THE PhD SCHOOL OF SCIENCE - FACULTY OF SCIENCE
UNIVERSITY OF COPENHAGEN, DENMARK
Natural History Museum of Denmark

PHD THESIS

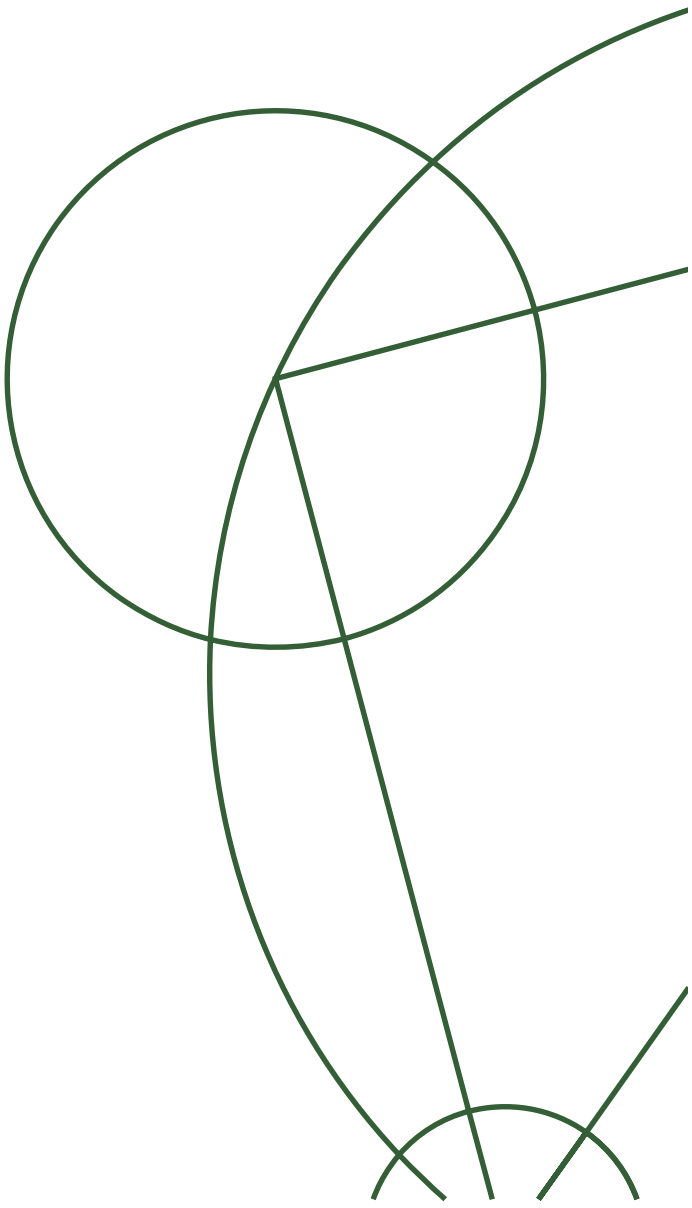
ASTRID M. Z. SCHMIDT

# ICE CORE GENETICS

## UNCOVERING BIODIVERSITY FROM ANCIENT ECOSYSTEMS IN GREENLAND AND ANTARCTICA

ACADEMIC ADVISOR: ESKE WILLERSLEV
CO-SUPERVISOR: JØRGEN PEDER STEFFENSEN

SUBMITTED: 16/12/2011



PhD Thesis

# Ice Core Genetics
# Uncovering Biodiversity from Ancient Ecosystems in Greenland and Antarctica

ASTRID M. Z. SCHMIDT

Centre for Ice and Climate, Niels Bohr Institute
Centre for GeoGenetics, Natural History Museum of Denmark
UNIVERSITY OF COPENHAGEN

Front page by Paula Campos

## Preface

This thesis is the product of my scientific efforts during three and a half years as a phD student. The work has been conducted as a shared project between the two research groups; Centre for GeoGenetics, Natural History Museum of Denmark and Centre for Ice and Climate, Niels Bohr Institute within the University of Copenhagen. Main supervisor on the project was Professor Eske Willerslev and co-supervisor was Associate Professor Jørgen Peder Steffensen. Funding was shared between the Faculty of Science, Copenhagen University and the two research groups funded by the Danish National Science Foundation.

The aim of the thesis has been to develop and expand the research field of ancient DNA in basal ice of ice cores by use of new sequencing and analytical techniques. The fairly young research field of studying ancient genetic remains in basal ice was initiated by previous studies on Holocene and Pleistocene ice from Greenland by the people involved in both research groups. These studies used regular cloning and Sanger sequencing and retrieved clonal sequence diversity in the order of a few hundred sequences. With the aid of next-generation sequencing on the GS-FLX (Roche) platform the data that can be produced are several orders of magnitude larger and the sensitivity to detect rare ancient DNA molecules has increased accordingly. In response to the large amounts of sequence data that can now be generated the demand increases for verification of the data in terms of sorting, filtering for damage and error signals as well as dealing with contamination issues. The majority of the work in the thesis has been divided between laboratory work and data analysis. Moreover the project has involved cross-disciplinary work with research groups conducting dating measures of the samples studied and a study on abiotic characteristics of basal ice samples.

The structure of the thesis follows the different aspects of reconstructing palaeo-environments from ancient DNA preserved in ice cores and permafrost to which my work has contributed. The thesis consists of seven chapters presented as manuscripts and two text chapters presenting and evaluating the methods applied to interpret the data. Chapter one introduces the broad field of research on biology within ice and reviews how this field has developed, what the implications are and the future of the joint fields of biology and climate research. Chapter two is a drafted manuscript on the chemical properties of soil that are indicative of the likelihood that a basal sample contains ancient genetic remains from a vegetated past ecosystem. In this chapter I did not conduct the laboratory work, but I have designed the overall study and written the manuscript. Chapters three and four present the methods applied to the basal ice and permafrost samples and critically discuss the analysis techniques applied for detecting past biodiversity by three different analytical approaches. Before the three main chapters presenting the ancient DNA findings in samples from Greenland and Antarctica there is a brief introduction to the studied locations. The seven chapters are followed by a general conclusions and future perspectives for this research field. Last, there is a section of appendices where additional data is presented for the three main chapters as well as a section introducing the reasons for choosing the different dating methods followed by thorough descriptions of these dating methods that were conducted by collaborating research groups.

Astrid M. Z. Schmidt, December 2011, Copenhagen

**Acknowledgements**

The last three and a half year have been a journey full of adventures, lessons and development. I feel lucky to have had the opportunity to walk this path in two such inspiring and creative environments as Centre for GeoGenetics and Centre for Ice and Climate.
Here is to all the people who contributed and provided support and interest in this project.

First of all thank you to my two supervisors Eske Willerslev and Jørgen Peder Steffensen for your enthusiasm, motivation and support, and for believing in this project. I have enjoyed the years of research and am thankful for your sharing of time and knowledge and for giving me the opportunity to do fieldwork on the NEEM ice core project on top of the Greenland Ice Sheet. It was an amazing experience and broadened my perception for the range of niches in Ice Core research. Moreover, it was an inspiring example on how teamwork produces high quality research even in the most extreme and remote areas.

Thank you to the team from Laboratoire de Glaciologie ULB, Brussels, Belgium for sharing sample material from Antarctica. Within this department I thank Jean-Louis Tison, Reginald Lorrain and Denis Samyn for valuable discussions on basal ice dynamics and processes and the local history of the samples, not to mention helpful sharing of previous findings and new ideas.

Within the Centre for Ice and Climate I thank all the people who created a wonderful work environment and in particular Henrik Clausen, Bo Vinter, Anders Svensson for valuable discussions on ice sheet dynamics, old ice core records and the local history of the Camp Century samples. Moreover, I thank Dorthe-Dahl Jensen for running temperature models to provide temperature histories on the ice sample sites of Greenland and Antarctica.

Within the Centre for GeoGenetics I would like to thank all the lovely people with whom I have spend many hours in the labs, behind the computers, in fruitful idea sharing and good laughs. In particular I would like to thank Tina Brand (Centre for GeoGenetics) for spending long hours in the freezer when helping me processing the samples. A special thank you to Morten Rasmussen, Aurélien Ginolhac and Ludovic Orlando for developing pipelines to sort, filter and assign the data and because you introduced and taught me basic skills within the world of perl and python scripts. Thank you Mikkel for all the hours we spend side by side testing scripts on our data. Moreover, Rute and Tobias I am grateful for valuable contributions and discussions on data. In particular I would like to add a special thank you to Per Hartvig for providing valuable considerations on the plant history and ecosystem niches. Last, I would like to thank Svend Funder and Kurt Kjær for discussions on the palaeo-environmental history of Greenland and for helpful comments.
Thank you to the "dirt DNA" people Mikkel, Kenneth, Mette and Tina for endless support, sharing of ideas, comments and discussions. Thank you to Tom for guiding me on the path of writing scientifically with valuable comments.
Last a special thank you to Maanasa who sat by my side writing her own thesis and provided great support and help in terms of edits and discussions on the final product. Also to Paula for the fine front cover and for your comments and edits, I am most grateful.

**Acknowledgements**

Thank you to Mike Bunce and Morten Allentoft for processing and replicating the samples in Murdoch University clean lab, Perth, Australia.

The cross-disciplinary work involved geochemical analysis of basal ice samples and dating of the samples involved in this study. I would like to thank Bo Elberling for valuable discussions and guidance from the collaboration we performed on the soil chemical properties of basal ice samples (Dep. of Geography and Geology, University of Copenhagen).
Several research groups involved in the different basal ice dating invested a huge effort and I would like to thank the following:

$^{10}Be/^{36}Cl$ dating team: Raimund Muscheler (Lund University), Ann-Marie Berggren and Ala Aldahan (Uppsala University).

OSL dating team: Christine Thiel, Andrew Murray and Jan Pieter Buylaert (The Luminescence Research Laboratory, Risø National Laboratory).

Uranium recoil dating: Sarah Aciego for her contribution (University of Michigan).

Amino Acid Racemization dating (attempt): Beatrice Demarchi, Matthew Collins, Enrico Cappellini and Kirsty Penkman, (University of York).

I'd especially like to thank Christine, Raimund, Ann-Marie and Beatrice for your openness and for your time to discuss and share your knowledge.

In addition I thank Jocelyne Bourgeois for conducting pollen analysis on the Camp Century basal ice samples and enjoyable discussions at the fieldwork on NEEM (Geological Survey of Canada, Ottawa).

I thank the Centres for supporting me with an extension during this PhD and in particular Dorthe-Dahl Jensen for providing funding for another 2 year postdoc where I can continue on the bridge between two such exciting research fields of aDNA and Ice Core studies. I am most grateful for this opportunity.

Above all I thank all my colleagues - for a wonderful time in your company during this period of my life. My friends and family – I owe you the deepest gratitude for standing by me with love, care and support in times of ups and downs. I could not have done this without you all.

TABLE OF CONTENTS

**Summary**

Reconstructions of palaeo-ecosystems have traditionally been approached by analysis of pollen and macrofossil remains from geological layers representing the past environments. These traditional methods are very specific in terms of the taxonomic level to which the fossil material can be assigned and can provide climatic and ecological details from the composition of the detected remains. However, the methods are inadequate when accessibility to sample material is limited or when the past biota did not leave a representative fossil record behind. Ancient DNA (aDNA) methods have the potential to detect past biodiversity in areas where sediments have adsorbed genetic degraded material from dead organisms and preserved it in a frozen state. Polar areas with low temperatures and dry climates offer optimal DNA and fossil preservation condition but access have only been possible on the margins of the ice sheets that cover Greenland and Antarctica. Amplification of ancient genetic material from ice cores can therefore contribute with details on past ecological characteristics, biodiversity and age of the periods where the ice sheets retreated in response to warming climates.

In this thesis I present a review of the major contributions and implications of studies on biological contents within ice. Then I present a study on the abiotic factors of basal ice from different deep drilling sites in Greenland that verify soil chemical characteristics of the sites producing aDNA results. Additionally, we find that within frozen basal ice, chemical factors that have formerly shown to increase the DNA degradation rate (non neutral pH and oxidation) are not of critical importance for DNA survival over a time scale of at least 400,000 years. The methods involved in analyzing aDNA sequence data through different filtering and assignment pipelines are evaluated and point to the critical importance of how to deal with possible false positive results. The major projects in this work have been the reconstruction of three ecosystems, one in Antarctica and two in Greenland. Ecological and climatic details were assessed from the composition of biodiversity detected by means of aDNA amplification of mini-barcodes and deep sequencing on the GS-FLX sequencing platform. Another side project has been the cross-disciplinary effort in age determination of the basal ice samples by three dating methods; cosmogenic $^{10}Be/^{36}Cl$ dating, recoil dating of $^{234}U/^{238}U$ and optical stimulated luminescence (OSL) dating.

Results from the thesis verified the existence of aDNA in the sample material and provided the following outcomes for each of the sites. Camp Century in Northwestern Greenland had a northern temperate to boreal ecosystem with indicator plant taxa constraining the mean annual temperature (MAT) conditions to 5 - 15°C. Moreover, the results from the Kap København formation in Northern Greenland supplemented and confirmed the climate defined from previous fossil findings and showed an overlap to the Camp Century taxa. Comparisons of the amplified biodiversity from the two sites in Greenland with fossil findings suggested a co-existence of these ecosystems and re-opened the discussion of when the Greenland Ice Sheet formed. We propose an age of around 1 million years (MY) for these sites.
In Antarctica did the aDNA results from basal ice in Suess Glacier in the Dry Valleys point to a mixed boreal/taiga palaeo-ecosystem with MAT around –10°C and summer temperatures that allowed tree growth prior to cooling and ice coverage. The preservation conditions at this location are optimal with average temperatures below –20°C. The results complemented a climate period suggested by fossil findings in the Transantarctic Mountains and thereby indirectly provide support for a debated warm climate period where the West Antarctic Ice Sheet was absent approximately 3MY.

**Resumé**

Rekonstruktion af palaeo-økosystemer har traditionelt været baseret på analyser af pollen og makrofossile fund fra stratigrafiske sediment lag der repræsenterer de forhistoriske miljøer. Disse traditionelle metoder er meget specifikke i forhold til det taksonomiske niveau man kan bestemme de fossile fund til. Desuden kan de indikere klimatiske og økologiske detaljer udfra sammensætningen af de fossile fund. Der er dog begrænsninger i de traditionelle metoders anvendelighed når tilgængeligheden til fossiler er begrænset ,eller hvis den forhistoriske biota ikke efterlod sig et repræsentativt fossilt materiale. Ancient DNA (aDNA) metoder har potentiale til at detektere forhistorisk biodiversitet i områder hvor sedimentet har absorberet genetisk nedbrudt materiale fra døde organismer og siden bevaret det under frosne forhold. Især i polare områder med lave temperaturer og et tørt klima er der optimale bevaringsforhold for DNA. Grønland og Antarktis er dækket af store iskapper som begrænser adgangen til fossile fund til randområderne af iskapperne. Derfor kan nye palaeo-økologiske genetiske data fra iskerner bidrage med detaljer om både biodiversitet og alder for de varme inter-glaciale perioder hvor iskapperne trak sig tilbage.

I denne afhandling præsenterer jeg et review over de vigtigste bidrag fra studier på iskerners biologiske indhold og implikationerne af disse fund. Herefter præsenterer jeg et studie af abiotiske faktorer i bundis fra forskellige dybe iskerne-boringer i Grønland som understøtter kemiske jord-karakteristikker af de lokaliteter der giver plante aDNA resultater. Dette studie viser ydermere at i frossen bundis, har de kemiske forhold som normalt kendetegner en øget DNA degradering (ikke-neutral pH og oxidation), ikke stor indflydelse på DNA bevaring på en tidsskala af mindst 400.000 år. Efterfølgende bliver de forskellige analyse metoder som er anvendt i studiet til at sortere og bestemme biodiversiteten i aDNA datasættet evalueret og jeg påpeger vigtigheden af at tage højde for falske positive resultater. Hovedprojekterne i denne phD har være rekonstruktionen af tre økosystemer, ét i Antarktis og to i Grønland. På baggrund af tre forskellige aDNA analyse metoder kunne økologiske og klimatiske detaljer af de forhistoriske miljøer beskrives ud fra sammensætningen af biodiversiteten. Et tværfagligt sideprojekt har været aldersbestemmelse af bundis-prøverne ved hjælp af tre forskellige dateringsmetoder; cosmogen $^{10}Be/^{36}Cl$ datering, Uran-Uran datering af $^{234}U/^{238}U$ og optisk stimuleret luminiscence (OSL) datering.

Resultaterne fra afhandlingen verificerede eksistensen af aDNA i prøvematerialet og gav følgende resultater for hvert af de tre lokaliteter. Camp Century i nordvest Grønland havde et nordligt tempereret til borealt økosystem med plante indikatorarter der indskrænker den årlige middel temperatur (ÅMT) til 5 - 15°C. Resultaterne fra Kap København formationen i Nordgrønland komplementerede og bekræftede det klima som tidligere blev defineret udfra fossile fund og viste yderligere et overlap til plante taxa fra Camp Century lokaliteten. Sammenligninger af den amplificerede biodiversitet fra de to Grønlandske lokaliteter med de fossile fund foreslår en sam-eksistens af disse to økosystemer, som derved genåbner en diskussion af alderen for den Grønlandske iskappe. Vi foreslår en alder omkring 1 million år for begge lokaliteter. Antarktis prøverne har haft optimale bevaringsforhold i Dry Valleys med temperaturer under –20°C. aDNA resultaterne fra bundisen i Suess gletcheren tyder på eksistensen af et forhistorisk blandet borealt økosystem med ÅMT omkring –10°C og sommer temperaturer der har begunstiget træ vækst før klimaet afkøledes og der kom isdække. Resultaterne sammenholdes med tidligere studier af fossile fund i de Transantarktiske Bjerge som beskriver en debatteret klimaperiode hvor den Vestantarktiske iskappe ikke eksisterede for ca. 3 millioner år siden.

# Chapter I

TITLE:

Ice Core Biology and its Implications for Ecology, Evolution and Climate

Contents

ABSTRACT

Within the last two decades a research field linking ecology, evolution and glaciology has emerged. This developing field of 'Ice Core Biology' has already challenged and expanded our understanding of ancient and present ecosystems in glaciated areas on Earth. Ice core biology is relevant for our understanding of climate fluctuations in the past and can add new constraints on climate and biodiversity models for the future climate scenarios. A microbial or genetic fingerprint of an ice sample specifies details about the environment it is representing, whether of local or distant origin. Such genetic fingerprints from higher taxonomic sources of local origin are linked to and test the robustness of niche and distribution models of vegetation for the areas prior to ice coverage.

When biology and glaciology are linked, they offer new understanding of physical properties in the ice affecting interpretations of climate signals in ice cores as biology might influence the chemical properties of these.

The broad applicability of results within this new field span areas within biology, climatology and astrobiology, which calls for closer collaboration between the involved supporting fields and for clear standards to avoid the risk of false positive results. We review the previous work done on biology in ice cores, highlighting the need to examine whether a habitat could sustain life for hundreds of thousands of years. We also review the development in sample and laboratory standards and discuss how possible biases are important to take into account. With the new era of high throughput sequencing technologies, a much more detailed level of research is possible on glaciological material. Therefore, we present new guidelines that should be considered from sample handling to deep sequencing and for data interpretation.

