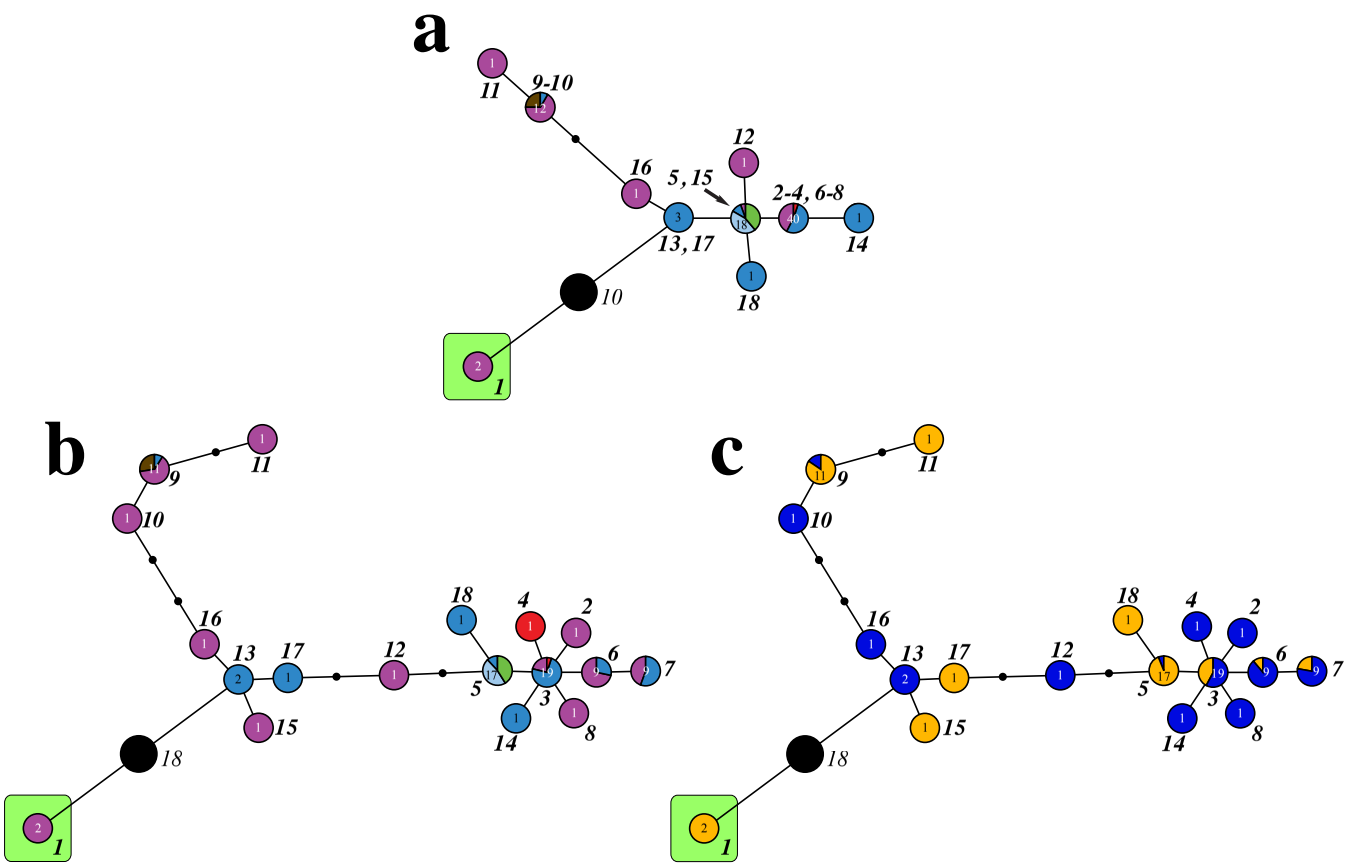


Figure 7.2 Mitochondrial haplotype networks of *P. brevicornis* constructed from, a: 336bps of COI, and b, c: 336 and 317bps of COI and COII, respectively. Within-circle colours for a and b correspond to geographic locality, whereas those for c correspond to presence (royal blue) or absence (amber) of COII sequence heterogeneity (fig. 7.3). Green boxes indicate Mt-Hg2, all other haplotypes belong to Mt-Hg1. Small black circles indicate missing haplotypes, and branches with >10 missing haplotypes are indicated by large black circles (exact number in italic next to circle). Numbers within circles indicate individuals. Haplotype numbers outside circles are in bold-italic.

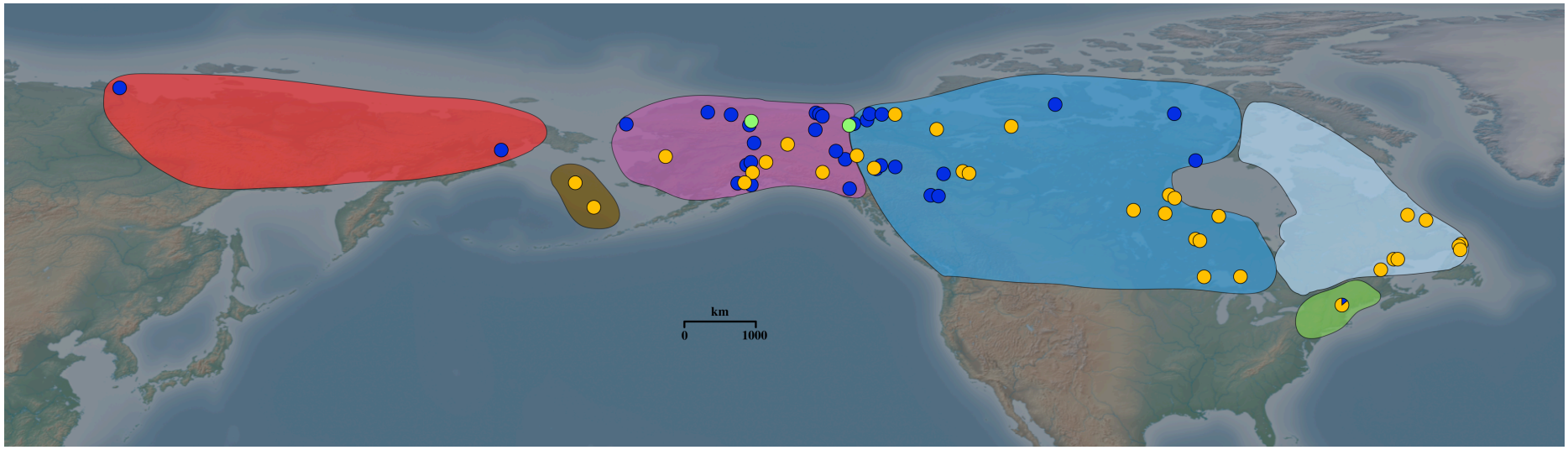


Figure 7.3 Geographic distribution of *P. brevicornis* individuals investigated. Within-circle colour corresponds to the presence (royal blue) or absence (amber) of COII sequence heterogeneity. The two Mt-Hg2 individuals (both with absent COII sequence heterogeneity) are indicated with green; all other individuals belong to Mt-Hg1. Geographic regions are indicated as (from left to right): region 1 (West Beringia; red), region 2 (Bering Islands; brown), region 3 (East Beringia; purple), region 4 (west of Hudson Bay; blue), region 5 (east of Hudson Bay; light blue), region 6 (Appalachian Mountains; green).

Region		1	2	3	4	5	6
	Population	**West Beringia**	**Bering Islands**	**East Beringia**	**W Hudson Bay**	**E Hudson Bay**	**Appalachians**
1	West Beringia	-	0.94	0.15	0.03	0.15	0.15
2	Bering Islands	0.100 (0.001)	-	0.40	0.78	0.78	0.78
3	East Beringia	0.209 (0.003)	0.675 (0.003)	-	0.06	0.17	0.17
4	W Hudson Bay	0.420 (0.004)	**0.021 (0.001)**	**0.002 (<0.001)**	-	0.10	0.10
5	E Hudson Bay	**0.022 (<0.001)**	**0.006 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	-	0.00
6	Appalachians	**0.028 (<0.001)**	**0.008 (<0.001)**	**<0.001 (<0.001)**	**<0.001 (<0.001)**	1.000 (1.000)	-

Region		3	3	1, 3-4, 6	2-6
	Population	**Mt-Hg1**	**Mt-Hg2**	**CSH**	**No CSH**
1-6	Mt-Hg1	-	3.61	-	-
3	Mt-Hg2	**0.017 (<0.001)**	-	-	-
1, 3-4, 6	CSH	-	-	-	0.07
2-6	No CSH	-	-	**<0.001 (<0.001)**	-

Table 7.2 Exact test of population differentiation and sequence divergence results for populations of *P. brevicornis*. Below diagonal: exact test *p*-values, with standard deviations in parentheses. Significant results ($p<0.05$) are in bold. Above diagonal: percentage sequence divergence between populations. Top: Data on populations within Mt-Hg1, with regions corresponding to fig. 7.3. Bottom: Data comparing the two mitochondrial haplogroup populations, and groups based on the presence or absence of COII sequence heterogeneity (CSH) in Mt-Hg1. W: west of, E: east of, CSH: COII sequence heterogeneity.

In the 28S dataset (fig. 7.4), 21 of the tested *P. brevicornis* samples (n=22) belonged to a single nuclear haplotype (Nu-Ht1). This haplotype consists of individuals belonging to Mt-Hg1 and includes the three individuals from the Bering Islands. The Mt-Hg2 individual analysed for 28S composes another distinct haplotype (Nu-Ht2), which differs by eight inserted and five substituted nucleotides. The majority (84.6%) of these differences occur within a variable section of the 28S D3 region. The insertions found within Nu-Ht2 are also found in *P. (Cryobius) nemoralis*, whereas the four other assessed species of *P. (Cryobius)* lack these insertions and therefore have a fragment length identical to Nu-Ht1.

7.4.2. Regional Analysis of mtDNA Data

The Beringian regions (1 to 3), which include the Bering Islands, are not significantly differentiated from each other (table 7.2). The west of Hudson Bay region (4) is significantly differentiated from all other regions, except West Beringia (region 1). The east of Hudson Bay and Appalachian Mountain regions (5, 6) are not differentiated, but both of these regions differ significantly from the remaining populations in their haplotype compositions.

7.4.3. Characterisation of COII Sequence Heterogeneity

CSH was detected in 48.7% of Mt-Hg1 individuals, but was absent from both individuals of Mt-Hg2 (fig. 7.2c). A total of 70 nucleotides, out of a total of 400, were found to be problematic across the five base positions (table 7.3). These five base positions spanned a 139bp region of COII. Four of the five problematic sites were third codon positions, which remained synonymous regardless of base variant. Genetic characterisation of CSH revealed that the CSH and non-CSH groups have a significantly different composition of mitochondrial haplotypes (table 7.2), although clusters of haplotypes with CSH occur throughout the haplotype network (fig. 7.2c), resulting in low sequence divergence between the two groups (0.07%). Longitude-based geographical characterisation of CSH revealed a significant difference in the distribution of CSH and non-CSH individuals (Two-tailed Mann-Whitney: U=359, n_1=40, n_2=38, p<0.001). CSH individuals tend to be found in the western regions (Beringia), whereas those without CSH are usually found in eastern regions (Canada

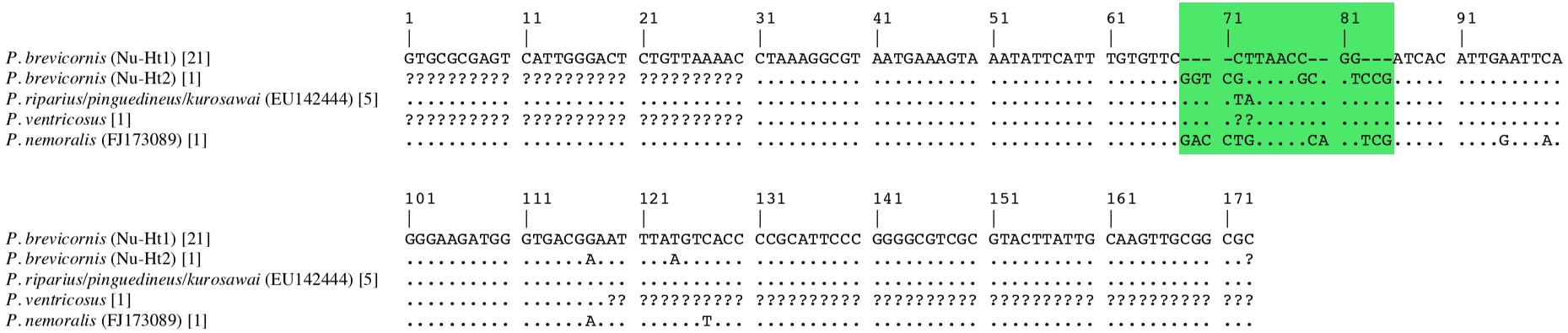


Figure 7.4 Alignment of the two nuDNA haplotypes of *P. brevicornis*, based on 164 to 172bps of the D3 region of 28S rDNA. The alignment is supplemented with two sequences from Genbank and *P. ventricosus* from this thesis. Parentheses indicate either haplotype designation or Genbank accession number. Square brackets indicate number of individuals. Area highlighted in green is a highly variable section of the D3 region identified in *P. (Cryobius)*. ?: missing data, .: identical to reference sequence, -: gap.

and the Appalachian Mountains). The two groups were not normally distributed, based on longitude (Two-tailed Shapiro-Wilk: CSH: W=0.857, d.f.=38, p<0.001; non-CSH: W=0.915, d.f.=40, p=0.005).

		Problematic Position						
Mt-Ht	**N**	**3460**	**3500**	**3529**	**3592**	**3598**	**CSH**	**Proportion**
1	2	0	0	0	0	0	0	0.00
2	1	0	0	1	1	0	1	1.00
3	19	1	7	11	6	2	11	0.58
4	1	0	0	1	1	0	1	1.00
5	17	0	1	1	0	0	1	0.06
6	9	1	0	8	4	4	8	0.89
7	9	0	0	7	1	2	7	0.78
8	1	0	0	1	0	0	1	1.00
9	11	2	0	0	0	0	2	0.18
10	1	1	0	0	0	0	1	1.00
11	1	0	0	0	0	0	0	0.00
12	1	1	0	0	0	0	1	1.00
13	2	2	0	0	0	0	2	1.00
14	1	0	1	1	0	0	1	1.00
15	1	0	0	0	0	0	0	0.00
16	1	1	0	0	0	0	1	1.00
17	1	0	0	0	0	0	0	0.00
18	1	0	0	0	0	0	0	0.00
Total	80	9	9	31	13	8	38	0.48
Codon position		3	1	3	3	3		
Nucleotides		C-T	C-T	C-T	A-T	C-T		
Amino acids		Tyr-Tyr	Pro-Ser	Phe-Phe	Val-Val	Ala-Ala		

Table 7.3 Problematic base positions resulting in sequence heterogeneity in the COII dataset (CSH). Base position was established from the *Calosoma* sp. mitogenome (Genbank: NC018339). Mitochondrial haplotypes (Mt-Hts) correspond to figs. 7.1b, 7.2b, c. N: Number of individuals per haplotype. CSH: Number of individuals with COII sequence heterogeneity.