# INTRODUCTION

Ice caps and glaciers contain valuable biological information that is relevant for a wide range of biological disciplines. For example, on the one hand environmental DNA recovered from such sources can add to the understanding of past and future climate scenarios where no other biological proxies are available (Willerslev, et al., 2007, Miteva, 2010, Segawa, et al., 2010). Alternatively, such information can be used in the disciplines of astrobiology and palaeoecology - polar and glacial environments can be used as models for how extraterrestrial life forms could be studied, and how past life leaves molecular level fossils behind (Abyzov, et al., 1998, Abyzov, et al., 1999, Gilichinsky, et al., 2003, Price, 2004, Christner, et al., 2005, Gilichinsky, et al., 2007). Furthermore, the documentation of putatively diverse microbial ecosystems in glacial ice that has been recovered from remote areas such as Antarctica and Greenland (Priscu, et al., 1999, Christner, 2002, Price, 2004, Castello, et al., 2005), provide unique insights into microbial ecosystems and biological scenarios that arise in environments that are under extreme weather conditions, cold temperatures and low levels of nutrients.

The majority of biological studies that have been done on glaciers or ice sheets have focused on microorganisms. Such studies have been used to investigate themes of research within evolutionary biology, origin of life, microbiology and possible survival strategies in limiting environments (Christner, 2002, Gilichinsky, et al., 2003, Bulat, et al., 2004, Fell, et al., 2006, Lavire, et al., 2006, Price, 2007, Rohde & Price, 2007, Simon, et al., 2009). Further, microbial diversity in ice cores is suggested to correlate with past climatic signals and anomalies of these, and bacteria have therefore been used as an additional climate indicator in ice core records (Yoshimura, et al., 2000, Tung, et al., 2005, Miteva, et al., 2007, Rohde, et al., 2008, Miteva, et al., 2009, Segawa, et al., 2010).In addition, a small, yet growing, number of studies has begun to investigate questions related to palaeo-environments, through the sequencing of DNA preserved at the base of these ice-covered areas.

Regardless of the underlying questions to be targeted, similar molecular and microbiological approaches are used when investigating the polar ice caps for past or present life (Willerslev, et al., 1999, Price, 2000, Christner, et al., 2003, Miteva, et al., 2004, Priscu & Christner, 2004, Willerslev, et al., 2004, Tung, et al., 2005, D'Elia, et al., 2008, Miteva, et al., 2009, Veerapaneni, 2009). As discussed in detail later, these methods are extremely susceptible to contamination.

Studies on microbial life in old ice started in the 1970 and -80's (Lyakh & Abyzov, 1974, Abyzov, et al., 1982) and have advanced greatly in detection methods and sampling

protocols (Bulat, et al., 2004, Rogers, et al., 2004, Willerslev, et al., 2004, Christner, et al., 2005, D'Elia, 2008). Although the diversity of results is increasing fast it should be discussed how such communities can survive in ice and to what degree they are living or dead. In addition, it is important to interpret the effect of a possible microbiota within the ice due to the biases these might infer in relation to climate signals, and is another critical reason for evaluating the field.

Both communities are therefore troubled by the risk of false positive results and no strict common lab procedures or independent replication have so far been implemented (Willerslev, et al., 2004). This could greatly enhance the strength of the results and provide for a broader collaborative interpretation.

In this review we discuss how biology in ice is studied, and evaluate the molecular and microbiological methods and data used to describe both microbial diversity and past terrestrial ecosystems. Moreover, we discuss whether a new theory is needed to explain the microbial habitat and its available nutrients, whether diversity can be linked to age of the ice and past climates, and how this link should be studied with high standards of precautions for false positive results. Lastly, we present some of the results of studies investigating palaeoecosystems, and discuss how such findings can be interpreted in an evolutionary, ecological and climatic perspective

**Expanding Applications of Ice**

Types of development.

The study of microbes in polar ice, some of which are believed to be many thousands of years old, using DNA tools has been undertaken for over twenty years and remains a controversial research field (Karl, et al., 1999, Priscu, et al., 1999, Ma, et al., 2000, Abyzov, et al., 2004, Castello, et al., 2005, Bidle, et al., 2007, Lewis, et al., 2008, Bulat, et al., 2009). The results from such studies raise many questions regarding survival and preservation of biological material in ice, which have implications on a broad scale and calls for a deeper understanding of the nature behind these outputs. Later studies have approached such questions by recovering DNA from different organisms in order to describe the biodiversity from past and present ecosystems in the ice (Bulat, et al., 2004, Lavire, et al., 2006, Willerslev, et al., 2007, Segawa, et al., 2010)

The development in research around glacial microbes has progressed through a number of phases. The first phase of studies simply reported the presence of high microbial diversity in ice, without discussing how these life forms could be sustained in such

environments (Abyzov, et al., 1982, Abyzov, et al., 1993, Castello, et al., 1999). Within this type of diversity studies, little focus was given to precautions that should be taken during sampling and other means to avoid laboratory-based contamination. The first studies on diversity were purely based on culturing, which only detects a very small diversity with as little as ~1% being culturable (Catranis, 1991, Karl, et al., 1999, Priscu, et al., 1999, Christner, et al., 2000). Later, it became possible to detect unculturable taxa with the use of genetic methods. This characterized the second phase of studies, in which genetic approaches were used to screen the samples for microbial diversity, either on its own or in combination with culturing (Willerslev, et al., 1999, Ma, et al., 2000, Willerslev, 2004a, Lavire, et al., 2006). With the advent of high throughput sequencing platforms as e.g. the Roche GS-FLX series, Illumina Genome Analyzer series and Illumina HiSeq (Shendure & Ji, 2008, Metzker, 2009), the power of such techniques has expanded, as the deep sequencing enabled by these platforms allows the recovery of extremely detailed information about the biodiversity present in the ice samples (Simon, et al., 2009). A benefit of such approaches is that they can be used to describe the total diversity of DNA in the ice; whether derived from microbes or higher organisms such as plants and animals – although it should be remembered that this diversity may represent both the endogenous DNA content and contaminants. This stresses the importance of following stringent and common decontamination and replication standards to interpret and study the new high levels of sequence data that become available. The new information level will greatly enhance understanding of preservation of biological diversity in the ice and its correlation to past climates and ecosystems in the polar and glaciated areas on Earth.

The third phase of microbial research has focused more on providing explanations for how microbial life forms can be sustained in the ice. Different habitats were suggested as candidates sites for survival of these microbial organisms: liquid veins at triple junctions between ice grain boundaries; in liquid around dust minerals in the ice; and within ice crystals (Price, 2000, Miteva & Brenchley, 2005, Tung, et al., 2005, Tung, et al., 2006, Price, 2007, Rohde & Price, 2007, Miteva, et al., 2009).

Descriptions on these glacial microbial habitats presented a new paradigm for the putative live biota in ice, and even provided a perspective on extra-terrestrial life forms and where these could be expected to occur (Christner, et al., 2003, Price, 2004, Priscu, et al., 2006, Poole & Willerslev, 2007).

In connection to this, a fourth phase of research has been emerged, that focuses on the concern of contamination, which promoted studies on decontamination procedures (Bulat, et

al., 2004, Willerslev, 2004a, Christner, et al., 2005, Veerapaneni, et al., 2007, D'Elia, 2008). There is a negative correlation between the level of precautions and decontamination taken and the number of organisms identified. Despite these conclusions, no common sampling protocol or lab procedure was established.

The fifth phase of studies has been the introduction of genetic techniques that were used to investigate how much, and what kind of, DNA damage occured in these ancient organisms (Bidle, et al., 2007, Johnson, et al., 2007). When relating DNA damage to age of organisms in the ice, it was shown that organisms that can repair their DNA survived longer in the ice in comparison to spore-forming bacteria, that were seen to undergo damage build up in their DNA through time. Studies using high standards for decontamination and replication of the genetic information were able to set an age limit of around half a million years for the detection of microbial life in frozen permafrost and up to one million years ancient DNA (Johnson, et al., 2007).

The sixth phase is investigating the genetic diversity within the ice with respect to age. Genetic diversity at the sequence level among different taxonomic phyla is then proposed to correlate both to age but also to past climate scenarios where biodiversity could act as a biological climate indicator (Hansen, et al., 2001, Willerslev, et al., 2004, Hebsgaard, et al., 2005, Bidle, et al., 2007, Johnson, et al., 2007, Willerslev, et al., 2007, Miteva, 2010, Segawa, et al., 2010)

The amount of literature available on microbial studies in ice is large, in comparison to those that used DNA for environmental reconstruction as this is a much more recent topic and thus remains relatively unexplored. In the following sections of this review, we will look at these two categories in turn. First, by focusing on those studies that relate predominantly to microbial analyses, and then secondly discussing those that relate to paleo-environmental and taxonomic reconstruction. Finally we present a section that discusses in detail a number of features that are common to both areas of research.

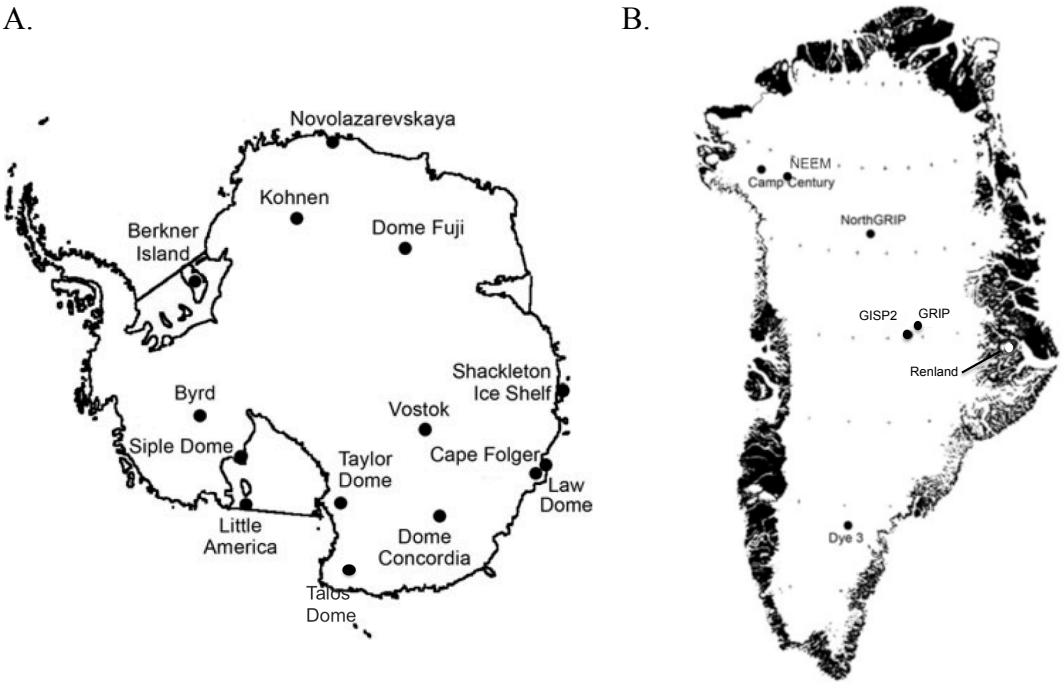
OBSERVATIONS OF MICROBIAL ORGANISMS IN ICE

**Biodiversity**

The diversity of DNA that has been observed within ice is high and previous results have included bacteria, archea, fungi, viruses, plants and insects (Castello, et al., 1999, Price, 2004, Tung, et al., 2005, Bidle, et al., 2007, Veerapaneni, et al., 2007, Willerslev, et al., 2007, D'Elia, et al., 2008, Rohde, et al., 2008, Miteva, et al., 2009, Price, 2009). The ice has different characteristics depending on how rich it is in impurities. Transition ice, also called basal ice, is found close to bedrock at the interface between either glacial ice and debris or glacial ice and water, that is refrozen and holds higher levels of impurities than clean glacial ice (Knight, 1997, Souchez, 2000, Samyn, 2005a). Some report that basal ice, where the ice is either close to bedrock or consists of accretion ice above a sub-glacial lake like Lake Vostok, contains higher levels of biodiversity than the clean ice (Price, 2000, Abyzov, et al., 2004, Willerslev, et al., 2007, D'Elia, et al., 2008, Miteva, et al., 2009). Others report a lower microbial diversity in the basal layers due to an older age of these bacteria if they correlate with the age of the basal ice and a signal of damaged DNA profiles (Christner, 2002, Bidle, et al., 2007, D'Elia, et al., 2008, Veerapaneni, 2009).

A combination of culturing, staining, scanning and genetic methods was used to detect the diversity of microbes reported to occur at various depths of the glacial ice in ice cores from both Greenland and Antarctica (Abyzov, et al., 1998, Abyzov, et al., 1999, Ma, et al., 2000, Christner, et al., 2003, Abyzov, et al., 2004, Rohde & Price, 2007, D'Elia, et al., 2008, Rohde, et al., 2008). However, most studies were not able to show any taxonomic results until after culturing, which is a critical step where contamination could be introduced and create false positive results. Also, very few studies have been replicated in an independent lab, which could increase the risk of reporting contaminants.

**Figure 1**. Maps of deep drilling sites in Antarctica (A) and Greenland (B). Modified from Bentley & Koci (2007).




**Contamination**

Development of common guidelines that deal with decontamination of the ice samples to a standardised level before a study reaches a level suitable for publishing has not yet been accomplished. Several proposals were published on decontamination protocols but most of these contained difficult steps that might leave possible false positive signals, as discussed below. Even if proper decontamination was applied to the samples, contamination could be introduced during downstream lab work and it is therefore of great importance to have strict clean lab facilities and high standards for verifying the results (Bulat, et al., 2004, Willerslev, 2004a).

Examples where thorough decontamination was applied to ice samples from various cold glaciers showed very low success in DNA amplification and low cell counts in clean ice from polar ice caps, especially for the old samples (>200 years) (Willerslev, et al., 1999, Christner, 2002, Willerslev, et al., 2004). This suggests that proper decontamination greatly reduces the amount of bacteria detected prior to culturing and amplification.

Contrasting results were also observed in studies on identical ice samples when different decontamination and clean laboratory protocols were applied. This was observed

with ancient samples from the accretion ice above Lake Vostok in Antarctica. When strict decontamination and independent replication criteria were applied, viable cells were not obtained from ancient glacial ice from Vostok accretion ice (Lavire, et al., 2006). Moreover, this study only retrieved short DNA fragments from ancient microbes in 20.000 – 60.000 years old ice samples as opposed to the diverse results previously obtained from the same site (Priscu, et al., 1999, Christner, et al., 2001), which could not be reproduced in later findings (Bulat, et al., 2004, Lavire, et al., 2006). Another example from Greenland where contamination may have caused false positive results is the GISP2 core. In this study the authors were unable to successfully PCR amplify microbial DNA from the sample, even when high microbial cell numbers were reported, until whole genomic amplification or culturing was performed (Miteva, et al., 2009). This last step is however sensitive to modern contaminants. Another microbial and viral study on the GISP2 and Dye3 ice cores reported the detection of various pathogenic bacteria and tomato viruses from both glacial and basal ice (Castello, et al., 2005). As argued following phylogenetic reanalysis of the data, both these results and a similar study where virus strains were amplified from Siberian ice (Zhang, et al., 2007) were likely to be false positive results arising during lab work (Worobey, 2008). It is worth noting that the analyses in the first of these studies were performed in laboratories that worked with tomato viruses, fungi and bacteria from environmental samples (Castello, et al., 1993, Castello, et al., 1995). Given that the stability and preservation of RNA molecules is lower than that of DNA, and that the 1kb long RNA fragments reportedly amplified largely exceed the 160bp of DNA amplified from the Dye 3 ice core in later studies (Willerslev, et al., 2007), the level of viable organisms containing preserved DNA in cold based samples from deep clear and silty ice is put into question.

The volume of ice studied for microbes or DNA is another interesting parameter to compare. As the biological material seemed to be of very low concentration it was surprising that low volume samples could yield positive results when high volume samples from comparable sites only returned negative results (Miteva, et al., 2004, Willerslev, et al., 2007, Miteva, et al., 2009). Hence the volume is relevant to consider and could indicate a bias if high diversity was found in small volume studies. Two opposing outcomes were presented from the GRIP and GISP2 sites, a surprising result given that both cores were very similar in chemistry and are in close geographical proximity. While GISP2 presented a higly diverse and abundant microbial community, no biological remains were detected from the GRIP ice core. Further, for the GRIP core high volumes of ice were studied and both DNA and amino acid racemization results indicated poor preservation (Willerslev, et al., 2007). However, previous

studies on the GRIP and Dye3 cores showed different results (Ma, et al., 2000, Castello, et al., 2005) and again suggest that ice volume, strict decontamination, clean down stream lab protocols and independent replication are critical factors affecting the results. Bulat et al. (2004) investigated the false positive result dogma downstream of decontamination and within strict clean lab facilities and demonstrated that even under these strict conditions false positive results occurred. This was done as an attempt to reproduce previous results indicating high biodiversity from the lake Vostok core (Karl, et al., 1999, Priscu, et al., 1999, Abyzov, et al., 2001, Christner, et al., 2001), but only one apparently authentic phylotype was found, indicating many putative contaminant groups in the original results. Thus Bulat et al. (2004) suggested making a contaminant database to which possible new findings should be compared, before believing in new diverse results. Interestingly, many of these previously published results of various bacterial groups fall into this contaminant database, including groups within Firmicutes and Actinobacteria. Bulat et al. (2004) removed 37 – 61% of the samples in decontamination and ascribed the false positive sequence results detected as arising post-decontamination from dust and human sources in the labs. This illustrated the delicacy of such samples where the very low biodiversity existing is easily outcompeted by contaminants entering the samples during extraction, amplification or culturing steps.

## POTENTIAL MICROBIAL ENVIRONMENT IN THE ICE

### Incorporation and survival in the ice

The explanations for how the reported microbes should be able to survive and be revived in culturing experiments or sustain their genetic profile with long amplifiable products were very diverse. Some argued that bacteria found in the glacial ice of cores from GISP2, GRIP, Dye 3 and above Lake Vostok resided in liquid veins in the polar glacial ice where nutrients were believed to be available in solution (Price, 2000, Miteva & Brenchley, 2005, Tung, et al., 2005, Miteva, et al., 2007). An additional suggestion was that they survived within the crystals by mere diffusion of nutrients to their position in the ice, at a level where they repaired their DNA, defined as survival or maintenance metabolism, but did not metabolize at a level where they multiplied (Price, 2007, Rohde & Price, 2007, Rohde, et al., 2008, Price, 2009).

Microbes were believed to become incorporated into the glaciers or ice caps by co-transport on terrestrial dust, wind blown material, volcanic ashes and by marine surface aerosols. The level of microbiota in the ice was suggested by some to correlate with dust load,

so that in colder periods, a higher dust content and higher number of microbes was expected to be incorporated in the ice (Abyzov, et al., 1998, Christner, et al., 2000, Priscu, et al., 2006, Miteva, et al., 2009).