7.5. Discussion

7.5.1. Genetic Diversity of *P. brevicornis*

Given the entire COI dataset, the genetic diversity of *P. brevicornis* (4.11%) is greater than typical intraspecific sequence divergence estimates of ground beetle COI [0 to 3.8% (Gibson *et al*. 2012; Raupach *et al*. 2010)]. However, when excluding the divergent Mt-Hg2, this figure is reduced eightfold (0.50%). Sequence divergence estimates were similar in the COI and COI-COII datasets, which would suggest a comparable mutation rate between these two markers in this species.

The vast majority of *P. brevicornis* mitochondrial genetic diversity is found in the west of Hudson Bay and East Beringian regions. With the exception of the Bering Island population, all haplotypes from the remaining populations differed by only two substitutions, although the West Beringian population was poorly sampled (n=2). As with *A. alpina* (section 5.5.1), this would indicate that significant migration between regions has occurred, as suggested by the beetle fossil record (Elias 2010). The two divergent nuclear haplotypes identified in *P. brevicornis* are consistent with the mtDNA data, with the two nuDNA haplotypes represented by individuals from Mt-Hgs1 and 2, respectively.

7.5.2. Regional Analysis of *P. brevicornis*

The mtDNA data suggest that the region east of Hudson Bay was colonised from the lower 48 states of the USA after deglaciation initiated. This inference derives from the observation that only a single haplotype is shared between the east of Hudson Bay and Appalachian Mountain populations, the latter of which was likely derived from individuals south of the Laurentide ice sheet (Schwert & Ashworth 1988). If colonisation of the east of Hudson Bay was by populations originating from East Beringia, as suggested by the Schwert-Ashworth (1988) model, it would be expected that the haplotype composition of the east of Hudson Bay population would be closer to the East Beringian and west of Hudson Bay populations rather than the Appalachian Mountains population. However, the east of Hudson Bay population is significantly differentiated from both of these western populations, which would also suggest that Hudson Bay was a longitudinal barrier to eastward dispersal.

The west of Hudson Bay population was probably colonised from both the lower 48 states of the USA and East Beringia, therefore resulting in significant differentiation from either of these refugial regions. Additional evidence for this interpretation is that individuals from the west of Hudson Bay population belong to haplotypes found in East Beringia and the Appalachian Mountains. Non-significant differentiation between the west of Hudson Bay and West Beringian populations is probably an artifact of the small sample size of the latter (n=2). Even though both the West Beringian and the Bering Island populations had a small sample size, all three Beringian regions were found to be non-significantly differentiated. This is likely due

to these regions being biogeographically connected until around 13,300 cal. yrs BP (England & Furze 2008).

7.5.3. Characterisation of COII Sequence Heterogeneity

Based on the sequence data, the source of CSH was tentatively inferred as either mitochondrial heteroplasmy or NUMT contamination, through the elimination of other potential sources. Misincorporations due to miscoding lesions or other sources of error (enzymatic, sequencing) as the sole cause of CSH was considered unlikely. This is because only five base positions were repeatedly problematic and, although misincorporation 'hotspots' are known to occur (Gilbert *et al*. 2005; Gilbert *et al*. 2003), it would be suspicious for these to occur within a 139bp sequence and yet be absent from the remaining dataset. The possibility of sample cross-contamination being the source of CSH is considered unlikely because, in addition to rigorous enforcement of negative controls and contamination reduction measures, it would be expected for the COI and the remainder of the COII datasets to also display sequence heterogeneity. Given these considerations, a NUMT would be the most parsimonious explanation for the source of CSH. However, strong support for NUMT contamination, such as the presence of stop codons or frameshifting indels (Bensasson *et al*. 2001), were not detected. In addition, four of the five problematic bases were synonymous third codon substitutions, which would be more indicative of function-retaining heteroplasmy. This has been inferred to explain similar results in other recent insect mtDNA studies (Frey & Frey 2004; Magnacca & Brown 2010; Sword *et al*. 2007). However, a problem with this inference is that the problematic bases only occurred in a small section of the COI-COII dataset, whereas it may be expected that similar apparent problems would arise in other dataset regions (although see section 3.5.5). Alternatively, a NUMT may have arisen relatively recently and, assuming a higher mutation rate in the mtDNA, the observed differences are due to subsequent substitutions in the mtDNA (Sunnucks & Hales 1996). Based on the sequence data alone therefore, neither mitochondrial heteroplasmy nor NUMT contamination could be confidently identified as the CSH source.

Despite the conservative approach of assigning CSH status to individuals if only a single base was found to be problematic, CSH was only detected in around half of

individuals. It is possible that, due to differential DNA preservation or the stochastic nature of the PCR, CSH was present in some individuals but was not detected. Comparison of the mtDNA haplotype composition between the CSH and non-CSH groups revealed that the composition of these groups was significantly different, with apparent clustering of the groups within the haplotype network. If heteroplasmic mitochondria were the CSH source, and were transmitted maternally, then it would be expected that the CSH and mtDNA data would be congruent. This could therefore be used to argue for a NUMT being the CSH source. Interestingly, there is a geographical pattern in the distribution of the CSH and non-CSH groups, with the latter being more common in the regions both west and east of Hudson Bay, as well as in the Appalachian Mountains. These distributions contrast to the regional analysis based on mtDNA, and hint that the population history of *P. brevicornis* may be more complex than suggested by the mtDNA sequence data alone. Therefore, further work employing nuclear DNA markers would be required to assess any regional analysis comprehensively, and may also resolve the source of the CSH.

7.5.4. Taxonomic Considerations

The nuDNA fragment was too conserved to scrutinise the subspecific status of the Bering Islands individuals. However, based on mtDNA evidence, the fact that all three individuals from the Bering Islands compose a haplotype that also occurs in East Beringia and the NWT fails to support Ball's (1966) concept of *P. b. delicatus* as a distinct subspecies confined to the Bering Islands. Interestingly, this haplotype (9), along with two others (10, 11), form a well-supported clade that diverges near the base of *P. brevicornis* (excluding Mt-Hg2). Whether this clade of three haplotypes represents a distinct operational taxonomic unit is only hinted at within the present study and would require further investigation of both mitochondrial and nuclear DNA markers.

Two individuals from East Beringia (Mt-Hg2/Nu-Ht2; *Pb*2) were found to be highly divergent from the remaining individuals of *P. brevicornis* (Mt-Hg1/Nu-Ht1; *Pb*1). The mtDNA sequence divergence between these two groups is beyond the intraspecific range reported from ground beetle COI. Together with the high divergence at the nuclear 28S locus, it is doubtful that *Pb*2 should be considered *P.*

brevicornis. The two individuals that compose *Pb*2 were collected by different taxonomists on different collection trips, identified by expert taxonomists upon entry into the CNC collection, and further identified by S.A. Elias during specimen sampling. The possibility that *Pb*2 are an artifact of specimen misidentification is therefore considered highly unlikely. Although introgressive hybridisation events are known in carabid beetles (Sasakawa & Kubota 2005; Sota & Vogler 2001; Streiff *et al*. 2005; Zhang *et al*. 2005), the fact that *Pb*1 and *Pb*2 are very divergent at both mitochondrial and nuclear loci does not support this. This also suggests that the *Pb*2 mtDNA sequence is not an artifact of NUMT contamination. Furthermore, if *Pb*2 were the result of a very recent hybridisation event between *P. brevicornis* and another species, potential heterozygote hybrids may be expected to have a poor to unreadable 28S alignment, due to the sequence length difference between *Pb*1 and *Pb*2. However, this was not observed in the chromatograms. It is therefore suggested that *Pb*2 may represent a cryptic clade, a phenomenon that has previously been identified in *Pterostichus* (Sasakawa & Kubota 2005). Assuming that the insertions in the 28S sequences are homologous, this may indicate a closer phylogenetic relationship of *Pb*2 to *P. (Cryobius) nemoralis*, a species that has also been assigned to the subgenus *Argutor* in some literature (Bousquet 2003), than to the other four species of *P. (Cryobius)* under study. The mitochondrial and especially nuclear DNA data presented here suggest that *P. (Cryobius)* and allied subgenera would benefit from a comprehensive assessment of phylogeny and the prevalence of potential cryptic clades.

7.6. Conclusions

This study has assessed key questions related to the North American population history and taxonomy of the arctic ground beetle *P. brevicornis*, as inferred from a museum DNA based approach. *P. brevicornis* is genetically diverse, with the majority of this diversity found in East Beringia and the west of Hudson Bay. *P. brevicornis* consists of two distinct clades, differing at mitochondrial and nuclear DNA markers. *Pb*1 is found throughout the entire range of *P. brevicornis* and the geographic distribution of its mtDNA haplotypes suggest that postglacial colonisation was from both the Beringian and southern refugia with Hudson Bay acting as a barrier to eastward dispersal from Beringia. Mitochondrial COII sequences displayed

heterogeneity in around half of the individuals sampled, which probably resulted from either mitochondrial heteroplasmy or NUMT contamination. Two groups, based on the presence or absence of sequence heterogeneity, had differing compositions of mitochondrial haplotypes and were geographically distinct, with samples displaying heterogeneity being more common in Beringia and northwestern Canada. Based on mtDNA evidence, the concept of a Bering Island-confined subspecies *P. b. delicatus* was not supported, due to this haplotype being found in East Beringia and the Northwest Territories. The divergent *Pb*2 potentially does not belong to *P. brevicornis*, and may have closer affinities to other species of *P. (Cryobius)*, which justifies a comprehensive phylogenetic assessment or this and closely related subspecies.

7.7. References

Ball GE (1966) A revision of the North American species of the subgenus *Cryobius* Chaudoir (*Pterostichus*, Carabidae, Coleoptera). *Opuscula Entomologica Supplementum* **28**, 1-166.

Ball GE, Currie DC (1997) Ground Beetles (Coleoptera: Trachypachidae and Carabidae) of the Yukon: Geographical Distribution, Ecological Aspects, and Origin of the Extant Fauna. In: *Insects of the Yukon* (eds. Danks HV, Downes JA), pp. 445-489. Biological Survey of Canada (Terrestrial Arthropods), Ottawa.

Bensasson D, Zhang D, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**, 314-321.

Bousquet Y (2003) Tribe Pterostichini Bonelli, 1810. In: *Catalogue of Palaearctic Coleoptera, Volume 1: Archostemata–Myxophaga–Adephaga* (eds. Löbl I, Smetana A), pp. 469–520. Apollo Books, Stenstrup.

Elias SA (2010) *Advances in Quaternary Entomology* Elsevier, Oxford, UK.

Elias SA, Berman D, Alfimov A (2000) Late Pleistocene beetle faunas of Beringia: where east met west. *Journal of Biogeography* **27**, 1349–1363.

England JH, Furze MFA (2008) New evidence from the western Canadian Arctic Archipelago for the resubmergence of Bering Strait. *Quaternary Research* **70**, 60–67.

Frey JE, Frey B (2004) Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? *Hereditas* **140**, 92-98.

Gibson CM, Kao RH, Blevins KK, Travers PD (2012) Integrative taxonomy for continental-scale terrestrial insect observations. *PLoS One* **7**, e37528.

Gilbert MTP, Shapiro B, Drummond A, Cooper A (2005) Post-mortem DNA damage hotspots in Bison (*Bison bison*) provide evidence for both damage and mutational hotspots in human mitochondrial DNA. *Journal of Archaeological Science* **32**, 1053-1060.

Gilbert MTP, Willerslev E, Hansen AJ, *et al*. (2003) Distribution patterns of postmortem damage in human mitochondrial DNA. *American Journal of Human Genetics* **72**, 32-47.

Lindroth CH (1966) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 4. *Opuscula Entomologica, Supplementum* **29**, 409-648.

Mack L (2008) *Investigating mitochondrial DNA phylogenies of Arctic and European beetle taxa* MSc thesis, Mainz University.

Magnacca KN, Brown MJF (2010) Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus (Nesoprosopis)* bees (Hymenoptera: Colletidae). *BMC Evolutionary Biology* **10**, 174.

Raupach MJ, Astrin JJ, Hannig K, *et al*. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* **7**, 1-15.

Sasakawa K, Kubota K (2005) Cryptic species of the subgenus *Morphnosoma* Lutshnik (Coleoptera: Carabidae; genus *Pterostichus*) from Japan. *Entomological Science* **8**, 389–404.

Schwert DP, Ashworth AC (1988) Late Quaternary History of the Northern Beetle Fauna of North-America - a Synthesis of Fossil and Distributional Evidence. *Memoirs of the Entomological Society of Canada*, 93-107.

Sota T, Vogler AP (2001) Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus*. *Systematic Biology* **50**, 39-59.

Streiff R, Veyrier R, Audiot P, Meusnier S, Brouat C (2005) Introgression in natural populations of bioindicators: a case study of *Carabus splendens* and *Carabus punctatoauratus*. *Molecular Ecology* **14**, 3775-3786.

Sunnucks P, Hales DF (1996) Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Molecular Biology and Evolution* **13**, 510-524.

Sword GA, Senior LB, Gaskin JF, Joern A (2007) Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera : Acrididae). *Systematic Entomology* **32**, 420-428.

Warner BG, Morgan AV, Karrow PF (1988) A Wisconsinan Interstadial arctic flora and insect fauna from Claksburg, southwestern Ontario, Canada. *Palaeogeography, Palaeoclimatology, Palaeoecology* **68**, 27-47.

Zhang AB, Kubota K, Takami Y, *et al*. (2005) Species status and phylogeography of two closely related *Coptolabrus* species (Coleoptera: Carabidae) in South Korea inferred from mitochondrial and nuclear gene sequences. *Molecular Ecology* **14**, 3823-3841.

Chapter 8. Discussion and Conclusions

8.1. General Discussion

8.1.1. Potential of DNA from Dry-stored Museum Beetles

The largest and most comprehensive study of the potential of DNA from dry-stored insect specimens, and the first related to beetles, was conducted. Endogenous DNA was recovered from nearly all specimens of both study species (*Amara alpina*, *Pterostichus brevicornis*), including from >100yr specimens (table 3.1). A Sanger sequencing based approach demonstrated that the maximum amplifiable fragment length of mitochondrial (mt) DNA decreases with age (fig. 3.3), and that fragments of length useful for next-generation sequencing (NGS) should be retrievable from >200yr specimens. PCR amplification success was lower in >100yr specimens (fig. 3.1), which may be due to a lower endogenous DNA concentration in these specimens. However, using a shotgun NGS-based approach, it was shown that a >100yr specimen can preserve an amount of endogenous DNA comparable to <10yr specimens (tables 2.3, 4.2; fig. 5.1). This is suggestive of large between-specimen variability. Fragment length distributions broadly revealed that nuclear (nu) DNA had a shorter mean fragment length than mtDNA, indicating that mtDNA may be preferentially preserved. The rate of base misincorporation was greater (~5%), but still relatively low, in a >100yr specimen when compared to <100yr specimens (~2%) although, contrary to other degraded DNA studies [e.g. Briggs *et al*. (2007); Orlando *et al*. (2011); Rasmussen *et al*. (2011)], misincorporations did not accumulate at read termini (fig. 4.3). The proportion of DNA preserved in dry-stored museum beetle specimens that was considered endogenous (~25 to 40%; fig. 5.1), did not seem to be related to specimen age or collection of origin, indicating that the proportion of contaminants does not increase with age. Interestingly, it was also possible to recover sequences of arthropod parasites and commensals (fig. 5.1; table 5.2), which could potentially be used to examine these associations through time and space using museum DNA (Tsangarasa & Greenwood 2012).

Altogether, these data are congruent with other small-scale studies of the preservation of museum insect DNA (Andersen & Mills 2012; Strange *et al*. 2009; Ugelvig *et al*. 2011; van Houdt *et al*. 2010; Watts *et al*. 2007) and demonstrate the immense potential for museum beetle specimens in future genetic investigation. However, studies that require a NGS-based approach would benefit from having a suitable reference genome, as this would increase the breadth of identifiable sequences in these large datasets. International initiatives, such as the 5,000 insect genomes project (i5k; http://www.arthropodgenomes.org/wiki/i5K), are anticipated to provide such reference genomes over the next five years.

8.1.2. Potential of DNA from Ancient Permafrost-preserved Beetles

The first assessment of the potential of ancient (a) DNA from permafrost-preserved beetles was performed on two study species: *A. alpina* and *P. brevicornis*. Using a Sanger sequencing based approach, aDNA could be recovered from ~45% of *A. alpina* specimens, but could not be recovered from any specimens of *P. brevicornis* (table 3.1). This suggests that the successful recovery of aDNA may be dependent on the species under study. However, additional species would need to be investigated for verification of this inference. Around half of the *A. alpina* specimens that yielded aDNA had a high amplification success, whereby the majority of targeted fragments could be retrieved (fig. 3.2). This success was not related to specimen age or local preservation conditions, as measured by locality (figs. 3.2, 3.4). Using a NGS-based approach, two ancient *A. alpina* specimens were shown to have up to an order of magnitude less endogenous DNA than two museum specimens (table 4.2). NuDNA had a shorter mean fragment length than mtDNA, based on fragment length distributions (fig. 4.2), which was consistent with data from museum specimens. These data suggest that the preferential preservation of mtDNA may be a more general feature of degraded DNA (Allentoft *et al*. 2012; Schwarz *et al*. 2009). The base misincorporation rate was greater in ancient specimen DNA (~5 to 10%; fig. 4.3) than in DNA from museum specimens, although misincorporations were spread throughout the read length, in agreement with the museum DNA data, but contrary to previous studies of degraded DNA [e.g. Briggs *et al*. (2007); Orlando *et al*. (2011); Rasmussen *et al*. (2011)]. Less than 0.5% of DNA in ancient permafrost-preserved specimens was considered to be endogenous (fig. 5.1). Instead, nearly all of the DNA

present in these specimens was likely to have derived from bacteria of the preservational environment. This suggests that surface contamination may be a significant source of DNA extracted from these remains and that a shotgun-based approach may not be the optimal method for retrieving endogenous aDNA from these remains.

There are two methods that could be considered as ways to increase the amount of endogenous DNA in future NGS-based investigations of ancient beetles: specimen decontamination and endogenous DNA enrichment. The surface of specimens could be decontaminated, through sterilisation with bleach, prior to DNA extraction (Campos & Gilbert 2012). This approach would only be useful if the main source of aDNA were within the sclerite, with the aDNA therefore being protected by the chitinous cuticle. However, this situation is considered unlikely because the insect cuticle is acellular, being formed by secretions from epidermal cells (Gullan & Cranston 2010; Moussian 2010; Vincent & Wegst 2004). It is plausible that small amounts of cellular material, including DNA, may become incorporated during cuticle formation, but whether this would preserve enough DNA to be detectable in ancient remains seems doubtful. A more likely source of the aDNA may be the underside of the sclerite, to which desiccated epidermal cells or haemolymph may be attached. If the external surface is the source of aDNA, then sterilising the specimen with bleach would be likely to remove both endogenous and contaminant DNA. Investigation into the source of endogenous DNA in ancient permafrost-preserved specimens would therefore be required for the optimisation of DNA extraction procedures.

Alternatively, it may be possible to enrich the DNA extract for endogenous DNA prior to NGS. Such an approach, using target-capture based or other enrichment methods, has been applied to DNA extracted from ancient mammals and plants that had a small proportion of endogenous aDNA (Avila-Arcos *et al*. 2011; Briggs *et al*. 2009; Burbano *et al*. 2010; Stiller *et al*. 2009). Capture-based methods have varied in their success, with some enriching aDNA by four to five orders of magnitude (Briggs *et al*. 2009; Burbano *et al*. 2010) and others enriching by only 4 to 29-fold (Avila-Arcos *et al*. 2011). The discrepancy between these studies may be due to experimental differences in the number of capture and PCR reactions performed (Avila-Arcos *et al*.

2011). A caveat to this technique is that baseline knowledge of the genetic regions to be enriched may be required for either the target species or a closely related taxon. However, new systems have recently been developed that reduce this limitation by allowing 'cross-species' capture and enrichment (Mason *et al*. 2011). Therefore, there may be promise in applying such an approach to beetle aDNA.

Ancient DNA could only be recovered routinely from a minority of specimens using Sanger sequencing, suggesting that DNA preservation was poor in most specimens. NGS revealed that ancient specimens have a much lower amount of endogenous DNA compared to museum specimens. This result may be due to an overwhelming amount of bacterial DNA outcompeting endogenous DNA, and could be a consequence of the shotgun-based approach used in this study. However, if capture-based enrichment methods can be successfully applied to endogenous aDNA from ancient beetle specimens, then beetle aDNA may become viable for large-scale investigation.

8.1.3. Postglacial Colonisation Mode of Canada

Regional analyses, based on mtDNA haplotype differentiation between geographically and temporally defined populations, were employed to test the Schwert-Ashworth (1988) model (fig. 1.3) of the postglacial colonisation mode of Canada by *A. alpina* (chapter six) and *P. brevicornis* (chapter seven). The model proposed that Canada, both west and east of Hudson Bay, was primarily colonised from the Beringian refugium. Based on the evidence found here (summarised in fig. 8.1), the region west of Hudson Bay seems to have been colonised from both Beringia (primarily, in the case of *A. alpina*) and the lower 48 states of the USA, which is broadly in line with the Schwert-Ashworth (1988) model, but suggests that northward colonisation was more important than the model implied. However, the evidence suggests that Hudson Bay was a barrier to eastward dispersal from Beringia in both of the study species, with individuals found east of Hudson Bay colonising from the lower 48. This suggests that Hudson Bay was a stronger barrier to dispersal than suggested by the Schwert-Ashworth (1988) model, and may have implications for future migration patterns of the North American beetle fauna in response to predicted future climatically driven changes. *A. alpina* and *P. brevicornis* differ in their ecologies - as *A. alpina* is adapted to drier environments than *P. brevicornis* (Elias & Crocker 2008;

Lindroth 1968) - but exhibit a broadly similar mode of Canadian postglacial colonisation (fig. 8.1). This may imply that the response of these two species was broadly representative of the arctic ground beetle fauna.

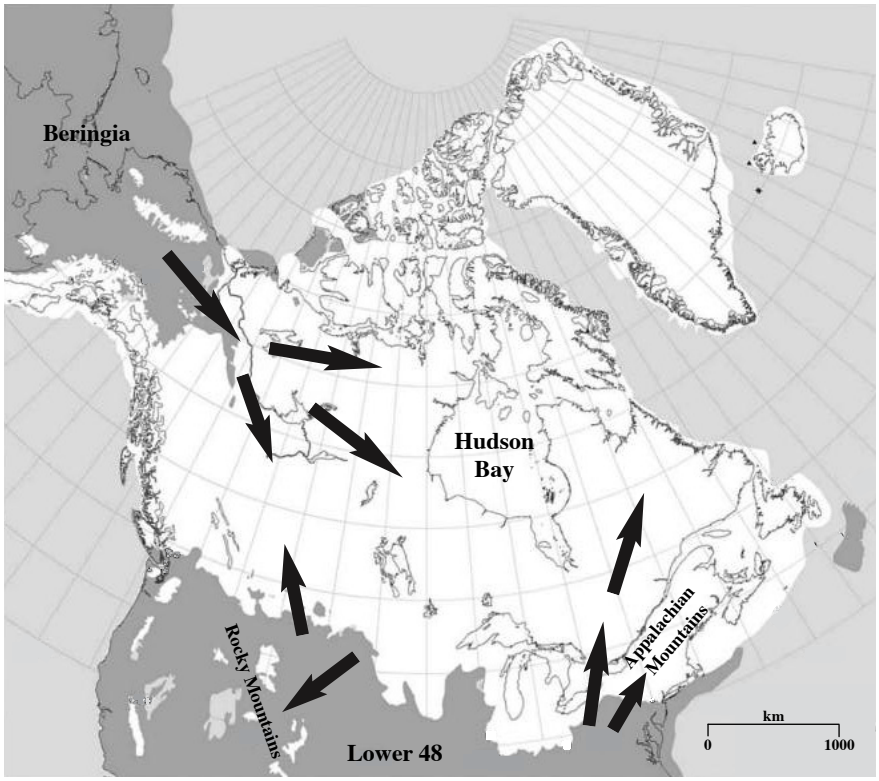


Figure 8.1 The modified Schwert-Ashworth (1988) model of the postglacial colonisation mode of Canada by the arctic beetle fauna, as exemplified by *Amara alpina* and *Pterostichus brevicornis*. Western refugium: Beringia, southern refugium: lower 48 of the USA, dispersal barrier: Hudson Bay. White: extent of the Laurentide and Cordilleran ice sheets at 18ka, just after maximal extent. Light grey: water bodies. Dark grey: exposed land. Map modified from Dyke (2004). Colonisation routes are based on data from this thesis (chapters six, seven).

8.1.4. Potential for Demographic Inference from Beetle mtDNA

Sophisticated coalescence-based phylogeographic and demographic analyses [e.g. (Anderson *et al*. 2005; Drummond & Rambaut 2007; Lemey *et al*. 2010)] have been used to infer the effects of past processes (such as climatically driven change) on components of the ice age fauna, using both modern and ancient mtDNA (Campos *et al*. 2010a; Campos *et al*. 2010b; Marske *et al*. 2011; Shapiro *et al*. 2004). Furthermore, these analyses have been used to estimate divergence dates between clades (Barnes *et al*. 2007). However, these analyses were not performed here. The

molecular clock of both species could not be calibrated, as no suitable comparable rate estimate could be identified. Additionally, the ages of ancient *A. alpina* individuals were not well constrained and so direct internal calibration of the molecular clock, using ancient sequences (Drummond *et al*. 2003), could not be performed.

The beetles have additional potential complicating factors for mtDNA-based phylogeographic and demographic analyses, some of which were encountered in this thesis. Mitochondrial heteroplasmy has been identified in a variety of insect taxa (Frey & Frey 2004; Magnacca & Brown 2010; Sword *et al*. 2007) and may violate the assumption of mtDNA sequence being of clonal, maternal descent, especially if the source is repeated paternal leakage (Galtier *et al*. 2009). Moreover, maternally inherited reproductive endosymbionts, such as *Wolbachia*, can cause problems for reconstructing demography, due to linkage disequilibrium with mtDNA (Hurst & Jiggins 2005). *Wolbachia* infections can induce selective sweeps, which indirectly select for mtDNA that hitchhikes with the endosymbiont (Hurst & Jiggins 2005). If a selective sweep occurred, the assumption of mtDNA being evolutionary neutral would be violated (Galtier *et al*. 2009), and the sweep could potentially be misinterpreted as another demographic process, such as a population bottleneck (Hurst & Jiggins 2005). For future studies wanting to employ phylogeographic and demographic analyses from modern and ancient beetles, these issues would need to be addressed through the examination of markers from mtDNA, nuDNA, and any endosymbionts present.

8.1.5. Status of Morphologically Cryptic Groups

Both study species (*A. alpina*, *P. brevicornis*) exhibited at least one additional morphologically cryptic group, based primarily on mtDNA (table 8.1). As discussed in chapters six and seven, it is considered unlikely that these groups are an artefact of specimen misidentification or introgression. *Aa*1 and *Pb*1 constitute the majority of specimens from across the geographic range examined, as well as the majority of mitochondrial haplotypes. *Aa*1 and *Pb*1 are therefore considered to be archetypal of their respective species.

Species	Cryptic group	MtDNA haplogroup	NuDNA haplotype	*Wolbachia* strain	Chapter reference
A. alpina	*Aa*1	1	1	*wRi* (A)	5, 6
A. alpina	*Aa*2	2	1	*wRi* (A)	5, 6
A. alpina	*Aa*3	3	2	*wPip* (B)	5, 6
P. brevicornis	*Pb*1	1	1	N/A	7
P. brevicornis	*Pb*2	2	2	N/A	7

Table 8.1 Morphologically cryptic groups of *Amara alpina* and *Pterostichus brevicornis*, based on mitochondrial DNA haplogroup designations in this thesis. Supergroup designations of *Wolbachia* strains are indicated in parentheses and follow Klasson and colleagues (2009). Data are from chapters five, six, and seven.

Compared to *Aa*1, *Aa2* constitutes a divergent mtDNA haplogroup, but has a similar, if not identical, strain of *Wolbachia* infection. Additionally, *Aa*1 and *Aa*2 did not differ in the short fragment of nuDNA analysed (183bp; table 8.1). *Aa*2 may have arisen through a recent population bottleneck or *Wolbachia*-induced selective sweep (Hurst & Jiggins 2005), which could have removed mtDNA haplotypes that were intermediate between *Aa*1 and *Aa*2. Detailed characterisation of the *Wolbachia* strain in multiple individuals of both *Aa*1 and *Aa*2, as well as further analysis of nuDNA markers, would be required to tease apart these hypotheses.

*Aa*3 constitutes a divergent mtDNA haplogroup from the remainder of *A. alpina* (*Aa*1, *Aa*2) and differs in the short fragment of nuDNA analysed (table 8.1). In addition, *Aa*3 is infected with a *Wolbachia* strain (*wPip*) that belongs to a different *Wolbachia* supergroup than the strain (*wRi*) infecting individuals of *Aa*1 and *Aa*2. A similar situation, in which two *Wolbachia* supergroups were found to infect different mtDNA clades, has been found in the fire ant *Solenopsis invicta* (Shoemaker *et al*. 2003). These researchers suggested that the divergent mtDNA clades might represent cryptic species, due to assumed incompatibility preventing gene flow between individuals infected with different *Wolbachia* supergroups (Shoemaker *et al*. 2003). Crossbreeding experiments between individuals of *Aa*3 and those of *Aa*1 and *Aa*2, as well as further analysis of nuDNA markers, would be required to confirm the reproductive isolation, and therefore cryptic species status, of *Aa*3.