The livelihood for immured microbes in the ice depends on the availability of nutrients and water in the ice. Diffusion was argued to account for the supply of nutrients, carbon, energy and water, all needed to maintain life in the ice. This was discussed in scenarios for both liquid veins and pure diffusion through the ice lattice (Price, 2000, Rohde & Price, 2007, Price, 2009). Not only in ice have ancient bacteria been proposed to survive for hundreds of thousands to millions of years. Ancient halophilic bacteria have been grown and isolated from Cretaceous and Permian salt crystals (Vreeland, et al., 2006, Vreeland, 2007). However, similar results have previously been debated due to limited sequence divergence between ancient and modern representatives (Graur & Pupko, 2001, Nickle, et al., 2002). Another critical issue, which was reported in neither the salt crystal nor the ice studies, was the lack of detected sequence damage or long recovery times for repairing fragmented DNA before growth. This should be expected to be important to regain growth after long periods of time where mutations or damage could have accumulated for ancient organisms many hundred thousands or million years old. The parallels between the studies of bacteria and archaea in salt crystals and the studies on microbes in ice were that the environments for survival are extreme, and the studied bacteria occurred in very low numbers, which made the studies very contamination prone. The majority of these studies have also neither been independently replicated nor addressed the question as to where the nutrient supply was, that would allow repair thus halting the accumulation of damage.

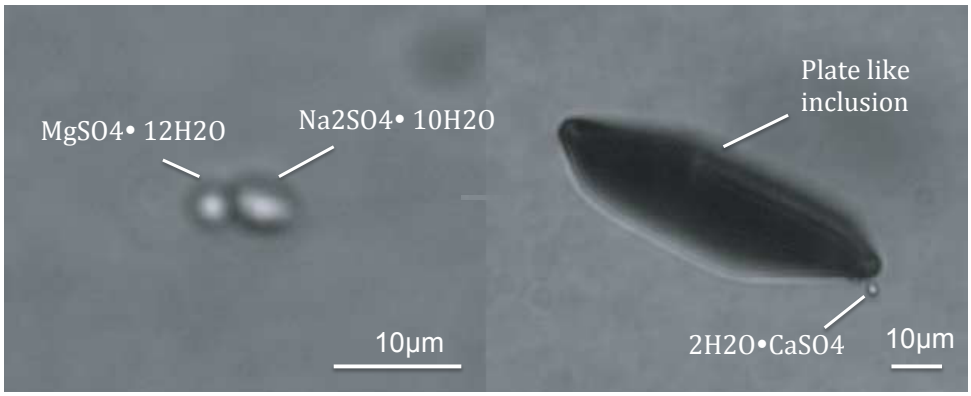
**Diffusion of nutrients**

The diffusion rate of nutrients is a critical factor with regards to the nutrient availability necessary to sustain microbial communities in ice. The liquid flow through veins is generally believed to be very slow, but the diffusion velocities of different impurities inside the liquid will depend on the individual impurities and temperature, and could be rapid (Dash, et al., 2006).

Rohde & Price (2007) discussed diffusion velocities for different small molecules at different temperatures, in order to describe how microbes within ice crystals in the cold solid ice would theoretically be able to sustain life by diffusion of nutrients. They reported diffusion velocities for temperatures at −10°C and −32°C, representing different depths in ice

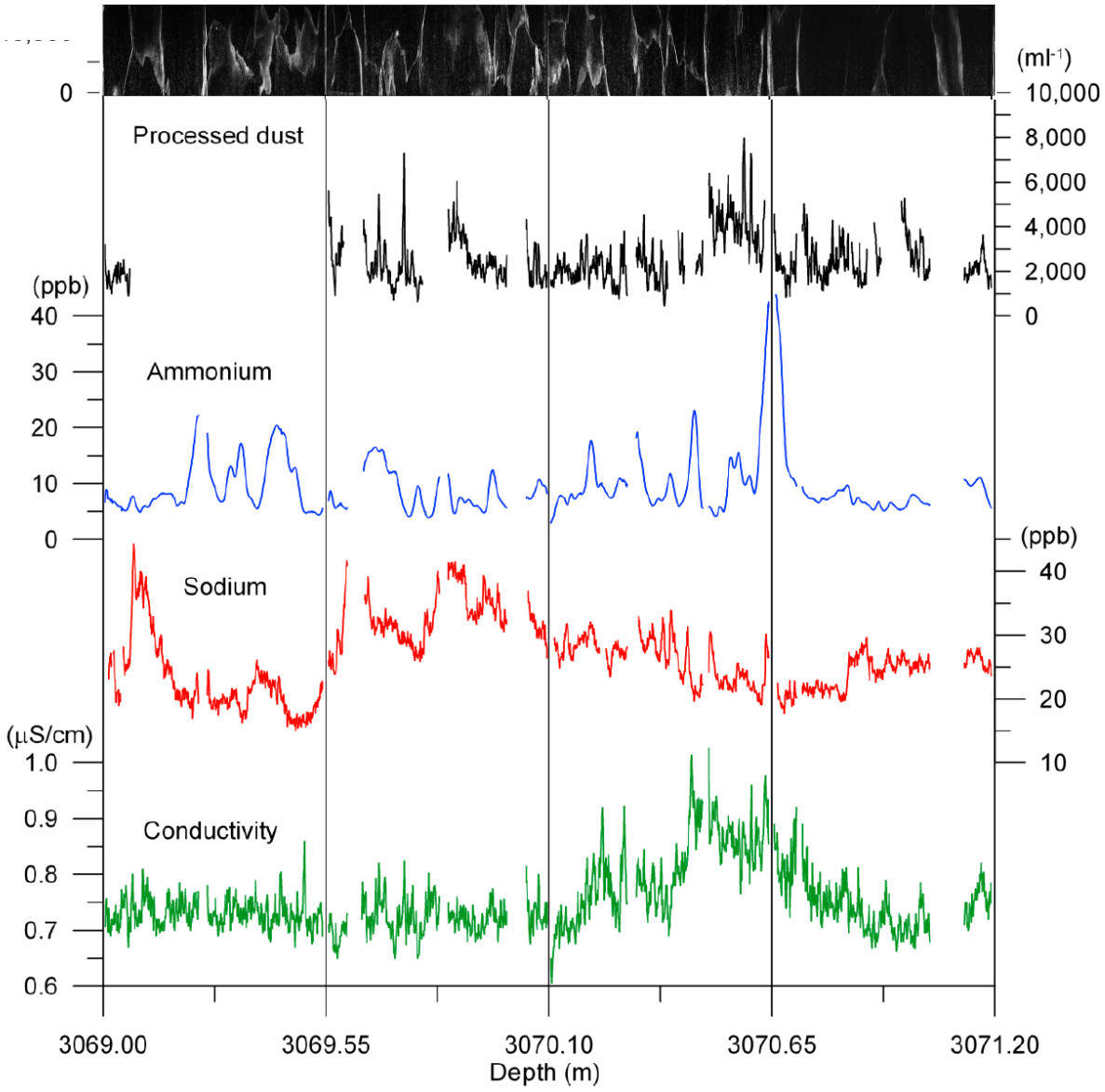
caps. However the velocities calculated apply to small molecules and many of the large microbes detected use larger molecules in their metabolism. It is therefore questioned whether they can switch to small molecules for maintenance of their DNA without active metabolism (Rohde & Price, 2007). However, even the diffusion rate of these small molecules depends on the chemical form they have in the ice. This varies not only in relation to temperature but also to what post-depositional reactions occur within the ice between different chemical compounds (Iizuka, et al., 2008). Such reactions depend on factors like the amount of terrestrial and marine inputs in the ice caps, and to what depth the reactions occur (Ohno, et al., 2005). This is because reactions in firn will be different from those deeper in the solid ice, where the compounds exist within clathrate hydrates, which again could be different from the scenario in deep ice close to bedrock where temperatures are generally warmer (Johnsen, et al., 2001, Iizuka, et al., 2008, Lambert, et al., 2008, Faria, et al., 2010, Ohno, et al., 2010). The diffusion rates that Rohde and Price (2007) reported for small compounds like the hydrochloric and nitric acids (HCl and $HNO_3$), which would act as nutrients for microbes, are challenged by the results of Ohno et al. (2008) who found that most of the $Cl^-$ and $NO_3^-$ occur as salts like NaCl, $Mg(NO_3)_2$ and $Ca(NO_3)_2$ that are much less diffusible than acids. These salts were found as micro-inclusions within the ice crystals or at grain boundaries, e.g. in the form of the clathrate hydrate ($CaSO_4 * 2H_2O$) and nitrate salts (Iizuka, et al., 2008), see Figure 2.

**Figure 2**. Micro-inclusions of different salts positioned within clathrate hydrates in warm Dome Fuji glacial ice. Raman spectra of two different ice crystals with three different clathrate hydrate salts. From Ohno et al. (2005).




Ice deposited during a warm time period has a lower dust and carbonate content and is more acidic. Acids have been observed built into the crystal lattice as clathrates in cold ice where these can be observed in ice cores from Greenland and Antarctica (Svensson, et al., 2005). In the study by Faria et al. (2008) on the Antarctic EPICA-DML and EPICA-Dome C cores, the authors reported that the micro-inclusions were found within the crystals and not along grain boundaries until they reached warm ice ( > −7°C). Nevertheless in Dome C they did not find liquid veins but saw a change in micro-inclusions of salt particles occurring at grain boundaries in this deep ice. This was in agreement with a study that also found higher sulfate peaks at grain boundaries in the deepest ice from the Dome C core (>2800 m), but not in the above lying colder ice (<−10°C) (Traversi, et al., 2009). Therefore, it is difficult to generalize about the nutrient availability in the ice in terms of compound diversity, concentration and location. This makes it difficult to verify the basis for microbes encapsulated in the ice and their long-term survival from a pure theoretical point of view. The microbes might be present in the ice but not viable when they have exhausted the nutrients available in their proximity. How fast they exhaust the nutrients depends on the temperature at the relevant depth as the chemical reactions driving the diffusion of different ions correlate to temperature.

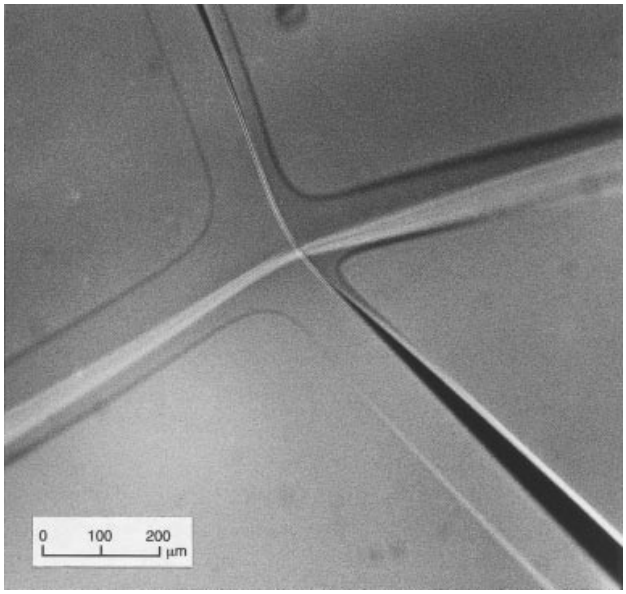
**Figure 3**. A visual stratigraphy picture of deep warm ice close to melting point from the NGRIP core at depth 3069-3071m with measured records of dust, ammonium, sodium, and conductivity. The figure illustrates dark areas of large size ice crystals with white crystal boundaries. Crystal sizes are of the size of around 10cm in diameter. The measures of ion and dust content along the core provide evidence that these compounds have not all migrated to the crystal boundaries. Indeed, the annual layering is conserved in dust across ice crystals that are tens of centimeters across, and there are no elevated impurity concentrations within the ice crystal boundaries. Not even the strongly diffused ammonium appears to have concentrated in grain boundaries, which challenges the ideas of an active liquid vein system in the deep central Greenland ice. If liquid veins were present the conductivity signal would not be detectable nor would the diffusible ions within crystals. From Svensson et al., (2011).



**Habitats**

Glacial ice. The first habitat within ice that was suggested as an environment that could supply microbes with life-sustaining requirements, was liquid veins (Figure 4) (Price, 2000). The theory about liquid veins at triple junctions at ice grain boundaries was first introduced as a condition in a temperate glacier (Nye & Frank, 1973). An interconnected network of such veins in cold polar glacier ice has so far not been detected empirically and their existence and the implications they would have on the interpretations of climate proxies have been greatly debated (Mulvaney, et al., 1988, Rempel, et al., 2001, Barnes, et al., 2003, Ohno, et al., 2005, Traversi, et al., 2009, Faria, et al., 2010). The empirical studies where liquid veins have been detected were from temperate glaciers, sea ice, coastal glacier ice (with high $H_2SO_4$, $HNO_3$ and HCl content) or in the deep warmer ice sections of polar ice cores. All these hold different characteristics than continental polar glacier ice from where the microbial studies have been conducted and sampled (GRIP, GISP2, DYE 2, DYE 3, Lake Vostok, Taylor Dome, Siple Dome, Dyer plateau – except for the Yamato and Mizuho cores that are coastal (Willerslev, et al., 1999, Ma, et al., 2000, Christner, 2002, Bulat, et al., 2004, Miteva, et al., 2004, Lavire, et al., 2006, Willerslev, et al., 2007, D'Elia, et al., 2008, D'Elia, et al., 2009, Miteva, et al., 2009, Segawa, et al., 2010). The size of these veins are believed to vary from less than 1μm up to 100μm in cross section, depending on temperature, solutes and pressure in the ice (Mulvaney, et al., 1988, Nye, 1989, Dash, et al., 1995, Fukazawa, et al., 1998). The physical properties of these veins are believed to conform to different liquid levels, varying from aqueous acids to thin gel structures, continuously changing according to re-crystallization dynamics (Dash, et al., 1995, Dash, et al., 2006). These veins have been studied by different means, e.g. directly by sensitive spectrometry methods or indirectly through chemical measurements and grain structure studies (Souchez, et al., 1978, Mulvaney, et al., 1988, Fukazawa, et al., 1998, Souchez, et al., 2004, Ohno, et al., 2005, Iizuka, et al., 2008, Faria, et al., 2010, Samyn, et al., 2010, Svensson, 2011).

**Figure 4**. Illustration of the hydraulic system of liquid veins in acidic ice close to pressure melting point intersecting at grain boundary between four ice crystals. From Mulvaney et al. (1988).




Studies from Antarctic coastal ice showed that liquid veins contained liquid acids as hydrochloric acid, sulfuric acid and nitrous acids that have eutectic temperatures down to −88ºC, which would keep the liquid unfrozen even within solid pure ice (Mulvaney, et al., 1988, Fukazawa, et al., 1998). The acids should be confined to the liquid veins and be insoluble in ice crystals or at least have much higher affinity for the veins than the crystals or planar boundaries (Price, 2000). However, this was later questioned by other results (Ohno, et al., 2005, Iizuka, et al., 2008, Faria, et al., 2010) where the authors measured ionic compounds by ion chromatography, Raman spectrometry or microstructure mapping and found acid clathrates and salts within ice crystals and no liquid veins. Moreover, the study by Iizuka et al. (2008) was on an ice core drilled from the central part of Antarctica, whereas the studies by Fukazawa et al. (1998) were from Antarctic cores close to the coast, which contain a much higher content of HCl and thereby resemble other studies where brine pockets have been detected at grain boundaries (Wolff, et al., 1988, Dash, et al., 1995, Junge, et al., 2001). Iizuka and colleagues also reported, using ion balance measurements on the chemical compounds found in Greenland and Antarctic ice cores, that most of the $Cl^-$, $NO3^-$ and $SO4^{2-}$ were found as salts rather than in acids. Ohno et al. (2005) further found that the soluble impurities in the ice were mainly on the form of sulfate salts and were formed in the upper layers of the ice and preserved their composition as they were build into the ice cap, even at deep positions. Therefore, the depth profile and climate proxies should be preserved

when interpreting the isotope signals. Ohno et al. (2005) and Faria et al. (2008) further showed that the impurities did not always occur at grain boundaries or triple junctions but were found within the crystals as inclusions. This was also shown in studies of polycrystalline ice from the Byrd core, Antarctica and from the GISP2 core from Greenland where impurities occurred at grain boundaries and as inclusions (Barnes & Wolff, 2004). In a recent study by Svensson et al. (2011) the annual signals in dust and even annual signal of sodium in the Eemian ice from NGRIP was preserved in the deep ice, and the impurities were not concentrated at grain boundaries. This is known because the crystals are up to 10cm across and still contains 1cm thick annual layers within the grains. Furthermore, even the small $NO_3$ ions were clearly diffused significantly in this ice and were not concentrated in the grain boundaries (Svensson, 2011). If liquid veins existed in the cold ice they would cause high diffusion of chemical compounds, which could therefore not show the regular fluctuations. Liquid veins in polar glaciers have not been detected empirically *in situ*, nor has the significant amount of water these veins would hold if they existed widespread through the ice cap in response to the high pressure and the low freezing point of eutectic acids (H.B. Clausen, J.P. Steffensen, A. Svensson, S. Kipfstuhl, pers comm. 2011). Thus, the evidence for liquid veins in polar glacier ice is less well supported as an interconnected network distributed through the ice cap than evident from previous papers on microbes in polar ice (Price, 2000, Christner, 2002, Miteva & Brenchley, 2005, Tung, et al., 2005, Tung, et al., 2006, Rohde & Price, 2007, Rohde, et al., 2008, Miteva, et al., 2009, Price, 2009, Simon, et al., 2009, Veerapaneni, 2009, Christner, 2010).

Basal ice. Basal ice from ice cores or glacier samples with high contents of debris, contain much higher levels of minerals than the clean glacial ice (Samyn, et al., 2005b). Furthermore, if the temperature of basal ice below an icecap is warm due to geothermal heat flux from bedrock, then water might be present. Bottom melting has been observed in several cases, such as NGRIP in Greenland and Siple Dome and Byrd station in Antarctica (Dahl-Jensen, 2003, Vogel, 2003, Andersen, et al., 2004). It is also a common phenomenon in temperate glaciers that water is present when pressure melting and regelation occurs, as the ice moves around obstacles of large or minor sizes (Samyn, et al., 2005b). The dynamic existence of liquid that can exist between ice crystals or around mineral inclusions in basal ice depends on salinity, pressure and temperature. Water in ice at the base of a glacier can exist as water channels that can drain the melt water and re-freeze to basal ice at the ice-rock interface (Jouzel & Souchez, 1982, Souchez, et al., 2004). In such locations basal samples with a

combination of water, minerals, nutrients and higher temperatures provide a more attractive environment for microbes than the clean cold ice.

However, in cold polar glaciers where geothermal heating is absent, temperatures do not rise above −10°C, liquid veins have not been found empirically in basal or debris rich ice from polar glaciers or ice core samples. Traces of local melting and re-freezing dynamics of thin liquid films as ice reforms and incorporates debris are however discussed for Antarctic basal glacier ice (Samyn, et al., 2010). Here the thin liquid films are not a constant factor present in the ice but caused transient due to strain movement, re-crystallization around debris clasts and pressure. Overall, the in situ melting is strongly limited and the polar glaciers are generally considered dry (Knight, 1997, Souchez, 2000, Samyn, 2005a, Samyn, et al., 2010). The liquid availability for life within basal cold ice is thus extremely low and the presence of living ancient microbes very limited. On the other hand the preservation and abundance of their remains should be much higher in the basal ice due to a previous life supporting habitat prior to ice coverage, and due to a high binding affinity between DNA and soil particles (Lorenz & Wackernagel, 1994).

Permafrost. In permafrost the existence of liquid depends mainly on temperature and salinity, as opposed to the scenario in glaciers or ice caps where pressure plays an important role. In permafrost locations where seasonal temperature fluctuations raise the surface temperatures above freezing, water and melting is common. Permafrost is able to contain 2-7% liquid water within thin films (Gilichinsky, et al., 2007). Liquid films within permafrost have been observed to surround soil particles and have been discussed in relation to microbial habitats (Gilichinsky, et al., 2003, Steven, et al., 2006). In particular liquid films and cryopegs can be detected in areas where permafrost has high salinity levels. The size of liquid films decreases with temperature to a few nanometers thickness until −10°C, below which standard analytical instruments cannot detect them. Water is therefore considered bound in a biological "dry environment" (Gilichinsky, et al., 2007). It has nevertheless been demonstrated theoretically that nanometer thick liquid films could surround mineral grains or possibly microbes in permafrost soils at temperatures down to −30°C (Mohlmann, 2009). However, even in a frozen cold environment with access to nutrients and with limited available liquid, microbial life seems to show a decreasing trend with age. This has been observed in permafrost samples of varying age where a decrease in number of cultivable cells and microbial diversity occurs with increasing age of the permafrost (Willerslev, et al., 2004b, Johnson, et al., 2007).

This supports a theory of an upper age limit for how long microorganisms can survive in frozen environments, even with access to nutrients but in a water limited system (Steven, et al., 2006). Nevertheless, Bidle et al. (2007) presented the possibility of 5-8 million year old living bacteria from an Antarctic glacier ice in Beacon Valley below a debris layer. This is a stratigraphic age based on $^{40}Ar/^{39}Ar$ dating of an ash layer just above the ice material, and it has been questioned whether this is representative of the ice beneath it (Gilichinsky, et al., 2007). The age has also been dated with a combination of measurements of cosmogenically deposited $Be^{10}$ and embedded quartz grains within the ice, and modelling of the measured cosmogenic $^{3}He$ in the ice, which gives an age of maximum 300.000 years (Gilichinsky, et al., 2007). Still, the consensus shows that the oldest viable bacteria have been found in Antarctica and the age remains to be settled upon. In the Arctic the temporal limit for bacteria that can repair their DNA and metabolize at a very limited rate has been argued to reach as far back as 400.000 to 600.000 years (Johnson, et al., 2007). Plant, microbial and animal DNA have been amplified and independently replicated from permafrost and basal ice and dates back to 400,000 – 800,000 years (Willerslev, 2004a, Haile, et al., 2007, Johnson, et al., 2007, Willerslev, et al., 2007). Altogether, the preservation of DNA from dead organisms is supposed to be high in these cold environments that provide stable cold temperatures and limited amounts of water (Willerslev, 2003).

## GAS ANOMALIES IN GLACIAL ICE AND BACTERIA

If microbes stay alive in the ice and metabolize, they should produce excess gasses, which could give rise to gas anomalies in the climatic profiles. This scenario has been discussed in several studies for both clean and silty ice (Skidmore, et al., 2000, Price & Sowers, 2004, Tung, et al., 2005, Priscu, et al., 2006, Miteva, et al., 2007, Rohde, et al., 2008, Miteva, et al., 2009). Elevated levels of gasses in silty or basal ice has been observed in several cases in ice samples from Dye 3, GRIP, GISP2, JEG, Lake Vostok and Taylor glacier (Gow & Meese, 1996, Souchez, et al., 1998, Skidmore, et al., 2000, Sowers, 2001, Samyn, et al., 2005b, Bender, et al., 2010)

It has been suggested that the elevated levels of gasses in GRIP/GISP2 comes from local degradation of organic sources from a marsh or peat like environment that when frozen got incorporated to an advancing ice sheet and caused the elevated levels of $CO_2$ and $CH_4$ (Gow & Meese, 1996, Tison, et al., 1998, Tung, et al., 2006). A similar scenario for the evolution of the gas properties in the GISP2 site was presented in a study that reported a great

difference in the gas levels between the silty basal ice and the above lying clean ice, with the former containing gas levels indicative of a metabolizing source (Bender, et al., 2010). The elevated gas levels in the basal ice at Taylor Glacier in the Dry Valleys of Antarctica show highly increased $CO_2$ levels, at the expense of $O_2$, in layers with high debris content compared to the clean ice layers in between these. Such gas composition changes within the ice is suggested to be due to interactions between gas bubbles in the ice and liquid that is locally produced through melting-refreezing mechanisms. A biological source causing these anomalies was rejected as the gas sum of $CO_2$ and $O_2$ deviated from that of a biological signal (Samyn, et al., 2005b).

These productions could nevertheless be the frozen preserved signature of past biota metabolizing in the permafrost or peat environment before it was incorporated into the ice sheet, as suggested by Miteva et al. (2009). Extensive studies on the GRIP core provided no positive DNA results when stringent decontamination protocols were applied, even after studying large amounts of ice (Willerslev, et al., 2007). The amino acid racemization profiles for these basal samples indicated low levels of preserved biomolecules, supporting the signature of no living biota. Moreover, both the GRIP (−8.59°C) and GISP2 (−9.22°C) sites are cold at the base, and no basal melting occurs, which indicates a biologically dry environment, prior to the build up of the Greenland Ice Sheet, that could have been similar to the scenario described below the John Evans Glacier in Northern Canada.

The correlation between microbes and elevated levels of $CH_4$ and $N_2O$ gas has been studied in the GISP2 silty ice, Vostok clear ice and from the younger John Evans Glacier (JEG) in northern Canada (Skidmore, et al., 2000, Miteva, et al., 2007, Miteva, et al., 2009). All of these studies only detected the microbial diversity by culturing and did not follow strict decontamination protocols. The samples from the JEG glacier are considered modern, thus much younger than those from the GISP2 site and Vostok ice core. The temperature of this silty ice is at the pressure melting point (Skidmore, et al., 2000), and as basal melting occurs, both water and oxygen are available to support living microbial communities, which is therefore very different from the GISP2 and GRIP sites. However, the decontamination conditions reported were not among the most stringent and no replication was mentioned. The diversity reported from JEG belonged to aerobic chemoheterotrophs, anaerobic nitrate reducers, sulfate reducers and methanogens (Skidmore, et al., 2000).

The elevated levels of $N_2O$ at various depths in the Vostok core have been linked to the presence of higher dust contents and higher cell counts at equivalent depths, presumably by viable nitrifying bacteria (Abyzov, et al., 1998, Sowers, 2001, Price & Sowers, 2004).

Nitrifying bacteria require an aqueous media and as the glacial ice in Vostok is –40°C (Price & Sowers, 2004), the environment should be considered dry following above discussion on liquid veins. In addition, another study followed stringent decontamination and replication protocols on the Vostok ice, and found that the microbial phylogroup detected, did not belong to nitrifying groups (Bulat, et al., 2004). Moreover, this microbial phylotype was ascribed to be preserved in a frozen state at a grain boundary and not thriving alive in a liquid vein (Bulat, et al., 2004). Further, the cell counts reported previously for accretion ice in Vostok (Karl, et al., 1999, Priscu, et al., 1999) could not be replicated in a later study on the Vostok ice with high standards for decontamination and showed cell numbers 2-3 orders of magnitude less than reported in previous studies (Bulat, et al., 2009). Therefore, the $N_2O$ production must have another origin than from in situ production in the old ice. There is so far no clear evidence that viable bacteria are metabolizing, producing excess gasses in situ in the glacial or silty ice from the old and cold glacier ice from Antarctica or Greenland. If these reported excess gasses are due to microbial activity it should have been produced relatively shortly after deposition while the microbes were still viable and had enough nutrients and water to metabolize from. A new theory is needed to explain such a scenario.

So far there are no clear indications of correlation between gas anomalies and viable metabolizing microbes in deep cold ice cores from Antarctica and Greenland. If a correlation exists between gas excesses and microbial past activity, it remains to be investigated by means of ancient DNA techniques. Further, a clear positive result is difficult to prove as the physical mechanisms occurring, e.g. during melting will blur the signal from a microbial source. A new paradigm is needed to explain if the gas anomalies correlate with microbial phyla-groups that metabolize these gasses.

## AGE LINK TO BIODIVERSITY

A negative correlation between diversity or abundance of organisms and sample age is expected when studying ancient samples such that the diversity detected decreases the older the sample material becomes (Hebsgaard, et al., 2005, Poole & Willerslev, 2007). It is therefore interesting that different correlation patterns appear when comparing the biological diversity between studies from Greenland and Antarctica at various depths and of different ages. No correlation was observed between age and diversity from studies on Dye 3 and GISP2 cores and Lake Vostok (Ma, et al., 2000, Abyzov, et al., 2004, Miteva, et al., 2009). Contrary to this, correlations between age and diversity were found in other studies on the

Dye 3, GISP2 core and Antarctic glacier samples where higher level of decontamination criteria have been applied (Bidle, et al., 2007, Johnson, et al., 2007, Willerslev, et al., 2007, D'Elia, 2008, Veerapaneni, 2009). In all of the samples studied from the Dry Valleys in Antarctica by Bidle et al. (2007), there was neither evidence of liquid films around sediment grains in the samples when subjected to scanning electron micrography, nor was in situ active metabolism found for the microbial communities that were observed, although degraded DNA was detected. Johnson et al. (2007) also observed a clear decreasing trend between DNA fragment length, metabolism and age, and determined a maximum age for when DNA was detectable of 600.000 years in their samples. This finding presented non-spore forming bacteria as the oldest living bacteria in ice and set an age limit of when microbial life is no longer detectable in glacier ice (Johnson, et al., 2007). Depending on the dating of the samples from Bidle et al. (2007) these might have pushed the oldest age further. Thus, a new theory is needed to describe how these ancient microbes preserve their DNA in a state where it can be detected.

Bulat et al. (2009) compared the number of cells down through the Vostok site from surface to 3659m's depth by the use of flow cytometry and staining. Very low cell counts in the order of 0 – 19 cell/ml were found in the glacial ice. From these results the biodiversity of the central part of the Antarctic ice sheet was measured to be very low. Therefore the atmospheric contribution of airborne microbes as well as the input from the subglacial lake below the ice is limited in central Antarctica (Bulat, et al., 2009).

Miteva et al. (2008) and Abyzov et al. (1998) found correlations between cell counts and dust layers and Abyzov et al. (2004) found the highest number of cells in the deepest ice. It was previously shown that the number of cells found in young and old samples (up to 1 mill years) did not vary, even though the old samples did not have amplifiable DNA, so the preservation of dead cells is high in permafrost and ice (Johnson, et al., 2007). This could perhaps explain the high number of cells detected in studies where there is no amplification success on meltwater but high cell counts (Miteva & Brenchley, 2005, D'Elia, 2008). Moreover, staining of bacteria with fluorescence dyes, as the widely used Live/Dead BacLight Bacterial Viability kit (Molecular Probes, Eugene, OR) in environmental samples can be complicated, as these also bind to substrate material as well as the biological target and can therefore bias the signal detected (Biggerstaff, et al., 2006).

Another study aimed to link microbial diversity and abundance to climate in glacial coastal ice from Antarctica. The authors found a correlation between age and diversity and

diversity and climate (Segawa, et al., 2010). These results are based on samples from the coastal cores of Yamato and Mizuho sites, which date from the Holocene (2000-4000 years ago) and the Marine Isotope Stage 3 MIS3 (55.000-60.000 years ago), respectively. The highest diversity was found in the youngest sample where the diversity was believed linked to the interglacial climate at deposition and differed from the older glacial sample. The diversity reported was detected by PCR amplifications of ~460 bp fragments from 16S rRNA region, using high volumes of melted samples. The cell counts were high but did not support a viable signal as only short DNA fragments could be amplified. Unfortunately, the authors did not test the bacterial sequences for a damage pattern, which could support the older age of the diversity found and the endogenous nature of the results. Different geographic sources were believed to give rise to the variation in diversity found at the different locations (Segawa, et al., 2010). Therefore such diversity studies could be useful as environmental markers for past climates if high standard decontamination protocols are followed and lab procedures continuously updated to follow strict clean protocols.

## OBSERVATIONS ON ENVIRONMENTAL STUDIES IN ICE

Glaciated polar areas are sparse in fossils that can be used as proxies to describe past environments, climate and their ecosystems before ice coverage because their remains are buried deep below glaciers and ice sheets with limited access to sampling. Consequently, theories about past ecosystems in Antarctica and Greenland and their taxa are based on the often-limited fossil material available at the margins of the present glaciated areas (Funder, 1989, Bennike, 2002, Haywood, et al., 2002, Convey & Stevens, 2007, Francis, et al., 2007a). Uncertainties then exist about how the climate was at the time when these areas were last ice-free, and when the change in climate occurred that caused the ice sheet to build-up in both hemispheres (Huybrechts, 1993, Marchant, et al., 1993, Sugden, 1996, Salzmann, et al., 2011). Ancient genetic methods have been applied on permafrost and sediment samples to reconstruct past environments based on plant and animal molecular remains (Hofreiter, et al., 2003, Willerslev, 2003, Lydolph, et al., 2005, Hansen, et al., 2006, Haile, et al., 2007, Corinaldesi, et al., 2008, Sønstebø, et al., 2010, Anderson-Carpenter, et al., 2011). DNA amplified these sediments is believed to derive from tissues or faecal deposits from fauna (Haile, et al., 2007) and plant DNA most likely derives from fine rootlets, leaves and seeds or pollen (Hebsgaard, et al., 2009, Anderson-Carpenter, et al., 2011). DNA amplified from these sediment samples is believed to derive from degraded tissues or faecal remains from the

organisms living in the area at the time of deposition. The ancient DNA is moreover thought to be at least the age the sediments represent and leaching between sediment strata is considered insignificant and the chemical degradation rate of DNA slow due to the low temperatures present (Lindahl, 1993, Hebsgaard, et al., 2005)

Ancient environmental DNA has been found preserved in the basal debris rich ice in ice cores (Willerslev, et al., 2007). The DNA was extracted from the soil particles and represents "molecular fossils" from organisms that once thrived in these deglaciated environments. The exact chemical and physical factors involved in adsorption of the DNA in the basal ice or sediments are not fully understood, but involves interactions between temperature, soil particle surface area, and chemistry and differs among sediment types (Cai, et al., 2006, Hansen, et al., 2006, Corinaldesi, et al., 2008). Another aspect when understanding what the DNA represent in the basal ice samples, are the dynamics whereby the soil gets incorporated into the ice. From the low content of macrofossil or pollen found in the ice it is presumed that the DNA must be bound to the soil particles before it becomes incorporated into the ice matrix (Willerslev, et al., 2007). The incorporation of the soil is influenced by the ice sheet dynamics or glacier movements and depends on the surface elevations and character of the landscape beneath the ice (Knight, 1997). The basal ice samples are therefore representative of the past environment from one or several regions depending on these dynamics (Souchez, 2000).

Access to such basal ice samples from the polar areas is limited but exist from glacier samples that have incorporated debris or from deep ice core drillings that penetrated through the ice sheet to bedrock. DNA preservation is considered optimal in polar areas due to the constant cold temperatures (Willerslev, 2004a) present for the past hundreds of thousands to million years (Sugden, et al., 1995, Andersen, et al., 2004). As the environmental approach for studying ancient DNA in ice is a relatively new research field, few studies have been performed and as such its potential remains relatively unexplored. However, one such example of an environmental study was that performed on basal ice from the Dye 3 ice core in Southern Greenland (Willerslev, et al., 2007). The DNA results provided insights into a past ecosystem indicative of a forested southern Greenland almost 2 million years younger than was presumed from studies on macrofossils found in Northern Greenland (Funder, 1989, Souchez, et al., 1994). Climate models had described the southern range of the Greenland Ice Sheet around Dye 3 as being ice-free during the Eemian (114-130 ky)(Otto-Bliesner, et al., 2006) but these ancient genetic results proved that this part of the ice sheet was much older. Data such as this could potentially provide valuable information for a range of disciplines

because it may help constrain climate models, thus increasing their accuracy, as well as adding to the understanding of the past palaeo-ecosystems and the evolutionary history and distribution of its taxa groups, in situations where data is otherwise unavailable.

Such genetic environmental studies on basal ice can also contribute with new knowledge about past biodiversity and climate changes by providing access to a detailed molecular fossil record of the organisms existing prior to ice coverage. Such data can be used to elucidate more of the complex interactions between climate and biodiversity and how fast responses are detected. The data can contribute to the understanding of evolutionary rates within certain organisms if comparing the genetic diversity from the past with the present representatives of these organisms. However, the risk of contamination when doing environmental studies on ice needs to be approached with the same caution as for microbial studies where standards for clean laboratory work needs to be high at all steps during sampling, processing and analyses. Otherwise, false positives will provide an erroneous picture of the past ecosystems.

## METHODS FOR ANALYZING DNA/BIOLOGY FROM ANCIENT ICE

### Sample Processing

Decontamination. Guidelines for obtaining authentic DNA or culturing results from old samples have been published several times e.g (Nickle, et al., 2002, Pääbo, et al., 2004, Willerslev, 2004a, Gilbert, et al., 2005, Hebsgaard, et al., 2005, Hofreiter, 2008).

Here we present additional guidelines when working on ice:
Any decontamination of the ice cores or samples should be done in positive flow benches in cold rooms where melting is prevented. The removal of outer contaminated surface layers should be done mechanically with sterilized tools as wet surfaces might adhere new airborne contaminants. This is due to the strong adhesion forces between a humid or wet surface and micron-sized particles available in the air (Davies, 1983). The amount of ice that needs to be removed could vary between ice samples and depends on the drilling or cutting technique used. Christner et al. (2005) recommended removing 3 cm of the surface of Lake Vostok core to eliminate contaminants but less could also be adequate if the core quality is high and the surface without cracks. However, as pure diffusion of impurities or gasses can reach about 10mm over 10 years at -20C, due to ageing effect where the ice is changing physical

properties because of changed pressure and temperature conditions during storage (Schwander, et al., 1983, Christner, et al., 2005), the removal of at least 1.5cm is recommended for older cores. Sterile gloves, facemasks, body suit and sterilized tools should be used when decontaminating the samples. The clean tools used for removing the surface should be exchanged for new clean ones at each consecutive scrape, as more material is removed. If a melting device is used for melting the inner part of the core, it is critical to pre-clean all tubes and meltparts of the device that is to be used to remove DNA (e.g. by radiation, acid and heat) to avoid contamination. The basal part of the ice core should be cleaned according to above dry criteria and the melting should be in a clean cold facility. The surface of the samples should always be spiked with a recognizable tracer to monitor the penetration of modern contaminants or drilling fluids. The tracer should as a minimum be of DNA form, but could also include recognizable molecular and cellular contaminants, as suggested in Christner et al. (2005). It is important to apply the tracer on the sample in question and not merely on sham cores that are made of frozen water in similar shape to the ice cores, as these are not fully representative of the sampled ice. If the sham core shows negative contamination results after removing the surface containing contaminants, it is not equally adaptable to the sampled ice.
After decontamination, the samples should be transferred to a clean vessel and melted in a sealed system. They should now be transferred to clean laboratory facilities for DNA extraction and other downstream steps. It is recommended to have duplicates of all samples so that independent replication can be conducted on the samples to compare results.