As discussed in chapter seven, *Pb*2 differed substantially from *Pb*1 at mtDNA and nuDNA markers. Based on the nuDNA data, *Pb*2 may be a cryptic species that is morphologically indistinguishable from *Pb*1, but phylogenetically closer to other

species of *P. (Cryobius)*. Additional nuDNA data, as well as assessment of other *P. (Cryobius)* species would be required to confirm these inferences.

If the prevalence of morphologically cryptic groups in the arctic beetle fauna were as common as suggested by the two taxa examined in this thesis, then large-scale phylogenetic and taxonomic investigation may be required.

8.1.6. Suggestions for Future Study

- Assessment of the source of aDNA in permafrost-preserved beetle specimens.
- Exploration of the potential of using a capture-based enrichment approach to increase the yield of endogenous aDNA from beetle specimens.
- Examination of ancient beetle specimen DNA from other species and regions would provide further insight into the findings of this study.
- Investigation into the origins of the morphologically cryptic groups within *A. alpina* and *P. brevicornis*, through further analysis of mtDNA, nuDNA, *Wolbachia* markers, and closely related taxa.

8.2. Conclusions

This thesis has provided the most comprehensive characterisation of DNA extracted from dry-stored museum and ancient permafrost-preserved beetle specimens. Previously, research on DNA from permafrost-preserved beetles had been limited to proof-of-concept. It was shown that museum specimens have massive potential for future genetic research, although further methodological development may be required before the full potential of DNA from ancient specimens can be realised. The DNA data from museum and ancient specimens were used to assess biogeographic and extinction hypotheses. Biogeographic inferences were broadly in line with existing hypotheses, and the data suggested that there had not been major extinction of populations during the last ice age. However, consideration of endosymbionts, and other insect-related concerns, would be required before further detailed analyses are conducted. This thesis demonstrates the great potential for

museum and ancient DNA from beetle specimens, although, in the words of Reiss (2006), 'proceed with caution'.

8.3. References

Allentoft ME, Collins M, Harker D, *et al*. (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society of London Series B-Biological Sciences* **279**, 4724-4733.

Andersen JC, Mills NJ (2012) DNA extraction from museum specimens of parasitic hymenoptera. *PLoS One* **7**, e45549.

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

Avila-Arcos MC, Cappellini E, Romero-Navarro JA, *et al*. (2011) Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports* **1**, 74.

Barnes I, Shapiro B, Lister A, *et al*. (2007) Genetic structure and extinction of the woolly mammoth, *Mammuthus primigenius*. *Current Biology* **17**, 1072-1075.

Briggs AW, Good JM, Green RE, *et al*. (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318-321.

Briggs AW, Stenzel U, Johnson PL, *et al*. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14616-14621.

Burbano HA, Hodges E, Green RE, *et al*. (2010) Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* **328**, 723-725.

Campos PF, Gilbert TMP (2012) DNA extraction from keratin and chitin. In: *Ancient DNA: Methods and Protocols, Methods in Molecular Biology* (eds. Shapiro B, Hofreiter M), pp. 43-49. Springer.

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

Campos PF, Willerslev E, Sher A, *et al*. (2010b) Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (*Ovibos*

moschatus) population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 5675-5680.

Drummond AJ, Pybus OG, Rambaut A, Forsberg R, Rodrigo AG (2003) Measurably evolving populations. *Trends in Ecology and Evolution* **18**, 481-488.

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

Dyke AS (2004) An outline of North American deglaciation with emphasis on central and northern Canada. In: *Quaternary Glaciations-Extent and Chronology — Part II: North America* (eds. Ehlers J, Gibbard PL), pp. 373-406. Elsevier, Amsterdam.

Elias SA, Crocker B (2008) The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota? *Quaternary Science Reviews* **27**, 2473-2483.

Frey JE, Frey B (2004) Origin of intra-individual variation in PCR-amplified mitochondrial cytochrome oxidase I of *Thrips tabaci* (Thysanoptera: Thripidae): mitochondrial heteroplasmy or nuclear integration? *Hereditas* **140**, 92-98.

Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **18**, 4541–4550.

Gullan PJ, Cranston PS (2010) *The Insects: An Outline of Entomology*, 4th edn. Wiley-Blackwell, UK.

Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B-Biological Sciences* **272**, 1525-1534.

Klasson L, Westberg J, Sapountzis P, *et al*. (2009) The mosaic genome structure of the *Wolbachia wRi* strain infecting *Drosophila simulans*. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5725-5730.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography Takes a Relaxed Random Walk in Continuous Space and Time. *Molecular Biology and Evolution* **27**, 1877-1885.

Lindroth CH (1968) The ground-beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska. Part 5. *Opuscula Entomologica, Supplementum* **33**, 649-944.

Magnacca KN, Brown MJF (2010) Mitochondrial heteroplasmy and DNA barcoding in Hawaiian *Hylaeus (Nesoprosopis)* bees (Hymenoptera: Colletidae). *BMC Evolutionary Biology* **10**, 174.

Marske KA, Leschen RAB, Buckley TR (2011) Reconciling phylogeography and ecological niche models for New Zealand beetles: Looking beyond glacial refugia. *Molecular Phylogenetics and Evolution* **59**, 89-102.

Mason VC, Li G, Helgen KM, Murphy WJ (2011) Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* **21**, 1695-1704.

Moussian B (2010) Recent advances in understanding mechanisms of insect cuticle differentiation. *Insect Biochemistry and Molecular Biology* **40**, 363-375.

Orlando L, Ginolhac A, Raghavan M, *et al*. (2011) True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Research* **21**, 1705-1719.

Rasmussen M, Guo X, Wang Y, *et al*. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94-98.

Reiss RA (2006) Ancient DNA from ice age insects: proceed with caution. *Quaternary Science Reviews* **25**, 1877-1893.

Schwarz C, Debruyne R, Kuch M, *et al*. (2009) New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. *Nucleic Acids Research* **37**, 3215-3229.

Schwert DP, Ashworth AC (1988) Late Quaternary History of the Northern Beetle Fauna of North-America - a Synthesis of Fossil and Distributional Evidence. *Memoirs of the Entomological Society of Canada*, 93-107.

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

Shoemaker DD, Keller G, Ross KG (2003) Effects of *Wolbachia* on mtDNA variation in two fire ant species. *Molecular Ecology* **12**, 1757–1771.

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* **19**, 1843-1848.

Strange JP, Knoblett J, Griswold T (2009) DNA amplification from pin-mounted bumble bees (*Bombus*) in a museum collection: effects of fragment size and specimen age on successful PCR. *Apidologie* **40**, 134–139.

Sword GA, Senior LB, Gaskin JF, Joern A (2007) Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera : Acrididae). *Systematic Entomology* **32**, 420-428.

Tsangarasa K, Greenwood AD (2012) Museums and disease: Using tissue archive and museum samples to study pathogens. *Annals of Anatomy* **194**, 58-73.

Ugelvig LV, Nielsen PS, Boomsma JJ, Nash DR (2011) Reconstructing eight decades of genetic variation in an isolated Danish population of the large blue butterfly *Maculinea arion*. *BMC Evolutionary Biology* **11**, 201.

van Houdt JK, Breman FC, Virgilio M, de Meyer M (2010) Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources* **10**, 459-465.

Vincent JFV, Wegst UGK (2004) Design and mechanical properties of insect cuticle. *Arthropod Structure and Development* **33**, 187-199.

Watts PC, Thompson DJ, Allen KA, Kemp SJ (2007) How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses? *Journal of Insect Conservation* **11**, 195–198.

Appendices

Appendix 1. Specimen Data

Data on all of the *Amara* and *Pterostichus (Cryobius)* specimens used in this thesis. CNC: Canadian National Collection of Insects, Arachnids, and Nematodes. NRM: Swedish Museum of Natural History. RHUL: Royal Holloway University of London. NRM museum references numbers are preceded by an asterisk, which is an abbreviation of 'NHRS-JLKB000020'. Mt-Ht: mitochondrial DNA haplotype (from chapters six, seven). Specimens that yielded mtDNA, but in insufficient quantities as to be used in chapters six and seven, are indicated by an asterisk. Nu-Ht: nuclear DNA haplotype (from chapters six, seven). Specimens of *A. alpina* and *P. brevicornis* that yielded nuDNA, but were not sequenced, are indicated by an asterisk. X: Extract exhausted prior to testing. LM specimens (n=42) were extracted by Mack (2008). CFx specimens (n=2) were extracted by Thomsen *et al*. (2009). E: East, W: West. Lat: latitude, Long: longitude. Lab: Laboratory in which specimens were analysed. Next Generation Sequencing (NGS) specimens were used in chapters four and five. Ancient localities are the sediment identifiers from table 2.1.