Laboratory based clean protocols. Generally laboratory procedures should include blanks at all possible steps and follow guidelines that are designed to prevent contamination of samples with low levels of microorganisms or DNA, for example those for ancient DNA work (Cooper & Poinar, 2000, Hofreiter, et al., 2001, Pääbo, et al., 2004, Gilbert, et al., 2005, Willerslev & Cooper, 2005). Replication in another ancient/clean laboratory, enabling independent generation of results that help provide support for the results may also be relevant for ice core studies. However, as discussed by Gilbert et al. (2005) replication does not always rule out false positives if the contamination was not removed prior to analyses, which again highlights the need of proper decontamination. Independent replication of the biodiversity can nevertheless be very useful for environmental samples to compare the number of exact matches or overlap of diversity in the samples, which can help elucidate what is likely to be true diversity found in these ice samples. Evaluating positive sequence

results from a functional point of view is only useful for organisms that have a known function. To minimize contaminant results, consideration about the source and environment from where the detected organisms have come from could easily be biased, as only 1-5% of bacterial diversity is known (Willerslev, et al., 2004, Poole & Willerslev, 2007). It is however informative to use the biological knowledge of the detected diversity as a guideline when interpreting results and also to be aware that the diversity detected could be biased from difference in ability to preserve DNA between organisms (Poole & Willerslev, 2007). A test for the presence of bacterial DNA by simple PCR of the melted sample prior to any downstream culturing, could act as a positive control for microbial material present in the samples.

**Confirmation of Results**

At the DNA sequence level several issues need to be addressed to verify the endogenous nature of these. If for example the samples are ancient it is relevant to investigate whether damage derived sequence modifications exist, as these would cause false sequence diversity. It is also relevant to look into age related differences among modern and ancient sequences to compare for substitution changes equivalent to the time difference between the samples. Further, if samples are either modern or ancient comparing the sequence results with contaminant sequences from various laboratory steps has been suggested to further reduce false positive results.

Damage Profile. To verify the endogenous nature of sampled ancient DNA from old ice samples, one could search for a signal of damage in the sequences. This is a robust indicator of the authenticity of the samples. Despite the cold preservation temperatures of ancient sequences in ice, nucleotide sites accumulate damage with time, most likely occurring due to hydrolytic or oxidative modifications, which principally cause deamination of cytosine and produce error containing sequences (Pääbo, 1989) Other age related modifications of the DNA involve spontaneous events of depurination and inter-strand crosslinks, which prevent amplification of the DNA (Mitchell, et al., 2005, Hansen, et al., 2006). The frequency of inter-strand cross-links (ICL) in permafrost samples has been shown to be high with at least one ICL in every double stranded molecule longer than 100-200bp found (Mitchell, et al., 2005). However, this was later questioned by a study where no such pattern was detected in ancient microbial sequences from Antarctic ice (Bidle, et al., 2007). Inter strand cross-links are critical damage types as they block the polymerase and hence decrease the number of

molecules available for amplification (Mitchell, et al., 2005). The particular damage type is relevant to consider as it might bias the diversity detected in the samples. This is indeed the case for microbes where it has been shown that some groups show better repair mechanisms than e.g. endospore forming groups and would then be preserved longer in the ancient samples (Johnson, et al., 2007). However, the inter-strand cross-links also serve a protective function in preserving the DNA as the double stranded structure is stabilized and less prone to further decay (Mitchell, et al., 2005).

Even though the cold temperatures in ice and permafrost provide similar temperature and dry preservation conditions they differ greatly in the chemistry and in the mineral particles surrounding the DNA molecules. Therefore one could expect to find different degradation processes in these environments and slower damage accumulation in DNA preserved in glacial ice as compared to permafrost. However, even if the damage that accumulates in ice is mainly due to internal spontaneous reactions, the damage that could arise in microbes during aeolian transport onto the ice sheets could be quite significant due to radiation during the atmospheric transport (Mitchell, et al., 2005).

Age approach. As many of the retrieved sequences from the ice cores are old, often >50.000 years old, a damage signal is expected as described above, even if preserved within microbial organisms (Willerslev, et al., 2004b, Bidle, et al., 2007, Johnson, et al., 2007). These damage signals can be statistically tested for and could be used to verify the ancient nature of the samples (Hansen, et al., 2001). Specific substitution patterns are more common than others in old sequences, e.g. mismatches between cytosine to thymine and guanine to adenine are common in ancient DNA due to degradation (Hansen, et al., 2001, Hofreiter, et al., 2001). Increased levels of such substitutions can then be tested for as they involve an excess of cytosine to thymine bases (C-T) in the 5' ends of the sequences due to enhanced cytosine deamination to uracil in single stranded 5' overhanging ends in damaged DNA strands, which cause the complementary pattern of guanine to adenine substitutions (G-A) in the 3'end (Briggs, et al., 2007, Green, et al., 2009, Ginolhac, et al., 2011). Another damage type that could be tested for is an increased level of purines at positions around strand breaks caused by depurination prior to fragmentation of the ancient DNA strand (Briggs, et al., 2007, Krause, et al., 2010).

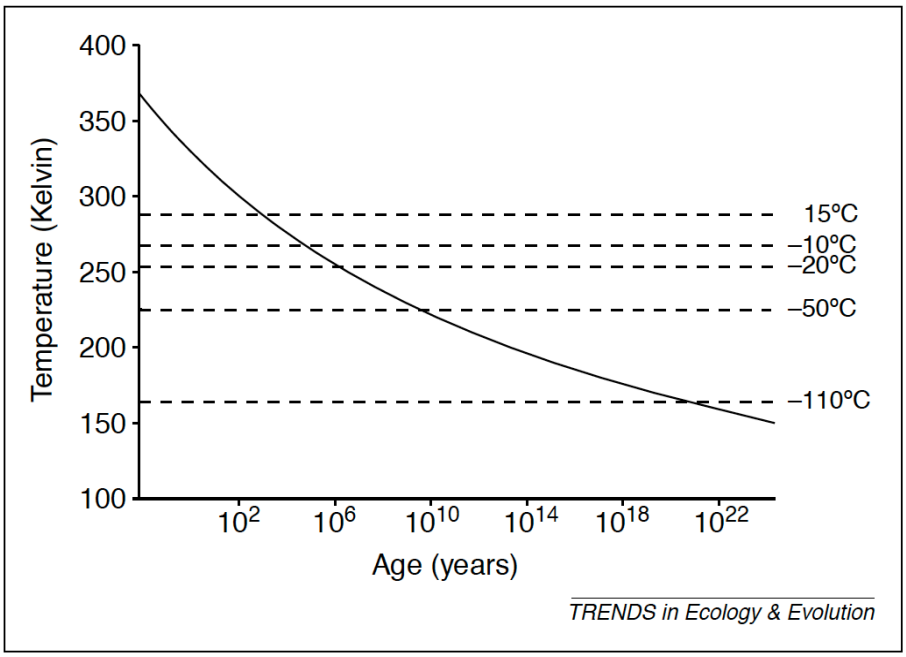
In addition to testing for accumulating damage in old sequences another approach could be to compare the supposed ancient sequences to related modern representatives by means of a relative rate test (Gutierrez & Marin, 1998, Nickle, et al., 2002). This method tests

whether the distance between the ancient sequence and an outgroup is significantly different from the distance between a modern closely related sequence and the outgroup. This difference would then be due to the diversity that would have accumulated in a clock-like manner in modern relatives with time (Hebsgaard, et al., 2005). Another approach that does not assume a molecular clock but takes into account that substitution rates can vary among related organisms could be the relaxed-clock rates test (Sanderson, 2002, Thorne & Kishino, 2002, Hebsgaard, et al., 2005). This approach relies on accurate dating of the sample material from which the sequences are retrieved, and a well-known divergence time for the tested sequences to modern relatives, and between modern sequences and fossils. However, this is often not the case for microbial specimens, which makes the last method more appropriate for higher organisms where better-preserved fossils exist. The relative rates test would be a very powerful way to reveal the nature of the source of DNA, whether representing contaminant or endogenous material. The disadvantage of this method would be that one has to find sequences that would both represent the taxonomic relationship and representing a mutation rate that would have accumulated changes within the timeframe studied. The 16S rRNA sequences commonly used to target most of the microbial sequences in such environmental samples would not be expected to differ more than 1-2% in 50 million years time or $2\text{-}4*10^{-10}$ substitutions per site per year (Wilson, et al., 1987, Moran, et al., 1993, Nickle, et al., 2002). Therefore the short sequences expected to be amplifiable in these old samples might not harbour enough differences and should be exchanged with more specific and faster evolving sequences, depending on the organism one wishes to compare. As mentioned by Hebsgaard et al. (2005) one can never prove the authenticity of the results obtained but they can be tested as thoroughly as available methods allow and then we can fail to disprove the authenticity. This is increasingly important when working with ancient material in low abundance that is prone to contamination and errors.

Another criteria one could use when testing for an age signal in the sequences is to look at the fragmentation level, which will accumulate with time (Willerslev & Cooper, 2005). The expected rate of DNA degradation has been seen to roughly follow the Arrhenius equation for spontaneous chemical decay that depends on temperature: $k=A*\exp(-Ea/RT)$, where k is the rate constant, A is the pre-exponential factor that is depending on the reaction, Ea is the activation energy, R is the gas constant (8.31 KJ mol-1 at 1 atm) and T is the temperature in Kelvin (Lindahl & Nyberg, 1972). Hence, temperature should have an exponential effect in either increasing or decreasing the rate of decay depending on the

temperature changes in the environment. The expected survival time for a 100 bp bacterial genome has been suggested to follow a trend illustrated in Figure 5.

**Figure 5**. Decay trend of long-term DNA survival of 100bp long fragments at different temperatures, if damage is assumed to be equally distributed over the genome at all purine sites. The calculations were based on the Arrhenius equation and depurination kinetics at pH 7.4 and 70°C (Lindahl & Nyberg, 1972). From Willerslev et al. (2004a)




Despite the low temperatures in ice and permafrost samples a fragmentation pattern is expected to increase with age and could be detected by simple PCR. This would be evident in old samples, even if some viable cells still exist because the number of viable cells in old samples would be much lower than dead cells. An inverse relationship between length of PCR product and number of cycles would indicate degradation and old age of the samples tested (Hebsgaard, et al., 2005). Such degradation patterns have been shown for the bacterial permafrost and ice studies from both PCR and DNA quantification points of view but also in terms of diversity found (Willerslev, et al., 2004b, Bidle, et al., 2007, Johnson, et al., 2007, Veerapaneni, 2009). These studies also found a decrease in sequence diversity with time and Bidle et al. (2007) suggest using a DNA-size model as a new independent geomolecular age indicator in ice. However, a suite of other chemical and physical factors besides temperature cause slight deviations in the decay rate compared to the Arrhenius equation (Willerslev, 2004a). Such factors include pressure within the ice, pH, humidity,

presence of mono- and multivalent cations, redox potentials, surface area of the soil particles in the ice and radiation (Lindahl, 1993, Frostegard, et al., 1999, Cai, et al., 2006, Andersen, et al., 2011). The complexity of interactions between these different factors and to what degree they might be either synergistic or antagonistic is not understood in detail. Inter-strand crosslinks seem to have a higher impact on the decay process than single strand breaks from depurination reactions and speed up the degradation (Mitchell, et al., 2005). Still, local conditions at the sample site seem to be the most important factors to investigate for predicting the DNA preservation as these might vary significantly among sample sites.

Databases for contaminants. Contaminants of both plant, animal and microbial origin do occur in most laboratory facilities and need to be monitored.
Databases with contaminant sequences obtained from controls taken at all possible steps during the various handling and laboratory steps could therefore be used locally to hold any new results up against, to verify the quality of the obtained sequences (Bulat, et al., 2004). This applies for both studies of environmental reconstructions and descriptions of microbial ecosystems. However this strategy might also eliminate true positive results. The problem with contaminant databases arises as a true sample sequence might match exactly a common known contaminant, be very different from the contaminants or show close similarity to a contaminant. The two latter cases might show the true sample diversity but the one showing high similarity is difficult to prove as its diversity falls within the diversity of the contaminant database (Poole & Willerslev, 2007). Another difficulty when assigning true positive results from ancient ice is that new unknown sequences do not per se qualify as significant results, as a contaminant could also fall into this category. Further, comparing results to the diversity of previously published results that could also contain contaminants provide another risk of false positive signal. Other reasons underlying false positive results could be due to sequence artefacts caused by chimeric DNA strands, post mortem hydrolytic damage or secondary structures that cause deletion patterns during the PCR. Sequencing errors can also cause false diversity and are especially common for sequences with homopolymer levels when sequenced on the FLX sequencing platform while for sequences generated on the Illumina sequencing platform errors occur in the 3' end of the reads (Dohm, et al., 2008, Gilles, et al., 2011, Quince, et al., 2011).

Controls from DNA to microbes. The difference in outcome between studies that have or lack strict decontamination protocols and replicate or fail to replicate their results is significant

when comparing studies based on similar sample material. Very few studies do extractions on more than a few mL of ice (Karl, et al., 1999, Sheridan, et al., 2003, Miteva, et al., 2004, Tung, et al., 2006), which has been shown to contain so few DNA molecules that they are not detectable. It is therefore interesting that such samples provide positive culturing results when these organisms do not leave a genetic profile measurable by PCR (Abyzov, et al., 1998, Abyzov, et al., 2004, Tung, et al., 2005, Miteva, et al., 2009). One could argue that a pipeline where DNA amplification is used as a first step to indicate if preserved biological material is present in the sample followed by replication should always go before any culturing studies. As the PCR is potentially able to start on as few as 10 – 50 templates (Anderson-Carpenter, et al., 2011) (if inhibition is eliminated) this should be a sensitive test to reveal the actual diversity in the samples. Such amplifications could be done with universal primers targeting short fragments of conserved regions among microbes as e.g. the 18S or 16S rRNA regions.

## LINKING ICE CORE GENETICS AND PAST ENVIRONMENTS OR CLIMATES

With high standard clean protocols, DNA from past organisms preserved in the ice can add important new information on past climates and ecosystems. When DNA is obtained from a variety of past organisms the information can be used to describe ancient ecosystems on various levels; from microbes to higher organisms of plants and animals. Such information is highly relevant for our understanding of evolutionary potentials and adaptation to changing climates through time.

Before the second generation of sequencing techniques, metagenomics was only applied through whole genome shotgun amplification that requires a contamination prone amplification step. However, with the next generation sequence platforms it is possible to deep sequence directly from environmental samples and reveal the biodiversity present (Valentini, et al., 2009, Sønstebø, et al., 2010), which provide new means to study low abundance fossil biomolecules in ice-covered areas. The continuous development of second and third generation sequencing presents the possibility of sequencing single molecules of fragments up to more than 500 bp (PacBio, GS FLX) whereby it becomes possible to detect individuals within an environmental or ice sample. Hence, these high throughput-sequencing techniques facilitate new possibilities for environmental sample types that accommodate population genetic studies on several species within an environmental sample. This is a huge step forward in which analyses can be done on past populations dynamics through time.

Comparisons between diversity at different time periods could provide detailed understanding of population fluctuations through time in relation to climate change.

Temporal changes in biological diversity at the microbial level can further be related to differences in climate periods and linked to changed wind patterns through time. This is interesting because it provides information about the geographical and biological source of the material deposited, which could provide useful constraints for climate models.

Ice core genetics provide information for the reconstruction of the immediate past environment in Greenland and Antarctica just before glaciation. In addition, it also provides very important pieces of information for reconstruction of the history of the Greenland and Antarctic Ice Sheet since its inception, sudden or gradual. DNA extracted from the basal parts of the debris containing ice cores can be used to describe the past ecosystem and to identify ecological constraints on temperature ranges and precipitation based on the taxa groups present. Coupling the identification of biological material with the dating of these ice samples can reveal the timing and duration of the different communities and determine when was the last unglaciated period.

## FUTURE PERSPECTIVES

Biology in ice from the polar areas has been studied from two angles, the first focusing on characterizing the diversity of microorganisms within bacteria, fungi, archaea, viruses and eukaryotes, and the second studying the genetic diversity to infer past climate fluctuations or environmental reconstructions of past ecosystems through analyses of glacier and basal ice. Both approaches provide valuable knowledge relevant for a variety of research themes within, evolutionary biology, ecology and climate research. Applying biological based methods on ice samples is an unique opportunity to address questions either as an additional and independent climate or age proxy or as a new palaeo proxy where other methods have come to terms due to the limited access to fossil material available in these areas. The shared feature among these studies is that they are very contamination prone because of the low quantity of the biological material present, which can be easily overshadowed by modern DNA introduced during sampling, handling or processing of the ice samples. The molecular or microbial protocols therefore need to be of the highest available standards and should be continuously updated as the field develops. These steps are critical for validating the results and consolidate the research field with the advancing opportunities for studying low-level molecules or organisms in environmental samples.

Future questions that would be relevant to address involve a better understanding of how the DNA is preserved in these environments. This is relevant for more optimized protocols that both extract the endogenous DNA of interest and reduce the contamination. Such steps are also relevant for following sound guidelines when applying the new sequencing technologies on this delicate material to reduce a large quantity of false positives. Understanding more of the mechanisms in DNA adsorption and incorporation to the ice would help clarifying new sampling strategies focusing on the best places to search for preserved DNA in the basal and glacier ice. Last, it could be interesting to combine climate and biodiversity models to describe the past responses and ancient DNA in cross-disciplinary projects. This could promote new insights to the dynamics between ecosystem responses to climate changes, which has a preserved frozen signature in these areas and are relevant for future scenarios.

References

Abyzov S, Mitskevich IN, Poglazova MN, et al. (1999) Antarctic ice sheet as an object for methodological problems of exobiology. *Life Sciences: Exobiology* **23**: 371-376.

Abyzov SS, Mitskevich IN & Poglazova MN (1998) Microflora of the deep glacier horizons of Central Antarctica. *Microbiology* **67**: 451-458.

Abyzov SS, Lipenkov VY, Bobin NE & Koudryashov BB (1982) The Microflora of the Central Antarctic Glacier and the Control Methods of the Sterile Isolation of the Ice Core for Microbiological Analysis. *Izvestiya Akademii Nauk Sssr Seriya Biologicheskaya* 537-548.

Abyzov SS, Mitskevich IN, Zhukova TY, Kuzhinovskii VA & Poglazova MN (1993) Microorganism Numbers in the Deep Layers of the Central Antarctic Ice-Sheet. *Microbiology* **62**: 130-135.

Abyzov SS, Hoover RB, Imura S, Mitskevich IN, Naganuma T, Poglazova MN & Ivanov MV (2004) Use of different methods for discovery of ice-entrapped microorganisms in ancient layers of the Antarctic glacier. *Space Life Sciences: Search for Signatures of Life, and Space Flight Environmental Effects on the Nervous System* **33**: 1222-1230.

Abyzov SS, Mitskevich IN, Poglazova MN, et al. (1998) Antarctic ice sheet as a model in search of life on other planets. *Life Sciences: Exobiology* **22**: 363-368.

Abyzov SS, Mitskevich IN, Poglazova MN, et al. (2001) Microflora in the basal strata at Antarctic ice core above the Vostok Lake. *Space Life Sciences: Living Organisms, Biological Processes and the Limits of Life* **28**: 701-706.

Andersen K, Bird K-L, Rasmussen M, et al. (2011) Meta-Barcoding of 'Dirt' DNA from Soil Reflects Vertebrate Biodiversity. *Molecular Ecology*.

Andersen KK, Azuma N, Barnola JM, et al. (2004) High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**: 147-151.

Anderson-Carpenter LL, McLachlan JS, Jackson ST, Kuch M, Lumibao CY & Poinar HN (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evol Biol* **11**: 30.

Barnes PRF & Wolff EW (2004) Distribution of soluble impurities in cold glacial ice. *Journal of Glaciology* **50**: 311-324.

Barnes PRF, Wolff EW, Mader HM, Udisti R, Castellano E & Rothlisberger R (2003) Evolution of chemical peak shapes in the Dome C, Antarctica, ice core. *Journal of Geophysical Research-Atmospheres* **108**: 15.

Bender ML, Burgess E, Alley RB, Barnett B & Clow GD (2010) On the nature of the dirty ice at the bottom of the GISP2 ice core. *Earth and Planetary Science Letters* **299**: 466-473.

Bennike O (2002) Late Quaternary history of Washington Land, North Greenland. *Boreas* **31**: 260-272.

Bentley CR & Koci BR (2007) Drilling to the beds of the Greenland and Antarctic ice sheets: a review. *Annals of glaciology* **47**: 1-9.

Bidle KD, Lee S, Marchant DR & Falkowski PG (2007) Fossil genes and microbes in the oldest ice on Earth. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 13455-13460.

Biggerstaff JP, Le Puil M, Weidow BL, Prater J, Glass K, Radosevich M & White DC (2006) New methodology for viability testing in environmental samples. *Molecular and Cellular Probes* **20**: 141-146.

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.

Bulat S, Alekhina I, Lipenkov V, Lukin V, Marie D & Petit J (2009) Cell concentrations of microorganisms in glacial and lake ice of the Vostok ice core, East Antarctica. *Microbiology* **78**: 808-810.

Bulat SA, Alekhina IA, Blot M, et al. (2004) DNA signature of thermophilic bacteria from the aged accretion ice of Lake Vostok, Antarctica: implications for searching for life in extreme icy environments. *International Journal of Astrobiology* **3**: 1-12.

Cai P, Huang Q, Jiang D, Rong X & Liang W (2006) Microcalorimetric studies on the adsorption of DNA by soil colloidal particles. *Colloids and surfaces. B, Biointerfaces* **49**: 49-54.

Castello JD, Jacobi GDB & V (1993) Tomato mosaic virus of red spruce on Whiteface Mt. in New York. *Phytopathology* **83**: 1351.

Castello JD, Rogers SO, Smith JE, Starmer WT & Zhao Y (2005) *Plant and bacterial viruses in the Greenlandic Ice Sheet*. Princeton University Press, New Jersey, USA.

Castello JD, Lakshman DK, Tavantzis SM, et al. (1995) Detection of infectious tomato mosaic tobamovirus in fog and clouds. *Phytopathology* **85**: 1138.

Castello JD, Rogers SO, Starmer WT, et al. (1999) Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biology* **22**: 207-212.

Catranis C (1991) Microorganisms entrapped in glacial ice. *Antarctic journal of the United States* **26**: 234.

Christner B (2002) Detection, recovery, isolation and characterization of bacteria in glacial ice and Lake Vostok accretion ice. Thesis, Ohio University, USA.

Christner B, Mosley-Thompson E & Thompson… L (2000) Recovery and identification of viable bacteria immured in glacial ice. Thesis, Ohio University, USA.

Christner BC (2010) Bioprospecting for microbial products that affect ice crystal formation and growth. *Applied Microbiology and Biotechnology* **85**: 481-489.

Christner BC, Mosley-Thompson E, Thompson LG & Reeve JN (2001) Isolation of bacteria and 16S rDNAs from Lake Vostok accretion ice. *Environmental Microbiology* **3**: 570-577.

Christner BC, Mosley-Thompson E, Thompson LG & Reeve JN (2003) Bacterial recovery from ancient glacial ice. *Environmental Microbiology* **5**: 433-436.

Christner BC, Mikucki JA, Foreman CM, Denson J & Priscu JC (2005) Glacial ice cores: A model system for developing extraterrestrial decontamination protocols. *Icarus* **174**: 572-584.

Convey P & Stevens MI (2007) Antarctic Biodiversity. *Science* **317**: 1877-1878.

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

Corinaldesi C, Beolchini F & Dell'Anno A (2008) Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Molecular Ecology* **17**: 3939-3951.

D'Elia T (2008) Isolation of bacteria and fungi from Lake Vostok accretion ice. Thesis, Ohio University, USA.

D'Elia T, Veerapaneni R & Rogers SO (2008) Isolation of microbes from Lake Vostok accretion ice. *Applied and Environmental Microbiology* **74**: 4962-4965.

D'Elia T, Veerapaneni R, Theraisnathan V & Rogers SO (2009) Isolation of fungi from Lake Vostok accretion ice. *Mycologia* **101**: 751-763.

Dahl-Jensen D (2003) Basal melt at NorthGRIP modeled from borehole, ice-core and radio-echo sounder observations. *Annals of glaciology* **37**: 207.

Dash JG, Fu HY & Wettlaufer JS (1995) The Premelting of Ice and Its Environmental Consequences. *Reports on Progress in Physics* **58**: 115-167.

Dash JG, Rempel AW & Wettlaufer JS (2006) The physics of premelted ice and its geophysical consequences. *Reviews of Modern Physics* **78**: 695-741.

Davies CN (1983) Filtration of aerosols. *Journal of Aerosol Science* **14**: 147-161.

Dohm JC, Lottaz C, Borodina T & Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research* **36**: e105.

Faria SH, Freitag J & Kipfstuhl S (2010) Polar ice structure and the integrity of ice-core paleoclimate records. *Quaternary Science Reviews* **29**: 338-351.

Fell JW, Scorzetti G, Connell L & Craig S (2006) Biodiversity of micro-eukaryotes in Antarctic Dry Valley soils with $<$ 5% soil moisture. *Soil Biology & Biochemistry* **38**: 3107-3119.

Francis JE, Haywood AM, Ashworth AC & Valdes PJ (2007a) Tundra environments in the Neogene Sirius Group, Antarctica: evidence from the geological record and coupled atmosphere-vegetation models. *Journal of the Geological Society* **164**: 317-322.

Frostegard A, Courtois S, Ramisse V, et al. (1999) Quantification of bias related to the extraction of DNA directly from soils. *Applied and Environmental Microbiology* **65**: 5409-5420.

Fukazawa H, Sugiyama K, Mae S, Narita H & Hondoh T (1998) Acid ions at triple junction of Antarctic ice observed by Raman scattering. *Geophysical research letters* **25**: 2845-2848.

Funder S (1989) Quaternary Geology of the Ice-Free Areas and Adjacent Shelves of Greenland. *Quaternary Geology of Canada and Greenland* **K-1**: 741-792.

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

Gilichinsky D, Rivkina E, Shcherbakova V, Laurinavichuis K & Tiedje J (2003) Supercooled water brines within permafrost - An unknown ecological niche for microorganisms: A model for astrobiology. *Astrobiology* **3**: 331-341.

Gilichinsky D, Wilson G, Friedmann E, McKay C, Sletten R & Rivkina E (2007) Microbial Populations in Antarctic permafrost: Biodiversity, state, age, and implication for astrobiology. *Astrobiology* **7**: 275-311.

Gilles A, Meglecz E, Pech N & Ferreira S (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Bioinformatics* 30.

Ginolhac A, Rasmussen M, Gilbert MT, Willerslev E & Orlando L (2011) mapDamage: testing for damage patterns in ancient DNA sequences. *Bioinformatics*.

Gow A & Meese D (1996) Nature of basal debris in the GISP2 and Byrd ice cores and its relevance to bed processes. *Annals of glaciology* **22**: 134-140.

Graur D & Pupko T (2001) The permian bacterium that isn't. *Molecular Biology and Evolution* **18**: 1143-1146.

Green RE, Briggs AW, Krause J, et al. (2009) The Neandertal genome and ancient DNA authenticity. *The EMBO journal* **28**: 2494-2502.

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.

Haile J, Holdaway R, Oliver K, et al. (2007) Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? *Molecular Biology and Evolution* **24**: 982-989.

Hansen A, 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.

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.

Haywood AM, Valdes PJ, Sellwood BW & Kaplan JO (2002) Antarctic climate during the middle Pliocene: model sensitivity to ice sheet variation. *Palaeogeography Palaeoclimatology Palaeoecology* **182**: 93-115.

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

Hebsgaard MB, Gilbert MTP, Arneborg J, et al. (2009) 'The Farm Beneath the Sand'- an archaeological case study on ancient 'dirt' DNA. *Antiquity* **83**: 430-444.

Hofreiter M (2008) Palaeogenomics. *Comptes Rendus Palevol* **7**: 113-124.

Hofreiter M, Mead JI, Martin P & Poinar HN (2003) Molecular caving. *Current Biology* **13**: R693-R695.

Hofreiter M, Serre D, Poinar HN, Kuch M & Paabo S (2001) Ancient DNA. *Nat Rev Genet* **2**: 353-359.

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

Huybrechts P (1993) Glaciological Modelling of the Late Cenozoic East Antarctic Ice Sheet: Stability or Dynamism? *Geografiska Annaler. Series A, Physical Geography* **75**: 221-238.

Iizuka Y, Horikawa S, Sakurai T, Johnson S, Dahl-Jensen D, Steffensen JP & Hondoh T (2008) A relationship between ion balance and the chemical compounds of salt inclusions found in the Greenland Ice Core Project and Dome Fuji ice cores. *Journal of Geophysical Research-Atmospheres* **113**.

Johnsen SJ, Dahl-Jensen D, Gundestrup N, et al. (2001) Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP. *Journal of Quaternary Science* **16**: 299-307.

Johnson SS, Hebsgaard MB, Christensen TR, et al. (2007) Ancient bacteria show evidence of DNA repair. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 14401-14405.

Jouzel J & Souchez R (1982) Melting-refreezing at the glacier sole and the isotopic composition of the ice'. ed.^eds.), p.^pp.

Junge K, Krembs C, Deming J, Stierle A & Eicken H (2001) A microscopic approach to investigate bacteria under in situ conditions in sea-ice samples. *Annals of Glaciology, Vol 33* **33**: 304-310.

Karl DM, Bird DF, Bjorkman K, Houlihan T, Shackelford R & Tupas L (1999) Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* **286**: 2144-2147.

Knight PG (1997) The basal ice layer of glaciers and ice sheets. *Quaternary Science Reviews* **16**: 975-993.

Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A & Paabo S (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current biology : CB* **20**: 231-236.

Lambert F, Delmonte B, Petit JR, et al. (2008) Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452**: 616-619.

Lavire C, Normand P, Alekhina I, et al. (2006) Presence of Hydrogenophilus thermoluteolus DNA in accretion ice in the subglacial Lake Vostok, Antarctica, assessed using rrs, cbb and hox. *Environmental Microbiology* **8**: 2106-2114.

Lewis K, Epstein S, Godoy VG & Hong SH (2008) Intact DNA in ancient permafrost. *Trends in Microbiology* **16**: 92-94.

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

Lindahl T & Nyberg B (1972) Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **11**: 3610-3618.

Lorenz MG & Wackernagel W (1994) Bacterial Gene-Transfer by Natural Genetic-Transformation in the Environment. *Microbiological Reviews* **58**: 563-602.

Lyakh SP & Abyzov SS (1974) Adaptation of Antarctic Microflore to Frost and Dryness as Basic Extreme Factors of Medium. *Izvestiya Akademii Nauk Sssr Seriya Biologicheskaya* 688-698.

Lydolph MC, Jacobsen J, Arctander P, et al. (2005) Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and Environmental Microbiology* **71**: 1012-1017.

Ma LJ, Rogers SO, Catranis CM & Starmer WT (2000) Detection and characterization of ancient fungi entrapped in glacial ice. *Mycologia* **92**: 286-295.

Marchant DR, Swisher CC, Lux DR, West DP & Denton GH (1993) Pliocene Paleoclimate and East Antarctic Ice-Sheet History from Surficial Ash Deposits. *Science* **260**: 667-670.

Metzker ML (2009) Sequencing technologies—the next generation. *Nat Rev Genet* **11**: 31-46.

Mitchell D, Willerslev E & Hansen A (2005) Damage and repair of ancient DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**: 265-276.

Miteva V (2010) Diversity and ecology of psychrophilic microorganisms. *Research in Microbiology*.

Miteva V, Sowers T & Brenchley J (2007) Production of N2O by ammonia oxidizing bacteria at subfreezing temperatures as a model for assessing the N2O anomalies in the vostok ice core. *Geomicrobiology Journal* **24**: 451-459.

Miteva V, Teacher C, Sowers T & Brenchley J (2009) Comparison of the microbial diversity at different depths of the GISP2 Greenland ice core in relationship to deposition climates. *Environmental Microbiology* **11**: 640-656.

Miteva VI & Brenchley JE (2005) Detection and isolation of ultrasmall microorganisms from a 120,000-year-old Greenland glacier ice core. *Applied and Environmental Microbiology* **71**: 7806-7818.

Miteva VI, Sheridan PP & Brenchley JE (2004) Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Applied and Environmental Microbiology* **70**: 202-213.

Mohlmann DTF (2009) Are nanometric films of liquid undercooled interfacial water bio-relevant? *Cryobiology* **58**: 256-261.

Moran NA, Munson MA, Baumann P & Ishikawa H (1993) A Molecular Clock in Endosymbiotic Bacteria Is Calibrated Using the Insect Hosts. *Proceedings of the Royal Society of London Series B-Biological Sciences* **253**: 167-171.

Mulvaney R, Wolff EW & Oates K (1988) Sulfuric-Acid at Grain-Boundaries in Antarctic Ice. *Nature* **331**: 247-249.

Nickle DC, Learn GH, Rain MW, Mullins JI & Mittler JE (2002) Curiously modern DNA for a "250 million-year-old" bacterium. *Journal of Molecular Evolution* **54**: 134-137.

Nye JF (1989) The Geometry of Water Veins and Nodes in Polycrystalline Ice. *Journal of Glaciology* **35**: 17-22.

Nye JF & Frank FC (1973) Hydrology of the intergranualr veins in a temperate glacier. *Symposium on the Hydrology of Glaciers (Proceedings of the Cambridge Symposium,*
*September 1969)* pp. 157-161.

Ohno H, Igarashi M & Hondoh T (2005) Salt inclusions in polar ice core: Location and chemical form of water-soluble impurities. *Earth and Planetary Science Letters* **232**: 171-178.

Ohno H, Lipenkov VY & Hondoh T (2010) Formation of air clathrate hydrates in polar ice sheets: heterogeneous nucleation induced by micro-inclusions. *Journal of Glaciology* **56**: 917-921.

Otto-Bliesner BL, Marsha SJ, Overpeck JT, Miller GH, Hu AX & Mem CLIP (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**: 1751-1753.

Pääbo S (1989) Ancient DNA - Extraction, Characterization, Molecular-Cloning, and Enzymatic Amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 1939-1943.

Pääbo S, Poinar H, Serre D, et al. (2004) Genetic Analyses from Ancient DNA. *Annual Review of Genetics* **38**: 645-679.

Poole AM & Willerslev E (2007) Can identification of a fourth domain of life be made from sequence data alone, and could it be done on mars? *Astrobiology* **7**: 801-814.

Price P (2007) Microbial life in glacial ice and implications for a cold origin of life. *Fems Microbiology Ecology* **59**: 217-231.

Price PB (2000) A habitat for psychrophiles in deep Antarctic ice. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1247-1251.

Price PB (2004) Life in solid ice on Earth and other planetary bodies. *Bioastronomy 2002: Life among the Stars,*(Norris RP & Stootman FH, ed.^eds.), p.^pp. 363-366.

Price PB (2009) Microbial genesis, life and death in glacial ice. *Canadian Journal of Microbiology* **55**: 1-11.

Price PB & Sowers T (2004) Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 4631-4636.

Priscu JC & Christner BC (2004) *Earth's icy biosphere*.

Priscu JC, Christner B, Foreman C & Royston-Bishop G (2006) Biological material in ice cores. *Encyclopedia of Quaternary Sciences* **2**: 1156–1166.

Priscu JC, Adams EE, Lyons WB, et al. (1999) Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science* **286**: 2141-2144.

Quince C, Lanzen A, Davenport RJ & Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38.

Rempel A, Waddington E, Wettlaufer J & Worster M (2001) Possible displacement of the climate signal in ancient ice by premelting and anomalous diffusion. *Nature* **411**: 568-571.

Rogers SO, Theraisnathan V, Ma LJ, et al. (2004) Comparisons of protocols for decontamination of environmental ice samples for biological and molecular examinations. *Applied and Environmental Microbiology* **70**: 2540-2544.

Rohde RA & Price PB (2007) Diffusion-controlled metabolism for long-term survival of single isolated microorganisms trapped within ice crystals. *Proc Natl Acad Sci U S A* **104**: 16592-16597.

Rohde RA, Price PB, Bay RC & Bramall NE (2008) In situ microbial metabolism as a cause of gas anomalies in ice. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 8667-8672.

Salzmann U, Williams M, Haywood AM, Johnson ALA, Kender S & Zalasiewicz J (2011) Climate and environment of a Pliocene warm world. *Palaeogeography, Palaeoclimatology, Palaeoecology* **309**: 1-8.

Samyn D (2005a) Structural and Geochemical Analysis of
Basal Ice from Taylor Glacier, Antarctica:
On the Role and Behaviour of the Interstitial Fluid Phase. Phd
Thesis, Universit´e Libre de Bruxelles, Belgium, Bruxelles.

Samyn D, Fitzsimons SJ & Lorrain RD (2005b) Strain-induced phase changes within cold basal ice from Taylor Glacier, Antarctica, indicated by textural and gas analyses. *Journal of Glaciology* **51**: 611-619.

Samyn D, Fitzsimons SJ & Lorrain RD (2010) Rotating micro-structures in Antarctic cold basal ice: implications for glacier flow and its interpretation. *International Journal of Earth Sciences* **99**: 1849-1857.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* **19**: 101-109.

Schwander J, Neftel A, Oeschger H & Stauffer B (1983) Measurement of direct current conductivity on ice samples for climatological applications. *The Journal of Physical Chemistry* **87**: 4157-4160.