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	LM5	N/A	Modern	2003	E. Beringia	69.70	-155.20	Whole	7	*	S. Kuzmina	RHUL	NGS: Mo2
A. alpina	LM16	N/A	Modern	2002	W. Beringia	72.92	127.02	Whole	9	1	S. Kuzmina	RHUL	
A. alpina	LM16a	N/A	Modern	2002	W. Beringia	72.92	127.02	Whole	16	1	S. Kuzmina	RHUL	
A. alpina	LM22	N/A	Modern	2004	W. Beringia	64.73	177.48	Whole	1	2	S. Kuzmina	RHUL	NGS: Mo1
A. alpina	LM22a	N/A	Modern	2004	W. Beringia	64.73	177.48	Whole	1	2	S. Kuzmina	RHUL	
A. alpina	CFx1.1	N/A	Modern	2003	W. Beringia	69.70	-155.20	Whole	7	*	P. Thomsen	N/A	
A. alpina	LM67	N/A	Museum	1995	W. Beringia	64.17	171.03	Hind leg	13	1	S. Elias	RHUL	
A. alpina	LM68	N/A	Museum	1995	W. Beringia	64.17	171.03	Hind leg	13	1	S. Elias	RHUL	
A. alpina	LM69	N/A	Museum	1995	W. Beringia	64.73	177.48	Hind leg	16	1	S. Elias	RHUL	
A. alpina	LM70	N/A	Museum	1995	W. Beringia	64.73	177.48	Hind leg	16	1	S. Elias	RHUL	
A. alpina	LM71 PH-SwM69	N/A	Museum	1979	W. of Hudson Bay	58.77	-94.17	Hind leg Hind leg	49	*	S. Elias	RHUL NRM	Extracted twice
A. alpina	LM72	N/A	Museum	1979	W. of Hudson Bay	58.77	-94.17	Hind leg	34	1	S. Elias	RHUL	
A. alpina	LM73	N/A	Museum	1996	E. Beringia	66.90	-162.58	Hind leg	43	1	S. Elias	RHUL	
A. alpina	LM74	N/A	Museum	1996	E. Beringia	66.90	-162.58	Hind leg	45	1	S. Elias	RHUL	
A. alpina	PH-M1	OT 2-33	Museum	1954	E. of Hudson Bay	60.02	-70.02	Middle leg	33	1	CNC	RHUL	
A. alpina	PH-M2	OT 2-35	Museum	1958	E. of Hudson Bay	60.02	-70.02	Hind leg	33	1	CNC	RHUL	
A. alpina	PH-M3	OT 2-46	Museum	1971	E. of Hudson Bay	54.80	-66.82	Middle leg	33	*	CNC	RHUL	
A. alpina	PH-M4	OT 2-51	Museum	1970	E. of Hudson Bay	48.92	-66.20	Hind leg	12	1	CNC	RHUL	
A. alpina	PH-M5	OT 2-54	Museum	1979	"Hudson Bay"	54.73	-82.23	Hind leg			CNC	RHUL	
A. alpina	PH-M6	OT 3-15	Museum	1948	W. of Hudson Bay	68.70	-134.12	Hind leg	48	*	CNC	RHUL	
A. alpina	PH-M7	OT 3-32	Museum	1982	E. Beringia	65.07	-138.28	Hind leg	22	1	CNC	RHUL	
A. alpina	PH-M8	OT 3-67	Museum	1977	Lower 48	36.55	-105.42	Hind leg	29	1	CNC	RHUL	
A. alpina	PH-M9	OT 3-78	Museum	1986	Lower 48	45.90	-68.92	Hind leg	47	1	CNC	RHUL	
A. alpina	PH-M10	OT 3-84	Museum	1961	E. Beringia	63.87	-160.78	Hind leg	43	*	CNC	RHUL	
A. alpina	PH-M13	OT 2-24	Museum	1935	E. of Hudson Bay	60.42	-64.82	Hind leg	33	*	CNC	RHUL	
A. alpina	PH-M14	OT 2-39	Museum	1954	E. of Hudson Bay	56.25	-64.70	Hind leg	33	*	CNC	RHUL	
A. alpina	PH-M15	OT 2-62	Museum	1962	W. of Hudson Bay	58.77	-94.17	Hind leg	34	*	CNC	RHUL	
A. alpina	PH-M16	OT 2-70	Museum	1981	W. of Hudson Bay	57.05	-122.87	Hind leg	36	*	CNC	RHUL	
A. alpina	PH-M17	OT 2-83	Museum	1950	W. of Hudson Bay	61.10	-94.07	Hind leg	34	1	CNC	RHUL	
A. alpina	PH-M18	OT 2-26	Museum	1949	E. of Hudson Bay	58.45	-78.10	Hind leg	34	*	CNC	RHUL	
A. alpina	PH-M19	OT 2-55	Museum	1979	"Hudson Bay"	54.73	-82.23	Hind leg			CNC	RHUL	
A. alpina	PH-M20	OT 2-60	Museum	1999	W. of Hudson Bay	58.75	-93.38	Hind leg	34	1	CNC	RHUL	
A. alpina	PH-M21	OT 2-77	Museum	1972	W. of Hudson Bay	68.37	-133.73	Hind leg	43	*	CNC	RHUL	
A. alpina	PH-M22	OT 2-85	Museum	1947	W. of Hudson Bay	64.32	-96.02	Hind leg	34	1	CNC	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH-M23	OT 2-90	Museum	1951	W. of Hudson Bay	67.82	-115.10	Hind leg	19	1	CNC	RHUL	
A. alpina	PH-M24	OT 3-6	Museum	1950	W. of Hudson Bay	61.92	-96.67	Hind leg	34	1	CNC	RHUL	
A. alpina	PH-M25	OT 3-21	Museum	1964	E. of Hudson Bay	70.50	-74.83	Hind leg	40	1	CNC	RHUL	
A. alpina	PH-M26	OT 3-28	Museum	1984	E. Beringia	69.28	-140.05	Hind leg	27	1	CNC	RHUL	
A. alpina	PH-M27	OT 3-36	Museum	1971	E. Beringia	69.58	-139.08	Hind leg	13	1	CNC	RHUL	
A. alpina	PH-M28	OT 2-96	Museum	1952	W. of Hudson Bay	70.73	-117.77	Hind leg	7	*	CNC	RHUL	
A. alpina	PH-M29	OT 3-18	Museum	1948	W. of Hudson Bay	68.70	-134.12	Hind leg	43	1	CNC	RHUL	
A. alpina	PH-M30	OT 3-25	Museum	1982	E. Beringia	65.07	-138.28	Hind leg	23	*	CNC	RHUL	
A. alpina	PH-M31	OT 3-33	Museum	1982	E. Beringia	65.07	-138.28	Hind leg	26	*	CNC	RHUL	
A. alpina	PH-M32	OT 3-49	Museum	1961	E. Beringia	68.13	-165.97	Hind leg	43	*	CNC	RHUL	
A. alpina	PH-M37	OT 2-13	Museum	1949	E. of Hudson Bay	51.60	-55.87	Hind leg	38	*	CNC	RHUL	
A. alpina	PH-M38	OT 2-16	Museum	1954	E. of Hudson Bay	57.47	-61.87	Hind leg	41	*	CNC	RHUL	
A. alpina	PH-M39	OT 2-42	Museum	1966	Lower 48	44.27	-71.30	Hind leg	40	1	CNC	RHUL	
A. alpina	PH-M40	OT 2-57	Museum	1979	"Hudson Bay"	54.73	-82.23	Hind leg			CNC	RHUL	
A. alpina	PH-M41	OT 2-64	Museum	1937	W. of Hudson Bay	58.77	-94.17	Hind leg	36	*	CNC	RHUL	
A. alpina	PH-M42	OT 2-66	Museum	1959	W. of Hudson Bay	58.65	-124.67	Hind leg	36	1	CNC	RHUL	
A. alpina	PH-M43	OT 2-75	Museum	1935	E. of Hudson Bay	62.85	-69.87	Hind leg	40	*	CNC	RHUL	
A. alpina	PH-M44	OT 2-86	Museum	1956	W. of Hudson Bay	68.22	-135.00	Hind leg	44	*	CNC	RHUL	
A. alpina	PH-M45	OT 2-99	Museum	1951	W. of Hudson Bay	69.53	-93.53	Hind leg	7	*	CNC	RHUL	
A. alpina	PH-M46	OT 3-40	Museum	1987	E. Beringia	67.97	-136.48	Hind leg	43	*	CNC	RHUL	
A. alpina	PH-M47	OT 3-45	Museum	1947	E. Beringia	57.17	-170.27	Hind leg	17	*	CNC	RHUL	
A. alpina	PH-M48 PH-M71	OT 3-52 OT 3-51	Museum	1978	E. Beringia	68.27	-149.43	Middle leg Hind leg	16	*	CNC	RHUL	Extracted twice
A. alpina	PH-M49	OT 3-56	Museum	1961	Lower 48	39.58	-105.65	Hind leg	6	1	CNC	RHUL	
A. alpina	PH-M50	OT 3-61	Museum	1961	Lower 48	39.58	-105.65	Hind leg	6	*	CNC	RHUL	
A. alpina	PH-M51	OT 3-66	Museum	1976	Lower 48	38.47	-109.23	Hind leg	31	1	CNC	RHUL	
A. alpina	PH-M52	OT 3-70	Museum	1977	Lower 48	38.84	-105.04	Hind leg	50	*	CNC	RHUL	
A. alpina	PH-M53	OT 3-73	Museum	1973	Lower 48	39.37	-106.18	Hind leg	51	*	CNC	RHUL	
A. alpina	PH-M54	OT 3-76	Museum	1993	Lower 48	39.58	-105.65	Hind leg	6	*	CNC	RHUL	
A. alpina	PH-M55	OT 3-80	Museum	1986	Lower 48	45.90	-68.92	Hind leg	47	*	CNC	RHUL	
A. alpina	PH-M56	OT 3-82	Museum	1947	E. Beringia	57.17	-170.27	Hind leg	17	*	CNC	RHUL	
A. alpina	PH-M61	OT 2-15	Museum	1925	E. of Hudson Bay	55.47	-60.22	Hind leg	40	*	CNC	RHUL	
A. alpina	PH-M62	OT 2-29	Museum	1949	E. of Hudson Bay	55.28	-77.76	Hind leg	33	*	CNC	RHUL	
A. alpina	PH-M63	OT 2-44	Museum	1968	E. of Hudson Bay	58.45	-78.10	Hind leg	37	1	CNC	RHUL	
A. alpina	PH-M64	OT 2-61	Museum	1952	W. of Hudson Bay	58.77	-94.17	Hind leg	34	*	CNC	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH-M65	OT 2-65	Museum	1980	W. of Hudson Bay	50.95	-115.15	Hind leg	8	*	CNC	RHUL	
A. alpina	PH-M66	OT 2-71	Museum	1935	E. of Hudson Bay	62.85	-69.87	Hind leg	33	*	CNC	RHUL	
A. alpina	PH-M67	OT 2-78	Museum	1950	W. of Hudson Bay	66.53	-86.23	Hind leg	35	*	CNC	RHUL	NGS: Mu1
A. alpina	PH-M68	OT 2-98	Museum	1972	W. of Hudson Bay	69.00	-134.67	Hind leg	48	*	CNC	RHUL	
A. alpina	PH-M69	OT 3-23	Museum	1951	E. Beringia	67.42	-140.98	Hind leg	46	*	CNC	RHUL	
A. alpina	PH-M70	OT 3-43	Museum	1987	E. Beringia	68.02	-136.70	Hind leg	17	*	CNC	RHUL	
A. alpina	PH-M72	OT 3-54	Museum	1961	E. Beringia	68.13	-165.97	Hind leg	16	*	CNC	RHUL	
A. alpina	PH-M73	OT 3-58	Museum	1954	Lower 48	44.25	-71.32	Hind leg	12	*	CNC	RHUL	
A. alpina	PH-M74	OT 3-64	Museum	1986	Lower 48	45.90	-68.92	Hind leg	47	*	CNC	RHUL	
A. alpina	PH-M75	OT 3-69	Museum	1977	Lower 48	36.55	-105.42	Hind leg	28	*	CNC	RHUL	
A. alpina	PH-M76	OT 3-72	Museum	1977	Lower 48	37.28	-105.17	Hind leg	30	*	CNC	RHUL	
A. alpina	PH-M77	OT 3-74	Museum	1965	Lower 48	40.37	-105.57	Hind leg	38	*	CNC	RHUL	
A. alpina	PH-M78	OT 3-77	Museum	1979	Lower 48	45.07	-109.57	Hind leg	38	*	CNC	RHUL	
A. alpina	PH-M79	OT 3-81	Museum	1968	Lower 48	44.97	-109.47	Hind leg	38	*	CNC	RHUL	
A. alpina	PH-M80	OT 3-86	Museum	1959	E. Beringia	69.37	-152.13	Hind leg	7	*	CNC	RHUL	
A. alpina	PH-SwM1	*187	Museum	1988	Scandinavia	62.30	9.60	Hind leg	25	*	NRM	NRM	
A. alpina	PH-SwM2	*188	Museum	?	Scandinavia	67.13	16.35	Hind leg	42	6	NRM	NRM	
A. alpina	PH-SwM3	*189	Museum	1925	Scandinavia	62.40	11.00	Hind leg	16	*	NRM	NRM	
A. alpina	PH-SwM4	*190	Museum	1915	Scandinavia	62.28	13.83	Hind leg	16	*	NRM	NRM	
A. alpina	PH-SwM5	*191	Museum	1988	Scandinavia	62.57	12.23	Hind leg	25	*	NRM	NRM	
A. alpina	PH-SwM6	*192	Museum	1913	Scandinavia	62.45	13.85	Hind leg	16	*	NRM	NRM	
A. alpina	PH-SwM7	*193	Museum	1949	Scandinavia	62.60	12.17	Hind leg	16	*	NRM	NRM	
A. alpina	PH-SwM8	*194	Museum	1913	Scandinavia	62.45	13.85	Hind leg	16	*	NRM	NRM	
A. alpina	PH-SwM9	*195	Museum	1907	Scandinavia	63.32	14.52	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM10	*196	Museum	?	Scandinavia	62.90	15.12	Hind leg	24	*	NRM	NRM	
A. alpina	PH-SwM11	*197	Museum	1936	Scandinavia	63.32	13.25	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM12	*198	Museum	1970	Scandinavia	67.47	18.50	Middle leg	42	*	NRM	NRM	
A. alpina	PH-SwM13	*199	Museum	1900	Scandinavia	66.22	18.30	Hind leg	42	1	NRM	NRM	
A. alpina	PH-SwM14	*200	Museum	1916	Scandinavia	66.22	18.30	Hind leg	15	*	NRM	NRM	
A. alpina	PH-SwM15	*201	Museum	1916	Scandinavia	66.22	18.30	Hind leg	14	*	NRM	NRM	
A. alpina	PH-SwM16	*202	Museum	1900	Scandinavia	68.22	19.70	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM17	*203	Museum	1918	Scandinavia	68.35	18.80	Hind leg	2	1	NRM	NRM	
A. alpina	PH-SwM18	*204	Museum	1919	Scandinavia	68.42	18.45	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM19	*205	Museum	1932	Scandinavia	68.35	18.80	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM20	*206	Museum	1935	Scandinavia	68.43	22.47	Hind leg	14	*	NRM	NRM	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH-SwM25	*207	Museum	1941	Scandinavia	67.87	20.50	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM26	*208	Museum	1947	Scandinavia	68.12	19.90	Hind leg	18	*	NRM	NRM	
A. alpina	PH-SwM27	*209	Museum	1949	Scandinavia	68.40	18.32	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM28	*210	Museum	1973	Scandinavia	68.35	18.80	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM29	*211	Museum	1973	Scandinavia	68.35	18.80	Hind leg	42	*	NRM	NRM	
A. alpina	PH-SwM30	*212	Museum	?	Scandinavia	71.02	25.00	Hind leg	53	1	NRM	NRM	
A. alpina	PH-SwM31	*213	Museum	?	"Hudson Bay"	N/A	N/A	Hind leg	32	1	NRM	NRM	
A. alpina	PH-SwM32	*214	Museum	?	"Hudson Bay"	N/A	N/A	Hind leg	36	*	NRM	NRM	
A. alpina	PH-SwM33	*215	Museum	1951	E. of Hudson Bay	51.52	-56.83	Hind leg	39	1	NRM	NRM	
A. alpina	PH-SwM34	*216	Museum	1875	North Russia	71.92	82.65	Hind leg	2	*	NRM	NRM	
A. alpina	PH-SwM35	*217	Museum	1875	North Russia	71.82	51.48	Hind leg	5	*	NRM	NRM	
A. alpina	PH-SwM36	*218	Museum	1875	North Russia	71.82	51.48	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM37	*219	Museum	1875	North Russia	71.28	53.33	Middle leg	2	*	NRM	NRM	
A. alpina	PH-SwM38	*220	Museum	1875	North Russia	73.30	54.40	Hind leg	2	*	NRM	NRM	
A. alpina	PH-SwM39	*221	Museum	1875	North Russia	69.67	60.35	Hind leg	2	*	NRM	NRM	
A. alpina	PH-SwM40	*222	Museum	1875	North Russia	71.45	83.42	Hind leg	2	*	NRM	NRM	
A. alpina	PH-SwM41	*223	Museum	1875	North Russia	72.43	52.80	Hind leg	3	*	NRM	NRM	
A. alpina	PH-SwM42	*224	Museum	1875	North Russia	N/A	N/A	Middle leg	14	*	NRM	NRM	
A. alpina	PH-SwM43	*225	Museum	?	Scandinavia	67.48	15.50	Hind leg		*	NRM	NRM	
A. alpina	PH-SwM44	*226	Museum	?	Scandinavia	67.48	15.50	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM49	*227	Museum	1875	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM50	*228	Museum	1875	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM51	*229	Museum	1875	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM52	*230	Museum	1875	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM53	*231	Museum	?	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM54	*232	Museum	1875	North Russia	70.02	59.43	Hind leg	4	1	NRM	NRM	
A. alpina	PH-SwM55	*233	Museum	1875	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM56	*234	Museum	1875	North Russia	70.02	59.43	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM57	*235	Museum	1875	North Russia	72.87	53.38	Hind leg	2	*	NRM	NRM	NGS: Mu2
A. alpina	PH-SwM58	*236	Museum	1875	North Russia	72.87	53.38	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM59	*237	Museum	1875	North Russia	72.87	53.38	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM60	*238	Museum	1875	North Russia	72.43	52.80	Hind leg	52	*	NRM	NRM	
A. alpina	PH-SwM61	*239	Museum	1875	North Russia	72.87	53.38	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM62	*240	Museum	1875	North Russia	72.87	53.38	Hind leg	*	*	NRM	NRM	
A. alpina	PH-SwM63	*241	Museum	1875	North Russia	72.87	53.38	Hind leg		*	NRM	NRM	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH-SwM64	*242	Museum	1875	North Russia	72.87	53.38	Hind leg	*	*	NRM	NRM	
A. alpina	LM27	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	?	43	1	S. Elias	RHUL	
A. alpina	LM28	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	?	*	X	S. Elias	RHUL	
A. alpina	LM29	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	?			S. Elias	RHUL	
A. alpina	LM43	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron			S. Elias	RHUL	
A. alpina	LM44	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron		*	S. Elias	RHUL	
A. alpina	LM45	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 elytron			S. Elias	RHUL	
A. alpina	LM46	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron	*	*	S. Elias	RHUL	
A. alpina	LM47	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	43	7	S. Elias	RHUL	
A. alpina	LM59	N/A	Ancient	2005	AL-5-05-B28	64.47	-142.72	<0.5 elytron			S. Elias	RHUL	
A. alpina	LM82	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron	*	*	S. Elias	RHUL	
A. alpina	LM83	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum	*	*	S. Elias	RHUL	
A. alpina	LM84	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	43	1	S. Elias	RHUL	
A. alpina	LM85	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 elytron			S. Elias	RHUL	
A. alpina	LM86	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 elytron			S. Elias	RHUL	
A. alpina	LM87	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron	*	*	S. Elias	RHUL	
A. alpina	LM88	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	43	1	S. Elias	RHUL	
A. alpina	LM89	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 elytron			S. Elias	RHUL	
A. alpina	LM90	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron		*	S. Elias	RHUL	
A. alpina	LM91	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron	43	1	S. Elias	RHUL	
A. alpina	LM92	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 elytron		*	S. Elias	RHUL	
A. alpina	LM93	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron			S. Elias	RHUL	
A. alpina	LM94	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum		1	S. Elias	RHUL	
A. alpina	CFx3.8	N/A	Ancient	2004	PT-CFx	65.10	171.18	?	7	1	P. Thomsen	N/A	
A. alpina	PH1 PH2	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	0.5 elytron 0.5 elytron	*	*	S. Elias	RHUL	Extracted twice
A. alpina	PH3 PH4	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum 0.5 pronotum			S. Elias	RHUL	Extracted twice
A. alpina	PH5 PH6	N/A	Ancient	2003	T03-B08b	69.70	-155.20	0.5 pronotum 0.5 pronotum	*		S. Elias	RHUL	Extracted twice
A. alpina	PH7 PH8	N/A	Ancient	2003	T03-B17	69.70	-155.20	0.5 pronotum 0.5 pronotum	*		S. Elias	RHUL	Extracted twice
A. alpina	PH9 PH10	N/A	Ancient	2003	T03-B17	69.70	-155.20	0.5 elytron 0.5 elytron	16	1	S. Elias	RHUL	Extracted twice

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH13 PH14	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	0.5 pronotum 0.5 pronotum			S. Elias	RHUL	Extracted twice
A. alpina	PH15 PH16	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum 0.5 pronotum			S. Elias	RHUL	Extracted twice
A. alpina	PH17 PH18	N/A	Ancient	2003	T03-B08b	69.70	-155.20	0.5 pronotum 0.5 pronotum	11	1	S. Elias	RHUL	Extracted twice
A. alpina	PH19 PH20	N/A	Ancient	2003	T03-B17	69.70	-155.20	0.5 pronotum 0.5 pronotum			S. Elias	RHUL	Extracted twice
A. alpina	PH21 PH22	N/A	Ancient	2003	T03-B17	69.70	-155.20	0.5 elytron 0.5 elytron	*	*	S. Elias	RHUL	Extracted twice
A. alpina	PH25	N/A	Ancient	2005	AL-3-05-B13	66.02	-149.08	Elytron			S. Elias	RHUL	
A. alpina	PH26	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum			S. Elias	RHUL	
A. alpina	PH27	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Pronotum	*	*	S. Elias	RHUL	
A. alpina	PH28	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Pronotum	*		S. Elias	RHUL	
A. alpina	PH29	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron			S. Elias	RHUL	
A. alpina	PH30	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron			S. Elias	RHUL	
A. alpina	PH31	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum	*		S. Elias	RHUL	
A. alpina	PH32	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	RHUL	
A. alpina	PH33	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	RHUL	
A. alpina	PH34	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	RHUL	
A. alpina	PH35	N/A	Ancient	2006	DF-06-90	67.80	-130.85	Pronotum			S. Elias	RHUL	
A. alpina	PH36	N/A	Ancient	2006	DF-06-92	67.80	-130.85	Pronotum			S. Elias	RHUL	
A. alpina	PH37	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	RHUL	
A. alpina	PH38	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum	21	1	S. Elias	RHUL	
A. alpina	PH39	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	*		S. Elias	RHUL	
A. alpina	PH40	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	10	3	S. Elias	RHUL	
A. alpina	PH41	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum			S. Elias	RHUL	
A. alpina	PH42	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum	*		S. Elias	RHUL	
A. alpina	PH43	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum	*		S. Elias	RHUL	
A. alpina	PH44	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum			S. Elias	RHUL	
A. alpina	PH49	N/A	Ancient	2005	AL-3-05-B13	66.02	-149.08	Pronotum	*		S. Elias	RHUL	
A. alpina	PH50	N/A	Ancient	2005	AL-4-05-B22	65.10	-153.25	Pronotum			S. Elias	RHUL	
A. alpina	PH51	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum			S. Elias	RHUL	
A. alpina	PH52	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Pronotum			S. Elias	RHUL	
A. alpina	PH53	N/A	Ancient	2005	GZ-05-48	63.78	-139.22	Pronotum			S. Elias	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH54	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	RHUL	
A. alpina	PH55	N/A	Ancient	2006	DF-06-90	67.80	-130.85	Pronotum			S. Elias	RHUL	
A. alpina	PH56	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	10	1	S. Elias	RHUL	NGS: An2
A. alpina	PH57	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	16	1	S. Elias	RHUL	
A. alpina	PH58	N/A	Ancient	2003	T03-B17	69.70	-155.20	Elytron	10	1	S. Elias	RHUL	
A. alpina	PH61	N/A	Ancient	2005	AL-3-05-B13	66.02	-149.08	Elytron			S. Elias	RHUL	
A. alpina	PH62	N/A	Ancient	2005	AL-4-05-B22	65.10	-153.25	Pronotum			S. Elias	RHUL	
A. alpina	PH63	N/A	Ancient	2005	AL-4-05-B22	65.10	-153.25	Head capsule			S. Elias	RHUL	
A. alpina	PH64	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron			S. Elias	RHUL	
A. alpina	PH65	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum		1	S. Elias	RHUL	
A. alpina	PH66	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH67	N/A	Ancient	2006	DF-06-92	67.80	-130.85	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH68	N/A	Ancient	2006	DF-06-101	67.48	-140.10	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH69	N/A	Ancient	2004	DF-04-67	67.43	-140.37	Pronotum			S. Elias	RHUL	
A. alpina	PH70	N/A	Ancient	2004	DF-04-67	67.43	-140.37	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH71	N/A	Ancient	2004	DF-04-67	67.43	-140.37	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH72	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum			S. Elias	RHUL	
A. alpina	PH73	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	10	1	S. Elias	RHUL	
A. alpina	PH74	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	10	1	S. Elias	RHUL	
A. alpina	PH75	N/A	Ancient	2003	T03-B08b	69.70	-155.20	Pronotum	*	*	S. Elias	RHUL	
A. alpina	PH76	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum		1	S. Elias	RHUL	
A. alpina	PH77	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum		*	S. Elias	RHUL	
A. alpina	PH78	N/A	Ancient	2003	T03-B17	69.70	-155.20	Elytron	*	1	S. Elias	RHUL	
A. alpina	PH79	N/A	Ancient	2003	T03-B17	69.70	-155.20	Elytron			S. Elias	RHUL	
A. alpina	PH80	N/A	Ancient	2003	T03-B17	69.70	-155.20	Elytron			S. Elias	RHUL	
A. alpina	PH85	N/A	Ancient	2006	DF-06-92	67.80	-130.85	0.5 pronotum	*		S. Elias	RHUL	
A. alpina	PH86	N/A	Ancient	2006	DF-06-101	67.48	-140.10	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH87	N/A	Ancient	2003	T03-B17	69.70	-155.20	Elytron	17	1	S. Elias	RHUL	
A. alpina	PH88	N/A	Ancient	2003	T03-B17	69.70	-155.20	Elytron	*	*	S. Elias	RHUL	
A. alpina	PH89	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum			S. Elias	RHUL	
A. alpina	PH90	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum	*		S. Elias	RHUL	
A. alpina	PH91	N/A	Ancient	2003	T03-B17	69.70	-155.20	Pronotum		1	S. Elias	RHUL	
A. alpina	PH92	N/A	Ancient	2003	T03-B17	69.70	-155.20	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH93	N/A	Ancient	2003	T03-B17	69.70	-155.20	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH96	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum			S. Elias	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH97	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 elytron	*		S. Elias	RHUL	
A. alpina	PH98	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH99	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 elytron			S. Elias	RHUL	
A. alpina	PH100	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 pronotum			S. Elias	RHUL	
A. alpina	PH101	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 pronotum			S. Elias	RHUL	
A. alpina	PH102	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 elytron	43	1	S. Elias	RHUL	
A. alpina	PH103	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 elytron			S. Elias	RHUL	
A. alpina	PH-SwA1	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA2	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA3	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron		*	S. Elias	NRM	
A. alpina	PH-SwA4	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 elytron	*	*	S. Elias	NRM	
A. alpina	PH-SwA5	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron			S. Elias	NRM	
A. alpina	PH-SwA6	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron		5	S. Elias	NRM	
A. alpina	PH-SwA7	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	*		S. Elias	NRM	
A. alpina	PH-SwA8	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	43	*	S. Elias	NRM	
A. alpina	PH-SwA9	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum	7	*	S. Elias	NRM	
A. alpina	PH-SwA10	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA11	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA12	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum	*		S. Elias	NRM	
A. alpina	PH-SwA13	N/A	Ancient	2003	GZ-03-128	63.75	-139.30	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA14	N/A	Ancient	2006	DF-06-90	67.80	-130.85	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA15	N/A	Ancient	2006	DF-06-90	67.80	-130.85	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA16	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA17	N/A	Ancient	2006	DF-06-101	67.48	-140.10	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA18	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA19	N/A	Ancient	2006	DF-06-101	67.48	-140.10	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA20	N/A	Ancient	2005	AL-5-05-B28	64.47	-142.72	?pronotum			S. Elias	NRM	
A. alpina	PH-SwA25	N/A	Ancient	2006	DF-06-90	67.80	-130.85	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA26	N/A	Ancient	2005	AL-5-05-B28	64.47	-142.72	?pronotum			S. Elias	NRM	
A. alpina	PH-SwA27	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA28	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron	43	*	S. Elias	NRM	
A. alpina	PH-SwA29	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA30	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron	*	*	S. Elias	NRM	
A. alpina	PH-SwA31	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron			S. Elias	NRM	
A. alpina	PH-SwA32	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron			S. Elias	NRM	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
A. alpina	PH-SwA33	N/A	Ancient	2004	GZ-04-44b	63.97	-138.97	Elytron	43	*	S. Elias	NRM	
A. alpina	PH-SwA34	N/A	Ancient	2003	GZ-03-128	63.75	-139.30	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA35	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	NRM	
A. alpina	PH-SwA36	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum		*	S. Elias	NRM	
A. alpina	PH-SwA37	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	43	*	S. Elias	NRM	NGS: An1
A. alpina	PH-SwA38	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA39	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	*	*	S. Elias	NRM	
A. alpina	PH-SwA40	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	>0.5 elytron		4	S. Elias	NRM	
A. alpina	PH-SwA41	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron			S. Elias	NRM	
A. alpina	PH-SwA42	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum		1	S. Elias	NRM	
A. alpina	PH-SwA43	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum	*	*	S. Elias	NRM	
A. alpina	PH-SwA44	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum	20	1	S. Elias	NRM	
A. alpina	PH-SwA49	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum	16	*	S. Elias	NRM	
A. alpina	PH-SwA50	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA51	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron			S. Elias	NRM	
A. alpina	PH-SwA52	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Elytron	*		S. Elias	NRM	
A. alpina	PH-SwA53	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	<0.5 elytron	*		S. Elias	NRM	
A. alpina	PH-SwA54	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum	*		S. Elias	NRM	
A. alpina	PH-SwA55	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	0.5 pronotum			S. Elias	NRM	
A. alpina	PH-SwA56	N/A	Ancient	2004	GZ-04-44a	63.97	-138.97	Pronotum			S. Elias	NRM	
A. glacialis	LM15	N/A	Modern	2002	W. Beringia	72.92	127.02	Whole	*	*	S. Kuzmina	RHUL	
P. brevicornis	LM7	N/A	Modern	2003	E. Beringia	69.70	-155.20	Whole	2	*	S. Kuzmina	RHUL	
P. brevicornis	LM12	N/A	Modern	2002	W. Beringia	72.92	127.02	Whole	3	*	S. Kuzmina	RHUL	
P. brevicornis	LM20	N/A	Modern	2004	W. Beringia	64.73	177.48	Whole	4	*	S. Kuzmina	RHUL	
P. brevicornis	PH-M85	OT 1-39	Museum	1949	E. of Hudson Bay	51.57	-55.72	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M86	OT 1-41	Museum	1923	E. of Hudson Bay	55.47	-60.22	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M87	OT 1-46	Museum	1971	E. of Hudson Bay	50.30	-64.50	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M88	OT 1-49	Museum	1979	"Hudson Bay"	54.73	-82.23	Hind leg			CNC	RHUL	
P. brevicornis	PH-M89	OT 1-50	Museum	2001	W. of Hudson Bay	55.98	-87.63	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M90	OT 1-53	Museum	1971	W. of Hudson Bay	47.98	-84.77	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M91	OT 1-55	Museum	1950	W. of Hudson Bay	56.35	-94.70	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M92	OT 1-57	Museum	1937	W. of Hudson Bay	58.77	-94.17	Hind leg	18	*	CNC	RHUL	
P. brevicornis	PH-M93	OT 1-61	Museum	1978	W. of Hudson Bay	58.77	-125.70	Hind leg	7	*	CNC	RHUL	
P. brevicornis	PH-M94	OT 1-63	Museum	1950	W. of Hudson Bay	63.33	-90.70	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M95	OT 1-66	Museum	1948	W. of Hudson Bay	68.70	-134.12	Hind leg	6	*	CNC	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
P. brevicornis	PH-M96	OT 1-69	Museum	1980	W. of Hudson Bay	68.22	-135.90	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M97	OT 1-71	Museum	1971	W. of Hudson Bay	69.45	-132.20	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M98	OT 1-77	Museum	1951	E. Beringia	67.42	-140.98	Hind leg	6	*	CNC	RHUL	
P. brevicornis	PH-M99	OT 1-80	Museum	1987	E. Beringia	67.97	-136.48	Hind leg	1	*	CNC	RHUL	
P. brevicornis	PH-M100	OT 1-86	Museum	1951	E. Beringia	60.12	-149.43	Hind leg	8	1	CNC	RHUL	
P. brevicornis	PH-M101	OT 1-92	Museum	1959	E. Beringia	69.37	-152.13	Hind leg	3	1	CNC	RHUL	
P. brevicornis	PH-M102	OT 2-2	Museum	1980	Bering Islands	56.57	-169.62	Hind leg		*	CNC	RHUL	
P. brevicornis	PH-M103	OT 2-4	Museum	1973	Appalachians	44.27	-71.30	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M104	OT 2-10	Museum	1987	Appalachians	44.30	-71.32	Front leg	5	1	CNC	RHUL	
P. brevicornis	PH-M109	OT 1-43	Museum	1951	E. of Hudson Bay	52.27	-55.58	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M110	OT 1-44	Museum	1972	E. of Hudson Bay	48.92	-66.20	Hind leg	5	1	CNC	RHUL	
P. brevicornis	PH-M111	OT 1-52	Museum	1973	W. of Hudson Bay	52.95	-90.70	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M112	OT 1-59	Museum	1958	W. of Hudson Bay	59.63	-136.47	Hind leg	7	1	CNC	RHUL	
P. brevicornis	PH-M113	OT 1-64	Museum	1965	W. of Hudson Bay	68.70	-134.12	Hind leg	6	*	CNC	RHUL	
P. brevicornis	PH-M114	OT 1-68	Museum	1975	W. of Hudson Bay	70.75	-109.28	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M115	OT 1-73	Museum	1960	E. Beringia	62.19	-132.90	Hind leg	7	*	CNC	RHUL	
P. brevicornis	PH-M116	OT 1-76	Museum	1960	E. Beringia	62.48	-130.42	Hind leg	6	1	CNC	RHUL	
P. brevicornis	PH-M117	OT 1-82	Museum	1960	E. Beringia	62.68	-132.33	Hind leg	7	1	CNC	RHUL	
P. brevicornis	PH-M118	OT 1-85	Museum	1951	E. Beringia	60.33	-151.27	Hind leg	7	1	CNC	RHUL	
P. brevicornis	PH-M119	OT 1-87	Museum	1978	E. Beringia	63.12	-147.50	Hind leg	9	1	CNC	RHUL	
P. brevicornis	PH-M120	OT 1-90	Museum	1948	Bering Islands	57.17	-170.27	Hind leg	9	1	CNC	RHUL	
P. brevicornis	PH-M121	OT 1-94	Museum	1961	E. Beringia	63.87	-160.78	Hind leg	15	*	CNC	RHUL	
P. brevicornis	PH-M122	OT 1-97	Museum	1984	E. Beringia	65.48	-144.63	Hind leg	6	*	CNC	RHUL	
P. brevicornis	PH-M123	OT 1-98	Museum	1978	E. Beringia	63.10	-149.50	Hind leg	7	*	CNC	RHUL	
P. brevicornis	PH-M124	OT 2-7	Museum	1987	Appalachians	44.27	-71.30	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M125	OT 2-8	Museum	1991	W. of Hudson Bay	47.97	-89.58	Hind leg	3	1	CNC	RHUL	
P. brevicornis	PH-M126	OT 3-88	Museum	1972	W. of Hudson Bay	61.67	-120.72	Hind leg	7	*	CNC	RHUL	
P. brevicornis	PH-M127	OT 3-95	Museum	1951	W. of Hudson Bay	69.53	-93.53	Hind leg	14	*	CNC	RHUL	
P. brevicornis	PH-M128	OT 4-4	Museum	1984	E. Beringia	69.63	-140.92	Hind leg	12	1	CNC	RHUL	
P. brevicornis	PH-M133	OT 1-42	Museum	1951	E. of Hudson Bay	52.02	-55.87	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M134	OT 1-47	Museum	1971	E. of Hudson Bay	50.28	-64.02	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M135	OT 1-54	Museum	1977	W. of Hudson Bay	56.78	-98.93	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M136	OT 1-60	Museum	1959	W. of Hudson Bay	58.65	-124.67	Hind leg	7	*	CNC	RHUL	
P. brevicornis	PH-M137	OT 1-62	Museum	1948	W. of Hudson Bay	69.50	-133.78	Hind leg	13	1	CNC	RHUL	
P. brevicornis	PH-M138	OT 1-70	Museum	1972	W. of Hudson Bay	61.57	-124.03	Hind leg	9	*	CNC	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
P. brevicornis	PH-M139	OT 1-74	Museum	1960	E. Beringia	62.32	-133.20	Hind leg	9	*	CNC	RHUL	
P. brevicornis	PH-M140	OT 1-78	Museum	1968	E. Beringia	63.98	-135.48	Hind leg	9	*	CNC	RHUL	
P. brevicornis	PH-M141	OT 1-84	Museum	1950	E. Beringia	60.40	-150.33	Hind leg	9	*	CNC	RHUL	
P. brevicornis	PH-M142	OT 1-89	Museum	1978	E. Beringia	62.73	-150.10	Hind leg	10	1	CNC	RHUL	
P. brevicornis	PH-M143	OT 1-91	Museum	1948	Bering Islands	57.17	-170.27	Hind leg	9	1	CNC	RHUL	
P. brevicornis	PH-M144	OT 1-96	Museum	1978	E. Beringia	68.03	-149.68	Hind leg	16	1	CNC	RHUL	
P. brevicornis	PH-M145	OT 1-100	Museum	1978	E. Beringia	68.47	-149.43	Hind leg	1	2	CNC	RHUL	
P. brevicornis	PH-M146	OT 2-5	Museum	1976	Appalachians	44.27	-71.30	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M147	OT 2-12	Museum	1987	Appalachians	44.30	-71.32	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M148	OT 3-91	Museum	1975	W. of Hudson Bay	70.75	-109.28	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M149	OT 3-93	Museum	1969	W. of Hudson Bay	67.48	-124.95	Hind leg	17	*	CNC	RHUL	
P. brevicornis	PH-M150	OT 3-97	Museum	1958	W. of Hudson Bay	67.82	-115.10	Hind leg	3	1	CNC	RHUL	
P. brevicornis	PH-M151	OT 3-100	Museum	1984	E. Beringia	69.22	-140.08	Hind leg	6	*	CNC	RHUL	
P. brevicornis	PH-M152	OT 4-2	Museum	1984	E. Beringia	69.45	-140.38	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M157	OT 1-40	Museum	1965	E. of Hudson Bay	56.15	-62.67	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M158	OT 1-51	Museum	1973	W. of Hudson Bay	52.75	-90.18	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M159	OT 1-56	Museum	1999	W. of Hudson Bay	58.37	-93.47	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M160	OT 1-58	Museum	1950	W. of Hudson Bay	56.35	-94.70	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M161	OT 1-67	Museum	1972	W. of Hudson Bay	61.67	-120.72	Hind leg		*	CNC	RHUL	
P. brevicornis	PH-M162	OT 1-72	Museum	1978	E. Beringia	64.60	-138.32	Hind leg	6	1	CNC	RHUL	
P. brevicornis	PH-M163	OT 1-75	Museum	1968	E. Beringia	64.52	-138.23	Hind leg	6	*	CNC	RHUL	
P. brevicornis	PH-M164	OT 1-81	Museum	1984	E. Beringia	69.22	-140.07	Hind leg	6	*	CNC	RHUL	
P. brevicornis	PH-M165	OT 1-88	Museum	1978	E. Beringia	65.67	-149.10	Hind leg	3	1	CNC	RHUL	
P. brevicornis	PH-M166	OT 1-95	Museum	1961	E. Beringia	68.13	-165.97	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M167	OT 1-99	Museum	1978	E. Beringia	61.77	-149.32	Hind leg	9	*	CNC	RHUL	
P. brevicornis	PH-M168	OT 2-1	Museum	1954	Bering Islands	60.40	-172.72	Hind leg	9	1	CNC	RHUL	
P. brevicornis	PH-M169	OT 2-9	Museum	1987	Appalachians	44.25	-71.32	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M170	OT 2-11	Museum	1987	Appalachians	44.30	-71.32	Hind leg	5	*	CNC	RHUL	
P. brevicornis	PH-M171	OT 3-89	Museum	1972	W. of Hudson Bay	61.90	-121.52	Hind leg	7	*	CNC	RHUL	
P. brevicornis	PH-M172	OT 3-94	Museum	1972	W. of Hudson Bay	68.22	-135.90	Hind leg	13	*	CNC	RHUL	
P. brevicornis	PH-M173	OT 3-96	Museum	1948	W. of Hudson Bay	69.50	-133.78	Hind leg	3	*	CNC	RHUL	
P. brevicornis	PH-M174	OT 3-98	Museum	1991	E. Beringia	63.50	-137.03	Hind leg	9	*	CNC	RHUL	
P. brevicornis	PH-M175	OT 4-1	Museum	1984	E. Beringia	69.45	-130.43	Hind leg	11	1	CNC	RHUL	
P. brevicornis	PH-M176	OT 4-3	Museum	1968	E. Beringia	61.80	-140.03	Hind leg	9	*	CNC	RHUL	
P. brevicornis	PH106	N/A	Ancient	1985	85 IGU-M-5	58.90	-159.07	0.5 pronotum			S. Elias	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
P. brevicornis	PH107	N/A	Ancient	?	K/BR-24	61.88	-157.85	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH108	N/A	Ancient	?	NBI 22-25	70.12	-143.67	Pronotum			S. Elias	RHUL	
P. brevicornis	PH109	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	RHUL	
P. brevicornis	PH110	N/A	Ancient	?	CP 2-9	58.20	-157.45	Pronotum			S. Elias	RHUL	
P. brevicornis	PH111	N/A	Ancient	1983	83FF 3-M-13	58.73	-158.38	Pronotum			S. Elias	RHUL	
P. brevicornis	PH112	N/A	Ancient	2005	Al-3-05-B13	66.02	-149.08	Pronotum			S. Elias	RHUL	
P. brevicornis	PH113	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	RHUL	
P. brevicornis	PH114	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	RHUL	
P. brevicornis	PH115	N/A	Ancient	?	K/BR-22	61.88	-157.85	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH116	N/A	Ancient	2005	Al-4-05-B21	65.10	-153.25	Pronotum			S. Elias	RHUL	
P. brevicornis	PH117	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	RHUL	
P. brevicornis	PH118	N/A	Ancient	1984	84 IGU 1-5	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH119	N/A	Ancient	2006	DF-06-90	67.80	-130.85	Pronotum			S. Elias	RHUL	
P. brevicornis	PH120	N/A	Ancient	2005	Al-4-05-B21	65.10	-153.25	Pronotum			S. Elias	RHUL	
P. brevicornis	PH121	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	Pronotum			S. Elias	RHUL	
P. brevicornis	PH122	N/A	Ancient	?	K/BR-24	61.88	-157.85	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH123	N/A	Ancient	1984	84 IGU 1-5	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH124	N/A	Ancient	1984	84 IGU 1-4	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH125	N/A	Ancient	1983	83FF 3-M-13	58.73	-158.38	Pronotum			S. Elias	RHUL	
P. brevicornis	PH130	N/A	Ancient	1985	85 IGU-M-5A	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH131	N/A	Ancient	2005	Al-4-05-B21	65.10	-153.25	Pronotum			S. Elias	RHUL	
P. brevicornis	PH132	N/A	Ancient	2006	DF-06-101	67.48	-140.10	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH133	N/A	Ancient	?	12-M4	58.62	-158.25	Pronotum			S. Elias	RHUL	
P. brevicornis	PH134	N/A	Ancient	1983	83 IGU 11-3	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH135	N/A	Ancient	1983	83FF 1-5	58.73	-158.38	Pronotum			S. Elias	RHUL	
P. brevicornis	PH136	N/A	Ancient	2006	DF-06-90	67.80	-130.85	Pronotum			S. Elias	RHUL	
P. brevicornis	PH137	N/A	Ancient	1983	83 IGU 11-3	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH138	N/A	Ancient	1985	85 IGU-M-1	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH139	N/A	Ancient	?	K/BR-24	61.88	-157.85	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH140	N/A	Ancient	1983	83FF 3-M-13	58.73	-158.38	Pronotum			S. Elias	RHUL	
P. brevicornis	PH141	N/A	Ancient	2005	Al-4-05-B21	65.10	-153.25	Pronotum			S. Elias	RHUL	
P. brevicornis	PH142	N/A	Ancient	?	K/BR-22	61.88	-157.85	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH143	N/A	Ancient	2006	DF-06-101	67.48	-140.10	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH144	N/A	Ancient	1983	83 IGU 11-3	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH145	N/A	Ancient	1984	84 IGU 1-4	58.90	-159.07	0.5 pronotum			S. Elias	RHUL	