Segawa T, Ushida K, Narita H & Kanda… H (2010) Bacterial communities in two Antarctic ice cores analyzed by 16S rRNA gene sequencing analysis. ed.^eds.), p.^pp.

Shendure J & Ji H (2008) Next-generation DNA sequencing. Vol. 26 ed.^eds.), p.^pp. 1135-1145.

Sheridan PP, Miteva VI & Brenchley JE (2003) Phylogenetic analysis of anaerobic psychrophilic enrichment cultures obtained from a Greenland glacier ice core. *Applied and Environmental Microbiology* **69**: 2153-2160.

Simon C, Wiezer A, Strittmatter AW & Daniel R (2009) Phylogenetic Diversity and Metabolic Potential Revealed in a Glacier Ice Metagenome. *Appl. Environ. Microbiol.* **75**: 7519-7526.

Skidmore ML, Foght JM & Sharp MJ (2000) Microbial life beneath a high Arctic glacier. *Applied and Environmental Microbiology* **66**: 3214-3220.

Sønstebø JH, Gielly L, Brysting AK, et al. (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*.

Souchez R (2000) Basal ice formation and deformation in central Greenland: a review of existing and new ice core data. *Special publication - Geological Society of London* **176**: 13.

Souchez R, Lemmens M, Lorrain R & Tison JL (1978) Pressure-melting within a glacier indicated by the chemistry of re-gelation ice. *Nature* **273**: 454-456.

Souchez R, Bouzette A, Clausen HB, Jonsen SJ & Jouzel J (1998) A stacked mixing sequence at the base of the Dye 3 core, Greenland. *Geophysical research letters* **25**: 1943-1946.

Souchez R, Samyn D, Lorrain R, Pattyn F & Fitzsimons S (2004) An isotopic model for basal freeze-on associated with subglacial upward flow of pore water. *Geophysical research letters* **31**: 4.

Souchez R, Tison J, L, et al. (1994) Stable isotopes in the basal silty ice preserved in the Greenland Ice Sheet at summit; environmental implications. *Geophys. Res. Lett.* **21**: 693-696.

Sowers T (2001) N2O record spanning the penultimate dÈglaciation from the Vostok ice core. *J. Geophys. Res.* **106**: 31903-31914.

Steven B, Léveillé R, Pollard W & Whyte L (2006) Microbial ecology and biodiversity in permafrost. *Extremophiles* **10**: 259-267.

Sugden DE (1996) The East Antarctic Ice Sheet: unstable ice or unstable ideas? *Transactions of the Institute of British Geographers* 443-454.

Sugden DE, Marchant DR, Potter N, Souchez RA, Denton GH, Swisher CC & Tison JL (1995) Preservation of Miocene Glacier Ice in East Antarctica. *Nature* **376**: 412-414.

Svensson A (2011) Annual layering in the NGRIP ice core during the Eemian. *Climate of the past discussions* **7**: 749.

Svensson A, Nielsen SrW, Kipfstuhl S, et al. (2005) Visual stratigraphy of the North Greenland Ice Core Project (NorthGRIP) ice core during the last glacial period. *J. Geophys. Res.* **110**: D02108.

Thorne JL & Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51**: 689-702.

Tison JL, Souchez R, Wolff EW, Moore JC, Legrand MR & de Angelis M (1998) Is a periglacial biota responsible for enhanced dielectric response in basal ice from the Greenland Ice Core Project ice core? *Journal of Geophysical Research-Atmospheres* **103**: 18885-18894.

Traversi R, Becagli S, Castellano E, et al. (2009) Sulfate Spikes in the Deep Layers of EPICA-Dome C Ice Core: Evidence of Glaciological Artifacts. *Environmental Science & Technology* **43**: 8737-8743.

Tung HC, Bramall NE & Price PB (2005) Microbial origin of excess methane in glacial ice and implications for life on Mars. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 18292-18296.

Tung HC, Price PB, Bramall NE & Vrdoljak G (2006) Microorganisms metabolizing on clay grains in 3-km-deep Greenland basal ice. *Astrobiology* **6**: 69-86.

Valentini A, Pompanon F & Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution* **24**: 110-117.

Veerapaneni R (2009) Analysis and Characterization of Microbes from Ancient Glacial Ice. Thesis, Ohio University, USA.

Veerapaneni R, D'Elia T & Rogers S (2007) Analysis and characterization of microbes from ancient glacial ice. *Abstracts of the General Meeting of the American Society for Microbiology* **107**: 480.

Vogel SW (2003) Distribution of basal melting and freezing beneath tributaries of Ice Stream C: implication for the Holocene decay of the West Antarctic ice sheet. *Annals of glaciology* **36**: 273.

Vreeland RH (2007) Isolation of live Cretaceous (121-112 million years old) halophilic Archaea from primary salt crystals (vol 24, pg 275, 2007). *Geomicrobiology Journal* **24**: 545-545.

Vreeland RH, Rosenzweig WD, Lowenstein T, Satterfield C & Ventosa A (2006) Fatty acid and DNA analyses of Permian bacteria isolated from ancient salt crystals reveal differences with their modern relatives. *Extremophiles* **10**: 71-78.

Willerslev E (2003) Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. *Science* **300**: 791-795.

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

Willerslev E, Hansen AJ, Ronn R, et al. (2004b) Long-term persistence of bacterial DNA. *Current Biology* **14**: R9-R10

Willerslev E, Hansen AJ, Christensen B, Steffensen JP & Arctander P (1999) Diversity of Holocene life forms in fossil glacier ice. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 8017-8021.

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

Willerslev E, Cappellini E, Boomsma W, et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**: 111-114.

Wilson AC, Ochman H & Prager EM (1987) Molecular time scale for evolution. *Trends in Genetics* **3**: 241-247.

Wolff E, Mulvaney R & Oates K (1988) The location of impurities in Antarctic ice. *Ann. Glaciol* **11**: 194–197.

Worobey M (2008) Phylogenetic Evidence against Evolutionary Stasis and Natural Abiotic Reservoirs of Influenza A Virus. *J. Virol.* **82**: 3769-3774.

Yoshimura Y, Kohshima S, Takeuchi N, Seko K & Fujita K (2000) Himalayan ice-core dating with snow algae. *Journal of Glaciology* **46**: 335-340.

Zhang G, Shoham D, Gilichinsky D, Davydov S, Castello JD & Rogers SO (2007) Evidence of influenza A virus RNA in Siberian lake ice (vol 80, pg 12229, 2006). *Journal of Virology* **81**: 2538-2538.

# Chapter II

TITLE:

Long-term Survival of Diverse Ancient Plant DNA in Basal Glacial Ice at Contrasting Environmental Conditions

# Long-term Survival of Diverse Ancient Plant DNA in Basal Glacial Ice at Contrasting Environmental Conditions

Astrid M. Z. Schmidt[1,5], Bo Elberling[2,3], Anders Michelsen[4], . Kurt Kjær[1], M. Thomas P. Gilbert, Ludovic Orlando[1], Jørgen Peder Steffensen[5], Eske Willerslev[1*]

[1]Centre for GeoGenetics, Natural History Museum, University of Copenhagen, Øster Voldgade 5-7, DK-1350, Denmark

[2]Department of Geography and Geology, University of Copenhagen, Øster Voldgade 10, DK-1350, Denmark

[3]UNIS, The University Centre in Svalbard, N-9171Longyearbyen, Norway.

[4]Department of Biology, University of Copenhagen, Øster Farimagsgade 2D, DK-1353, Denmark

[5]Centre for Ice and Climate, Niels Bohr Institute, University of Copenhagen, Juliane Mariesvej 30, DK-2100, Denmark

Keywords: ancient DNA, soil geochemical characteristics, basal ice, preservation conditions.

*Author of correspondence: ewillerslev@snm.ku.dk

**Abstract**

**Diverse ancient DNA of plants and animals can be recovered directly from sediment inclusions embedded in the basal part of deep ice cores. However, ice core material is extremely precious, thus whether destructive sampling, such as that required for DNA analyses remains, should be undertaken is a critical issue. So far no methods exist, beyond trial and error, for determining whether a basal ice core sample is likely to yield DNA information of past plant communities. Thus, investigation into factors that determine long-term DNA survival and presence in basal ice is needed. Here, we characterise chemical and physical characteristics of four rare basal ice core samples from Greenland that either previously have, or have not, yielded ancient plant DNA sequences. We find that, when frozen, chemical factors that have formerly shown to increase the DNA degradation rate (non neutral pH and oxidation) are not of critical importance for DNA survival over a time scale of at least 400.000 years. In contrast, soil geochemical characteristics, such as C and N levels, and in particular stable isotope signatures that indicate whether sediment inclusions are of terrestrial origin appear indicative of plant DNA survival. We suggest the latter of these be performed prior to larger scale destructive sampling for future environmental ancient plant DNA studies from such samples.**

## Introduction

The composition of past plant and animal communities has been reconstructed from diverse environmental ancient plant and animal DNA sequences, recovered directly from sediments and basal glacial ice dating back up to *c.* ½ million years (½ Ma BP) (Hofreiter, et al., 2003, Willerslev, 2003, Lydolph, et al., 2005, Willerslev, et al., 2007, Haile, et al., 2009, Hebsgaard, et al., 2009, Thomsen, et al., 2009, Anderson-Carpenter, et al., 2011). Basal (sediment rich) ice in particular has revealed remarkable results, such as evidence for diverse conifer forest in southern central Greenland some 450 to 800 thousand years ago (Ka BP), almost 2 million years later than previously thought (Willerslev et al. 2007) – a result later confirmed by traditional pollen records (de Vernal & Hillaire-Marcel, 2008). However, some basal ice samples do not yield PCR amplifiable DNA. It has been previously assumed that this relates to their deep age or to chemical conditions in the ice that are unfavourable to long-term DNA survival (Willerslev, et al., 2004).

It is well established that in the absence of cell repair, DNA degrades over time due to spontaneous chemical reactions such as hydrolysis, oxidation, and molecular crosslinking (Lindahl, 1993, Poinar, et al., 1996, Hansen, et al., 2006, Heyn, et al., 2010). In sediments the speed of degradation is determined by factors such as temperature, pH, redox potential, cation content, water content and pressure (Mitchell, et al., 2005, Cai, et al., 2006, Pietramellara, et al., 2009, Andersen, et al., 2011). The factors not only directly affect the speed of spontaneous decay but also influence the ability of the DNA to bind to minerals that causes protection, by condensing the helix shape and making the molecule less suitable to nucleases (Pietramellara, et al., 2009). Furthermore, soil types that hold particles with large surface areas correlate with higher amounts of adsorbed DNA

68 compared to soils with larger particles that holds smaller surface areas and less DNA
69 (Lloyd-Jones & Hunter, 2001).
70
71 So far attempts at environmental reconstruction using DNA extracted from basal ice has
72 essentially been a matter of trial and error, that sometimes results in fruitless destruction
73 of up to several litres of rare and precious ice core samples (Willerslev et al. 2004). In this
74 study we investigate the chemical and physical properties of basal ice core samples that
75 have previously been characterised for PCR amplifiable ancient plant DNA content (that
76 is, either do, or do not yield, amplicons). The aim is to determine factors that appear
77 crucial for recovery of PCR amplifiable environmental plant DNA within the basal ice,
78 providing proxies that can be used to screen samples prior to future attempts at
79 palaeocommunity reconstructions.
80
81 **Methods**
82 Basal ice core samples from four different drillings in Greenland were used for
83 comparison: GRIP, NGRIP, Dye 3 and Camp Century (Fig. 1; Table 1), each has
84 previously been reproducibly tested for ancient plant DNA under strict criteria of
85 authenticity (Willerslev, et al., 1999, Willerslev, 2004, Willerslev, et al., 2007).
86 Measurements of pH and redox were made under anoxic conditions within one hour after
87 thawing at 0°C in a continuous flow of argon. The redox potential was measured using a
88 gel redox-electrode connected to an EUTECH WP 600. Measurements of pH were made
89 using a calomel pH-electrode connected to a Metrohm 744 pH meter. Subsequently, melt
90 water was collected after centrifuging and analysed for $NO_3$, $NH_4$ and total dissolved
91 nitrogen (using a FIAstar 5000 Flow Analyzer), total dissolved carbon (using a Shimadzu
92 Total Organic Carbon Analyzer TOC-5000A), Cl, $SO_4$ (using a high pressure liquid

93 chromatography as well as Ca, Mg, K, Na, Mn, Al and Fe (using atomic absorption
94 spectroscopy, ASS) after 0.45 μm filtration.
95
96 Particle analysis included total carbon and nitrogen as well as $\delta^{15}N$ and $\delta^{13}C$ (measured
97 using an Eurovector elemental analyzer coupled to a Micromass IsoPrime Isotope Ratio
98 Mass Spectrometre), and measurement of the grain size distribution (using a Malvern
99 Mastersizer). The specific surface area (BET) was measured using a Micromeritics
100 Gemini 2365). Non-silicate-bounded Mn, Al and Fe were extracted using citrate-
101 bicarbonate-dithionite treatment (CBD). Subsequently, concentrations of Mn, Al, Fe, Ca,
102 Mg, Na and K were analysed using atomic absorption spectroscopy (ASS) after 0.45μm
103 filtration.
104
105 **Results**
106 The four basal ice core samples represent contrasting sedimentary environments and ages,
107 with pH ranging from 8.3 to 9.2, redox potentials from -8 to 113 $E_H$ (mV), bottom
108 temperatures from 0ºC to -13ºC and ages ranging from 120 Ka to more than 450 Ka BP
109 (Table 1):
110
111 The GRIP samples have not yielded any plant ancient DNA sequences despite extensive
112 previous PCR attempts in *Taq* inhibitor free conditions (data not shown). The basal ice
113 samples investigated have tentative ages dated to 400-950 Ka BP (Muscheler & Beer,
114 personal communication, Willerslev et al., 2007). They consist largely of fine sand with a
115 small specific surface area, a low C content (0.2%) and a C:N ratio suggesting relatively
116 high N content. The degree of sediment sorting is low and the availability of Fe, Mn and
117 Al after extraction (not silicate bounded metal) is higher than expected from non-

weathered sediments. The source material is uncertain but could be aeolian (wind) deposits without indications of representing a buried near-surface soil surface. The availability of metals indicates ongoing weathering – possible exposure prior to burial. The low C:N ratio may suggest a cyanobacterial input of C to the site rather than from plants. The results are partly in line with previous analyses of stable isotopes, carbon dioxide, methane and ammonium in the silty ice, which was thought to point to a deglaciated area with significant biological activity at the time of ice formation (Tison, et al., 1998, Souchez, 2000).

The Dye3 samples have produced diverse ancient plant DNA sequences and have been dated to approximately 450 – 800 Ka BP (Willerslev, et al., 2007). They have a limited content of sediment and consist largely of clay (< 2 μm). The C content was 0.1% and the C:N ratio about 14 representing a range observed in sediments with low amounts of plant-derived C-sources but with relatively high N. The availability of Fe, Mn and Al after CBD extraction is low suggesting limited *in-situ* weathering prior to burial and in line with almost no exchangeable cations. The site could represent an environment with deposited glacial-derived clay size particles after strongly physical weathering of previous near-surface layers with a C source from plants ecosystem.

The Camp Century samples have produced diverse ancient plant DNA sequences and provide a tentative age estimate of at least 0.18 Ma BP (Johnsen, et al., 2001, Otto-Bliesner, et al., 2006, Aciego, S., unpublished results). The samples has a C content of 0.6% and a C:N ratio of 24 which is typically observed in soil profiles with partly decomposition of organic matter. The availability of Fe, Mn and Al after CBD extraction is similar to that of GRIP and suggests some *in-situ* weathering prior to burial being

consistent with higher exchangeable cations (mainly Ca) as compared to Dye3. Camp Century could represent an environment with deposited soil-derived sediments from near-surface layers with a C source derived from an ecosystem with vascular plants.

The NGRIP may contain putative ancient plant DNA sequences (the basal ice was contaminated heavily with drill fluid so its hard distinguishing putative and contaminate DNA). The site is dated to approximately 0.12 Ma BP (Andersen, et al., 2004). The basal ice consists of two sediment fractions: a clay fraction and a very fine sand fraction. The C content is 2.9% and the C:N ratio about 16. The site could represent an environment similar to Dye 3 with deposited glacial-derived clay after strongly physical weathering of previous near-surface layers mixed with wind-derived material. If the C content is not influenced by contamination, the C content suggests almost *in-situ* preservation of near-surface sediments from an ecosystem with relatively high organic matter input or accumulation.

The combination of C and N levels and stable isotope signatures suggest a terrestrial plant-associated source of C in the three sites except for GRIP where cyanobacteria, as suggested by the low C/N ratio, could be a more dominant C source as compared to plants. One of the replicates from Camp Century show a high signal of $\delta^{13}C$, which could be due to a high content of inorganic carbonates and/or due to a marine signal. All four samples reveal similar and alkaline pH-values suggesting equilibrium with carbonate (Table 1). Redox measurements indicate that all four samples have been partly oxidized, with Camp Century having the highest redox potential. If not a result of contamination, NGRIP is the only site, which may have been almost fully oxidized. The bottom temperatures varied between sites from -8ºC to -13ºC at the GRIP, Dye 3 and Camp

168 Century sites with NGRIP being the only site that experienced basal melting due to local
169 geothermal conditions (Table 1).
170
171 **Discussion and Conclusion**
172 It has been assumed that successful retrieval of putative ancient DNA from basal ice is
173 closely linked to sample age and chemical conditions including pH, oxidation, and cations
174 (Willerslev, et al. 2004; Willerslev, et al. 2007). The Dye 3 and Camp Century samples
175 containing PCR amplifiable ancient plant DNA represent relatively high pH and oxidation
176 levels (Table 1). This suggests that long-term DNA survival may not be as sensitive to
177 oxidative and alkaline damage in cold temperatures as previous thought, likely because
178 low temperatures decrease the rate of chemical reactions. Such conditions are otherwise
179 known to modify bases (Lindahl, 1993, Höss, et al., 1996), create inter-strand and single
180 strand cross-links (Poinar, et al. 1996; Hansen, et al. 2006), disrupt hydrogen bonding
181 (Ageno, 1969) and increase depurination rates (Lindahl & Nyberg, 1972), ultimately
182 causing single strand breaks (Mitchell, et al., 2005). Similarly, cations are known to have
183 an advantageous effect on adsorbing DNA to soil particles and reduce the rate of damage
184 (Cai, et al., 2006, Pietramellara, et al., 2009). However, the Dye 3 and Camp Century
185 samples, that both show presence of PCR amplifiable plant ancient DNA have a ten-fold
186 difference in cation content. As for pH and redox, we may also conclude that under frozen
187 conditions cation content appears of limited importance for DNA long-term survival. In
188 fact, the GRIP sample that does not produce positive DNA findings appears to contain pH
189 and cation levels that would otherwise be viewed as more favourable for preservation
190 than the Dye3 sample, and similarly has oxidation levels that would normally be viewed
191 as more favourable than the DNA positive Camp Century sample (Table 1). Also the
192 oldest of the GRIP samples is in the age range that is well within the predicted maximal

range of long-term DNA survival under frozen conditions (Willerslev, et al., 2004, Willerslev, et al., 2004b). Interestingly, the difference between the samples that yielded positive DNA results, Dye 3 and Camp Century, *versus* the sample that did not, GRIP, relates to the complex combination of factors like summed cation content, total carbon and nitrogen, C:N ratio and $\delta^{13}C$ within each sample, providing information on the history of the site formation (Table 1): NGRIP, Camp Century and Dye 3 all show chemical values supportive of a rich past terrestrial ecosystem. The GRIP sample, in contrast, has a lower C/N ratio and appears to have undergone weathering as it consists mainly of fine sand with a small specific surface area and judged from the Fe, Mn, and Al availability. When combined with the stable isotope signature, these factors do not support a past existence of a vegetated plant ecosystem.