Species	Specimen	Museum reference	Age class	Collection year	Region/ Locality	Lat.	Long.	Portion extracted	Mt Ht	Nu Ht	Origin	Lab	Notes
P. brevicornis	PH146	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	RHUL	
P. brevicornis	PH147	N/A	Ancient	1983	83 IGU 11-3	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH148	N/A	Ancient	1984	84 IGU 1-4	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH149	N/A	Ancient	1984	84 IGU 1-4	58.90	-159.07	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH154	N/A	Ancient	1984	84 IGU 1-5	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH155	N/A	Ancient	?	NBI 22-25	70.12	-143.67	Pronotum			S. Elias	RHUL	
P. brevicornis	PH156	N/A	Ancient	1983	83 IGU 1-7	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH157	N/A	Ancient	2006	DF-06-101	67.48	-140.10	Pronotum			S. Elias	RHUL	
P. brevicornis	PH158	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH159	N/A	Ancient	1984	84 IGU 1-5	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH160	N/A	Ancient	1983	83 IGU 11-3	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. brevicornis	PH161	N/A	Ancient	2005	Al-3-05-B13	66.02	-149.08	Pronotum			S. Elias	RHUL	
P. brevicornis	PH162	N/A	Ancient	2006	GZ-06-57	67.43	-140.37	0.5 pronotum			S. Elias	RHUL	
P. brevicornis	PH163	N/A	Ancient	1983	83 IGU 1-7	58.90	-159.07	Pronotum			S. Elias	RHUL	
P. pinguedineus	LM13	N/A	Modern	2002	W. Beringia	72.92	127.02	Whole	*	*	S. Kuzmina	RHUL	
P. pinguedineus	LM80	N/A	Museum	1995	W. Beringia	64.73	177.48	Hind leg	*	*	S. Elias	RHUL	
P. ventricosus	LM14	N/A	Modern	2002	W. Beringia	72.92	127.02	Whole	*	*	S. Kuzmina	RHUL	

Appendix 2. List of Primers and Primer Sets

2a. Primers

Primers used in this thesis. All primers are novel, except that highlighted in bold, which was from Gilbert *et al*. (2007).

Primer	Sequence (5' to 3')	Marker	Species
A_COI_0_F	CCTGAAGTTTATATTTTAATTTTA	COI	*A. alpina*
A_COI_0_R	CATCTATACCTACAGTAAATATATGA	COI	*A. alpina*
A_COI_1_F	GCCATTGGATTATTAGGATTTG	COI	*A. alpina*
A_COI_1_R	TTTTAATTCCTGTAGGGACAGC	COI	*A. alpina*
A_COI_3_F	CGAGCTTATTTTACTTCAGC	COI	*A. alpina*
A_COI_3b_R	CCTGTTAATCCTCCRACT	COI	*A. alpina*
A_COI_4_F	CTGTCCCTACAGGAATTAAAAT	COI	*A. alpina*
A_COI_4b_F	TGCTGTACCTACAGGAATTAA	COI	*A. alpina*
A_COI_4b_R	AATATCAAGGGATGAATTAGC	COI	*A. alpina*
A_COI_5b_F	TTCATTACTTTGAGCTTTAGG	COI	*A. alpina*
A_COI_5_R	AATCATTGAATAAATCCTGC	COI	*A. alpina*
A_COI_6_F	TATGTTGTTGCTCATTTTCATTATG	COI	*A. alpina*
A_COI_6_R	GGGAAAAAAGTTAAATTTACTCCAA	COI	*A. alpina*
A_COI_7_F	GCAGGATTTATTCAATGATT	COI	*A. alpina*
A_COI_7_R	ATCTGAATATCGTCGAGGTA	COI	*A. alpina*
A_COI_8_F	TTGGAGTAAATTTAACTTTTTTCC	COI	*A. alpina*
A_COI_8_R	AAAATTACTAGACGTTGAGAAATAAA	COI	*A. alpina*
A_COI_9b_F	GTTCAACTATTTCTTTTATTGG	COI	*A. alpina*
A_COI_9b_F	GGTAATTCAGAGTATCTATGTTC	COI	*A. alpina*
P_COI_1_F	ATAATTTATGCTATATTAGCTATTG	COI	*P. (Cryobius) sp.*
P_COI_1_R	CATGAAAAAATTTTAATTCC	COI	*P. (Cryobius) sp.*
P_COI_2_F	TCATATATTTACAGTTGGAATAGA	COI	*P. (Cryobius) sp.*
P_COI_2_R	TAAATAAAAAAACAAATCCTAAAG	COI	*P. (Cryobius) sp.*
P_COI_3_F	GAATTAAAATTTTTTCATGATTA	COI	*P. (Cryobius) sp.*
P_COI_3b_R	ACATAATAAGTATCATGAAGAATAAT	COI	*P. (Cryobius) sp.*
P_COI_3c_F	TTAGCAACTCTTCATGGTG	COI	*P. (Cryobius) sp.*
P_COI_3c_R	ATACATAATGAAAATGAGCAAC	COI	*P. (Cryobius) sp.*
P_COI_4_F	TACAGTAGGAGGATTAACTGGAGTA	COI	*P. (Cryobius) sp.*
P_COI_4_R	TGAATAAAGGAAATCATTGAATAAA	COI	*P. (Cryobius) sp.*
P_COII_0_F	TCGATATTTATTAGAAGGTCAAA	COII	*P. (Cryobius) sp.*
P_COII_0_R	TCATAACTTCAATATCATTGATG	COII	*P. (Cryobius) sp.*
P_COII_1_F	GTAAGAAATCCTTCAGTTACTTTAA	COII	*P. (Cryobius) sp.*
P_COII_1_R	AAGGTAATACAATTCGATTATCTAC	COII	*P. (Cryobius) sp.*
P_COII_2_F	TGATTCATATATAATTCCAACAAATG	COII	*P. (Cryobius) sp.*
P_COII_2_R	ATTTTTACACCTAAGGCAGGA	COII	*P. (Cryobius) sp.*
P_COII_3_F	TTACCTTTCAATACACAAATTCG	COII	*P. (Cryobius) sp.*
P_COII_3_R	AACATTGTCCATAAAATAATCCA	COII	*P. (Cryobius) sp.*
28S_1d_F	GACCAAGGAGTCTAGCAT	28S	*A. alpina* / *P. (Cryobius) sp.*
D3R	**GCATAGTTCACCATCTTTC**	**28S**	***A. alpina* / *P. (Cryobius) sp.***
A_28S_1_F	GCAAGTCATTGGGACTATT	28S	*A. alpina*
A_28S_1_R	ATGCGAGCCAACATAAA	28S	*A. alpina*
A_28S_2_F	GTAAATGTTTGTTGAATTTTCGT	28S	*A. alpina*
A_28S_2_R	GCGTGTACGCTCTTGGT	28S	*A. alpina*
P_28S_1_F	AGTCATTGGGACTCTGTTAAAAC	28S	*P. (Cryobius) sp.*
P_28S_1_R	ATGCGGGGTGACATAAAT	28S	*P. (Cryobius) sp.*
P_28S_2_F	TAACCGGATCACATTGAAT	28S	*P. (Cryobius) sp.*
P_28S_2_R	CGTGTACGCTCTTGGTG	28S	*P. (Cryobius) sp.*
ITS1_2_F	TACTAGTTCAACGGTTGGAAAT	ITS1	*A. alpina* / *P. (Cryobius) sp.*
ITS1_2_ R	CACATTATCTGGAGTTTCAGAATAC	ITS1	*A. alpina* / *P. (Cryobius) sp.*
ITS1_3_F	CAAAAAGCAAACGAAAGTCTACAAG	ITS1	*A. alpina* / *P. (Cryobius) sp.*
ITS1_3_ R	TACTCGGACGAGCCAGAATC	ITS1	*A. alpina* / *P. (Cryobius) sp.*

Primer	Sequence (5' to 3')	Marker	Species
A_ITS1_1_F	GCATCTGCAGCAGGTATAT	ITS1	*A. alpina*
A_ITS1_1_R	GTTTCAGAATACGCGAGG	ITS1	*A. alpina*
A_ITS1_2_F	AGTCTACAAGAATTTGAAGCTG	ITS1	*A. alpina*
A_ITS1_2_R	GTTTCTTTATCTCGGAGGC	ITS1	*A. alpina*
P_ITS1_1_F	TTATGGACACGCACATCT	ITS1	*P. (Cryobius) sp.*
P_ITS1_1_R	GAATATGCTAGGCGAACTAAA	ITS1	*P. (Cryobius) sp.*
P_ITS1_2_F	CTGAAAGTCTGCAAGAATTTG	ITS1	*P. (Cryobius) sp.*
P_ITS1_2_R	TTTATCTCGGAGACCATCAA	ITS1	*P. (Cryobius) sp.*

2b. Primer Sets

Primer sets used in this thesis. Forward and reverse primers are listed in appendix 2a. T_a: Annealing temperature. Product size does not include primer sequence length.

Set №	Forward	Reverse	T_a (C)	Product size (bps)	Species
1	A_COI_0_F	A_COI_0_R	55	125	*A. alpina*
2	A_COI_0_F	A_COI_3b_R	52	302	*A. alpina*
3	A_COI_1_F	A_COI_1_R	54	80	*A. alpina*
4	A_COI_1_F	A_COI_6_R	55	399	*A. alpina*
5	A_COI_3_F	A_COI_3b_R	58	123	*A. alpina*
6	A_COI_3_F	A_COI_5_R	54	247	*A. alpina*
7	A_COI_3_F	A_COI_7_R	50	383	*A. alpina*
8	A_COI_4b_F	A_COI_4b_R	54	118	*A. alpina*
9	A_COI_4_F	A_COI_8_R	55	466	*A. alpina*
10	A_COI_5b_F	A_COI_5_R	55	142	*A. alpina*
11	A_COI_5b_F	A_COI_7_R	50	278	*A. alpina*
12	A_COI_6_F	A_COI_6_R	55	117	*A. alpina*
13	A_COI_7_F	A_COI_7_R	54	116	*A. alpina*
14	A_COI_7_F	A_COI_9b_R	56	313	*A. alpina*
15	A_COI_8_F	A_COI_8_R	51	146	*A. alpina*
16	A_COI_9b_F	A_COI_9b _R	56	112	*A. alpina*
17	P_COI_1_F	P_COI_1_R	54	107	*P. (Cryobius) sp.*
18	P_COI_1_F	P_COI_3_R	48	251	*P. (Cryobius) sp.*
19	P_COI_1_F	P_COI_3b_R	49	251	*P. (Cryobius) sp.*
20	P_COI_2_F	P_COI_2_R	55	128	*P. (Cryobius) sp.*
21	P_COI_2_F	P_COI_3_R	49	199	*P. (Cryobius) sp.*
22	P_COI_2_F	P_COI_3b_R	54	199	*P. (Cryobius) sp.*
23	P_COI_3_F	P_COI_3_R	48	120	*P. (Cryobius) sp.*
24	P_COI_3_F	P_COI_3b_R	52	120	*P. (Cryobius) sp.*
25	P_COI_3c_F	P_COI_3c_R	58	131	*P. (Cryobius) sp.*
26	P_COI_4_F	P_COI_4_R	55	108	*P. (Cryobius) sp.*
27	P_COII_0_F	P_COII_0_R	58	125	*P. (Cryobius) sp.*
28	P_COII_0_F	P_COII_1_R	58	233	*P. (Cryobius) sp.*
29	P_COII_1_F	P_COII_1_R	58	119	*P. (Cryobius) sp.*
30	P_COII_1_F	P_COII_2_R	56	202	*P. (Cryobius) sp.*
31	P_COII_1_F	P_COII_3_R	56	274	*P. (Cryobius) sp.*
32	P_COII_2_F	P_COII_2_R	52	115	*P. (Cryobius) sp.*
33	P_COII_2_F	P_COII_3_R	48	187	*P. (Cryobius) sp.*
34	P_COII_3_F	P_COII_3_R	55	114	*P. (Cryobius) sp.*
35	28S_1d_F	D3R	60	183	*A. alpina*
36	28S_1d_F	D3R	60	189	*P. (Cryobius) sp.*
37	A_28S_1_F	A_28S_1_R	60	81	*A. alpina*
38	A_28S_2_F	A_28S_2_R	60	87	*A. alpina*
39	P_28S_1_F	P_28S_1_R	60	79	*P. (Cryobius) sp.*
40	P_28S_2_F	P_28S_2_R	60	75	*P. (Cryobius) sp.*
41	P_28S_1_F	P_28S_2_R	60	133	*P. (Cryobius) sp.*
42	ITS1_2_F	ITS1_2_R	60	120	*A. alpina*
43	ITS1_2_F	ITS1_2_R	48	120	*P. (Cryobius) sp.*
44	ITS1_3_F	ITS1_3_R	60	127	*A. alpina*
45	ITS1_3_F	ITS1_3_R	49	127	*P. (Cryobius) sp.*
46	A_ITS1_1_F	A_ITS1_1_R	60	80	*A. alpina*
47	A_ITS1_2_F	A_ITS1_2_R	55	84	*A. alpina*
48	A_ITS1_1_F	A_ITS1_2_R	55	145	*A. alpina*
49	P_ITS1_1_F	P_ITS1_1_R	59	84	*P. (Cryobius) sp.*
50	P_ITS1_2_F	P_ITS1_2_R	60	84	*P. (Cryobius) sp.*

Appendix 3. Approximate Permutation Test Script

Example of the java script used to calculate the probabilities of the differences observed between misincorporation rates in chapter three. The script was implemented in *Statistics101* (http://www.statistics101.net/).

```
DATA (0 1 2 3 4 5 6) combineddata
REPEAT 100000
  SHUFFLE combineddata combineddatashuffled
  TAKE combineddatashuffled 1,3 sample1
  MEAN sample1 sample1mean
  TAKE combineddatashuffled 4,7 sample2
  MEAN sample2 sample2mean
  SUBTRACT sample2mean sample1mean meandifference
  ABS meandifference absmeandiff
  SCORE absmeandiff meandifferences
END
COUNT meandifferences >= 4 greaterdifferences
LET probability = greaterdifferences / 100000
PRINT probability
```

Appendix 4. Parameter Optimisation for Bowtie2

Optimisation of Bowtie2 parameters for alignment of reads to the short, multi-copy reference sequences of chapter four. Optimal parameters were those that mapped the maximum number of reads in the minimum amount of CPU time, and are shown in bold. Parameters were tested against the youngest and oldest samples in the dataset [Modern (Mo) 1, Ancient (An) 2]. Optimal parameters were: -D 20 -R 3 -N 1 -L 20 -i S,1,0.50.

Bowtie	**Mo1**		**An2**	
Parameters	**Mapped reads**	**CPU Time**	**Mapped reads**	**CPU Time**
-N 1 -L 10	16455	7236.79	311	2215.89
-N 1 -L 15	16506	1470.77	325	488.83
-N 1 -L 20	**16506**	**388.39**	**325**	**145.46**
-N 1 -L 25	16493	357.71	**325**	**137.48**
-N 0 -L 5	14713	6931.13	289	2271.03
-N 0 -L 10	16506	1505.7	325	363.43
-N 0 -L 15	16478	249.65	**325**	**99.11**
-N 0 -L 20	16338	234.22	321	98.8

Appendix 5. NuDNA Accession Numbers

5a. Accession numbers for *Amara* nuDNA Alignment

Genbank accession numbers for the non-*alpina Amara* specimens that were used in the nuDNA alignment of chapter six. Bold indicates representative sequence used in the alignment. Specimens are listed in the same order as they appear in the alignment.

Species	Genbank	Reference
Amara (Curtonotus) sp.	**AF398694**	**Ober (2002)**
Amara (Curtonotus) aulica	**GU347386**	**Raupach *et al*. (2010)**
Amara (Curtonotus) aulica	GU347387	Raupach *et al*. (2010)
Amara chalcites	**AB243496**	**Sasakawa & Kubota (2007)**
Amara aenea	**FJ173093**	**Ruiz *et al*. (2009)**
Amara erratica	**GU347389**	**Raupach *et al*. (2010)**
Amara erratica	GU347388	Raupach *et al*. (2010)
Amara erratica	GU347390	Raupach *et al*. (2010)
Amara erratica	GU347391	Raupach *et al*. (2010)
Amara erratica	GU347392	Raupach *et al*. (2010)
Amara anthobia	**GU347385**	**Raupach *et al*. (2010)**
Amara anthobia	GU347384	Raupach *et al*. (2010)
Amara similata	GU347398	Raupach *et al*. (2010)
Amara similata	GU347399	Raupach *et al*. (2010)
Amara similata	GU347400	Raupach *et al*. (2010)
Amara glacialis	**N/A**	**This thesis (LM15)**
Amara quenseli	**GU347394**	**Raupach *et al*. (2010)**
Amara quenseli	GU347393	Raupach *et al*. (2010)
Amara quenseli	GU347395	Raupach *et al*. (2010)
Amara quenseli	GU347396	Raupach *et al*. (2010)
Amara quenseli	GU347397	Raupach *et al*. (2010)

5b. Accession numbers for *Pterostichus (Cryobius)* nuDNA Alignment

Genbank accession numbers for the non-*brevicornis Pterostichus (Cryobius)* specimens that were used in the nuDNA alignment of chapter seven. Bold indicates representative sequence used in the alignment. Specimens are listed in the same order as they appear in the alignment.

Species	Genbank	Reference
P. riparius	**EU142444**	**Will & Gill (2008)**
P. riparius	EU142445	Will & Gill (2008)
P. pinguedineus	Novel	This thesis (LM13)
P. pinguedineus	Novel	This thesis (LM80)
P. kurosawai	AB243485	Sasakawa & Kubota (2007)
P. ventricosus	**Novel**	**This thesis (LM14)**
P. nemoralis	**FJ173089**	**Ruiz *et al*. (2009)**

References for Appendices

Gilbert MTP, Moore W, Melchior L, Worobey M (2007) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS One* **2**, e272.

Mack L (2008) *Investigating mitochondrial DNA phylogenies of Arctic and European beetle taxa* MSc thesis, Mainz University.

Ober KA (2002) Phylogenetic relationships of the carabid subfamily Harpalinae (Coleoptera) based on molecular sequence data. *Molecular Phylogenetics and Evolution* **24**, 228-248.

Raupach MJ, Astrin JJ, Hannig K, *et al*. (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* **7**, 1-15.

Ruiz C, Jordal B, Serrano J (2009) Molecular phylogeny of the tribe Sphodrini (Coleoptera: Carabidae) based on mitochondrial and nuclear markers. *Molecular Phylogenetics and Evolution* **50**, 44-58.

Sasakawa K, Kubota K (2007) Phylogeny and genital evolution of carabid beetles in the genus *Pterostichus* and its allied genera (Coleoptera: Carabidae) inferred from two nuclear gene sequences. *Annals of the Entomological Society of America* **100**, 100-109.

Thomsen PF, Elias S, Gilbert MT, *et al*. (2009) Non-destructive sampling of ancient insect DNA. *PLoS One* **4**, e5048.

Will KW, Gill AS (2008) Phylogeny and classification of *Hypherpes auctorum* (Coleoptera : Carabidae : Pterostichini : *Pterostichus*). *Annals of Carnegie Museum* **77**, 93-127.