Thus, our study allows us to draw four main conclusions: i) Common among the sites that contains DNA is the fact that temperatures have remained well below freezing and have stayed so since incorporation. ii) As long as temperature stays low, pH, oxidation, and cations otherwise affecting long-term DNA survival, appear of less importance. iii) Age does not seem the limiting factor for long-term DNA survival in a frozen environment within the time span investigated here (hundred of thousands of years). iv) Rather the source of the sediments is a key factor. The latter can by investigated by soil geochemical and physical characteristics. Such analyses require approximately 50% of the ice volume required for DNA-based palaeoenvironmental studies, and more importantly this ice can be taken from the contaminated surface of the ice core, that is of no use for most other chemical/physical analyses that are carried out on ice cores (e.g. for glaciological analyses). These may likely be useful tests for the screening of promising basal ice samples prior to larger-scale destructive sampling for DNA studies.

We are aware that our study is based on a limited number of samples. Given the expense and high value of deep ice core samples, we encourage groups having access and studying ice core samples from other parts of the world to conducting similar test. Not only may it provide future solid ground for ice core sampling for ancient DNA studies, it will also make it easier getting access to the rare and precious material.

**Acknowledgements**

We thank the Danish National Research Foundation for financial support.

225 **Table 1.** Physical and chemical characteristics of the studied basal ice samples from
226 the deep drilling sites, GRIP, NGRIP, Camp Century and Dye 3 in Greenland

| | GRIP | Dye 3 | Camp Century | NGRIP [1] |
|---|---|---|---|---|
| Presence of plant DNA | ÷ | + | + | nd |
| Depth (m) | 3024 - 3028 | 2011 - 2129 | 1353 - 1365 | 3047 - < 3085 |
| Age | 400-950 ky | 450 – 800 ky | >180 ky | 123 ky |
| Basal temperature (ºC) | -8.6 | -13.2 | -12 | 0 |
| pH ± 1 standard dev, n=3 | 8.4 ± 0.2 | 8.6 ± 0.3 | 8.3 ± 0.1 | 9.2 ± 0.1 |
| Redox ± 1 standard dev, n=3 | 51 ± 15 | 52 ± 12 | 113 ± 9 | -8 ± 3 |
| Total C (% ) | 0.2 ± 0.03 | 0.1 ± nd | 0.6 ± 0.2 | 2.9 ± 0.3 |
| Total N (%) | 0.02 ± 0.002 | 0.01 ± nd | 0.03 ±0.021 | 0.19 ± 0.007 |
| C:N | 7.0 | 14.1 | 23.5 | 15.9 |
| $\delta^{13}C$, $\delta^{15}N$ (‰) | -26.5; 8.4 | -25.9; 2.7 | -22.2; 9.8 | -28.8; 6.0 |
| CBD (Fe; Al; Mn), mg $kg^{-1}$ | 4966, 374, 44 | 2793, 209, 29 | 5758, 552, 51 | 12346, 1436, 22 |
| ∑ exchangeable cations, cmol $kg^{-1}$ | 4.6 | 0.7 | 7 | 50 |
| $D_{50}$, µm | 151 | nd | 59 | 7 |
| BET, $m^2$ $g^{-1}$ | 0.4 | nd | 1 | 3 |

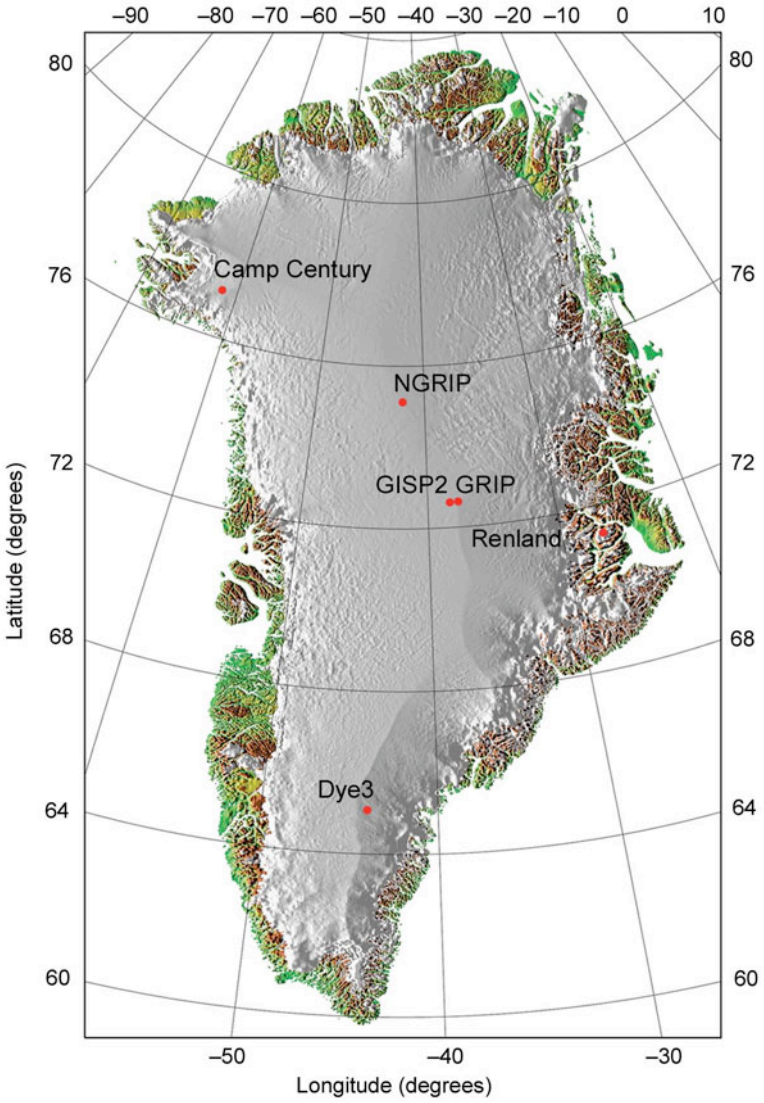
227

228 [1]Refrozen basal meltwater, nd= not determined, $D_{50}$ is a measure of the grain size
229 distribution and BET is a measure of the specific surface area of the soil particles
230 Green: plant DNA presence as shown from deep-sequencing of trnL/rbcL short
231 minibarcodes (Willerslev, et al., 2007); Blue: Physical characteristics; Yellow:
232 Chemical characterization.

233

234 **Figure 1**.

235



## References


Ageno M (1969) The alkaline denaturation of DNA. *Biophysical journal* **9**: 1281.

Andersen K, Bird K-L, Rasmussen M, et al. (2011) Meta-Barcoding of 'Dirt' DNA from Soil Reflects Vertebrate Biodiversity. *Molecular Ecology*. In Press.

Andersen KK, Azuma N, Barnola JM, et al. (2004) High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**: 147-151.

Anderson-Carpenter LL, McLachlan JS, Jackson ST, Kuch M, Lumibao CY & Poinar HN (2011) Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC Evol Biol* **11**: 30.

Cai P, Huang QY & Zhang XW (2006) Interactions of DNA with clay minerals and soil colloidal particles and protection against degradation by DNase. *Environmental Science & Technology* **40**: 2971-2976.

Cai P, Huang Q, Jiang D, Rong X & Liang W (2006) Microcalorimetric studies on the adsorption of DNA by soil colloidal particles. *Colloids and surfaces. B, Biointerfaces* **49**: 49-54.

de Vernal A & Hillaire-Marcel C (2008) Natural Variability of Greenland Climate, Vegetation, and Ice Volume During the Past Million Years. *Science* **320**: 1622-1625.

Haile J, Froese DG, MacPhee RDE, et al. (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 22352-22357.

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.

Hebsgaard MB, Gilbert MTP, Arneborg J, et al. (2009) 'The Farm Beneath the Sand'- an archaeological case study on ancient 'dirt' DNA. *Antiquity* **83**: 430-444.

Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M & Meyer M (2010) Road blocks on paleogenomes - polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Research* **38**: e161.

Hofreiter M, Mead JI, Martin P & Poinar HN (2003) Molecular caving. *Current Biology* **13**: R693-R695.

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M & Paabo S (1996) DNA Damage and DNA Sequence Retrieval from Ancient Tissues. *Nucleic Acids Research* **24**: 1304-1307.

Johnsen SJ, Dahl-Jensen D, Gundestrup N, et al. (2001) Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-

3, GRIP, GISP2, Renland and NorthGRIP. *Journal of Quaternary Science* **16**: 299-307.

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

Lindahl T & Nyberg B (1972) Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **11**: 3610-3618.

Lloyd-Jones G & Hunter DWF (2001) Comparison of rapid DNA extraction methods applied to contrasting New Zealand soils. *Soil Biology & Biochemistry* **33**: 2053-2059.

Lydolph MC, Jacobsen J, Arctander P, et al. (2005) Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and Environmental Microbiology* **71**: 1012-1017.

Mitchell D, Willerslev E & Hansen A (2005) Damage and repair of ancient DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**: 265-276.

Otto-Bliesner BL, Marsha SJ, Overpeck JT, Miller GH, Hu AX & Mem CLIP (2006) Simulating arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**: 1751-1753.

310 Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G & Nannipieri P
311 (2009) Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology*
312 *and Fertility of Soils* **45**: 219-235.
313
314 Poinar HN, Hoss M, Bada JL & Paabo S (1996) Amino acid racemization and the
315 preservation of ancient DNA. *Science* **272**: 864-866.
316
317 Souchez R (2000) Basal ice formation and deformation in central Greenland: a review
318 of existing and new ice core data. *Special publication - Geological Society of London*
319 **176**: 13.
320
321 Thomsen PF, Elias S, Gilbert MTP, et al. (2009) Non-Destructive Sampling of
322 Ancient Insect DNA. *PLoS ONE* **4**.
323
324 Tison J-L, Souchez R, Wolff EW, Moore JC, Legrand MR & de Angelis M (1998) Is
325 a periglacial biota responsible for enhanced dielectric response in basal ice from the
326 Greenland Ice Core Project ice core? *J. Geophys. Res.* **103**: 18885-18894.
327
328 Willerslev E (2003) Diverse Plant and Animal Genetic Records from Holocene and
329 Pleistocene Sediments. *Science* **300**: 791-795.
330
331 Willerslev E (2004) Isolation of nucleic acids and cultures from fossil ice and
332 permafrost. *Trends in Ecology & Evolution* **19**: 141-147.
333

334 Willerslev E, Hansen AJ & Poinar HN (2004) Isolation of nucleic acids and cultures
335 from fossil ice and permafrost. *Trends in Ecology & Evolution* **19**: 141-147.
336
337 Willerslev E, Hansen AJ, Christensen B, Steffensen JP & Arctander P (1999)
338 Diversity of Holocene life forms in fossil glacier ice. *Proceedings of the National*
339 *Academy of Sciences of the United States of America* **96**: 8017-8021.
340
341 Willerslev E, Hansen AJ, Ronn R, et al. (2004b) Long-term persistence of bacterial
342 DNA. *Current Biology* **14**: R9-R10.
343
344 Willerslev E, Cappellini E, Boomsma W, et al. (2007) Ancient Biomolecules from
345 Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**: 111-114.
346
347
348 **Figure 1**. Map showing deep drilling sites in Greenland. Samples have been analysed
349 from Camp Century, NGRIP, GRIP and Dye 3.
350

# Chapter III

METHODS

# Methods: Sampling, laboratory work and data processing

In this chapter, I will describe the different methods that have been applied to generate and analyse the DNA sequence data retrieved from ice and permafrost samples from Greenland and Antarctica. The chapter organisation will be as follows. First, I will introduce the strategies used for sample retrieval, sample processing and laboratory procedures that were performed to extract ancient DNA from these samples. Then, I will describe how sequence data generated on the Roche GS-FLX platform was analysed employing three different approaches, and lastly, I will describe how the sequence data was assigned to taxonomic groups also using three approaches. The results will be discussed in the following chapter.

## SAMPLING

### Sampling of Ice Cores:

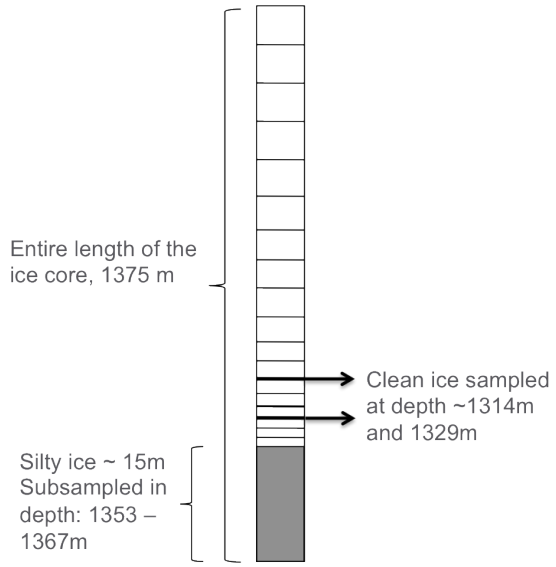
Samples were drilled using an electromechanical drill in 1966 at the Camp Century site on the Greenland Ice Sheet (GIS) at the location 77°10'N 61°08'W, and at 1885m above sea level. The core basal debris spanned 15.7m out of the entire length of 1375m (Herron & Langway, 1978). The bottom temperature in the drill hole was measured as -13°C. All samples have since been stored in cold rooms at -22°C. The basal ice sampled in this study was sub-sampled from depth 1353m to 1367m and covered the entire length of the silty section that is still available (see Figure 1). Clean ice samples have been sampled above the silty ice from depth 1314.65m – 1315.2m and from 1329.78m – 1330.33m to test if the DNA was of non-local origin and could have been wind born from an external source.

Cracks can occur during the drilling process whereby microbial or other modern DNA sources can contaminate the sample during handling or storing. It is therefore important to check for possible exogenous contamination to avoid false positive results. A recognizable bacterial vector suspended in 70% ethanol (TOPO® TA cloning Kit, ~1.7ng/µl) was applied to the surface of the samples as in (Willerslev, et al., 2007) to test for contamination during sampling or handling.

Microtome knives cleaned in a 20% bleach solution and UV irradiated in a UV-cross linker (UVP, CL-1000 model) at 254 nm for 20 minutes were used to scrape off the

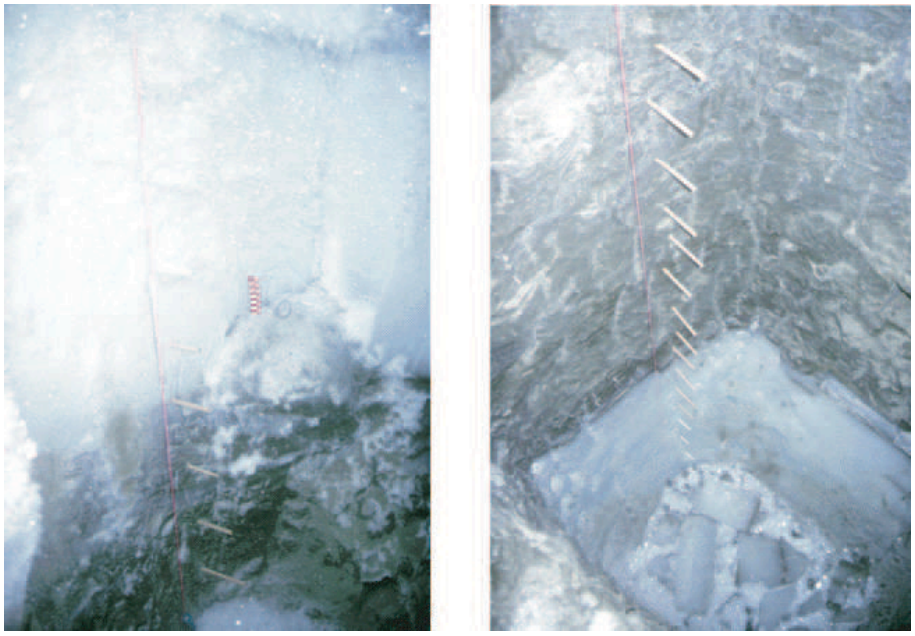
outer surface of the samples, removing the exposed layers. For each ½ cm removed the knife was exchanged for a clean one and the scrapes were collected to test for the presence of the vector. Approximately 2 cm of the surface was removed in a cold room at -18°C under a flowhood. The amount removed was approximately between 60 – 70% of the original weight of the sample.
Samples are listed in table 1.

**Figure 1**. Illustration of sampling positions from the Camp Century ice core.




**Sampling of Glacier Ice**:
Basal ice samples from the Taylor and Suess Glaciers in Dry Valleys of Antarctica (77°44′S 162°10′E and 77°38′S 162°40′E, respectively) were sampled in the austral summers in 1996/1997 to 2000/2001 and have since been kept in cold rooms at temperatures below -22°C. Samples were cut out from tunnels dug into the apron of the glaciers, approximately 10m and 25m respectively from the ice cliff at the two glaciers, with chain saws and then sampled from a vertical shaft within the apron (Samyn, et al., 2005b) (see Figure 2). The temperature at both places was -17°C. As above, clean samples were included to test for non-local origin of the DNA and were sampled above the basal ice in younger glacier ice but below the surface (see Table 2).

**Figure 2**. Illustration of the sampling strategy in the vertical shaft at Taylor glacier. Top (left) and bottom (right) view of the clean overlying ice and the lower debris rich ice, from (Samyn, 2005a).



**Sampling of Permafrost**:

Kap København:

Sediment samples were obtained from the Kap København formation in Northern Greenland (82° 24' 00" N 22° 12' 00" W) in the summer of 2006 (see Table 3). Sample material consisted of organic rich permafrost and dry permafrost. The samples were collected vertically from the slope of the hills either with a mechanical drill or directly into DNA-free 50mL falcon tubes. Disposable tools like gloves and scalpels were used and changed between samples to avoid cross-contamination. For DNA extractions, the material was further sub-sampled from the inner part of sediment cores, leaving 1.5 - 2 cm between the inner core and the surface that provided a subsample of approximately 6-10g. After sampling the samples were stored at temperatures below -22°C.

## SAMPLE DESCRIPTIONS

**Antarctica & Greenland**;

Basal ice from Taylor Glacier contained thick debris layers that consisted mainly of coarse grained particles having a small surface to volume ratio of 10.71 $m^2/g$ (Aciego, unpublished)

Basal ice from Suess Glacier contained a low concentration of very fine-grained red coloured silt, also termed amber ice. Amber ice has previously been described as containing a relatively high concentration of silt to sand particles with a relatively uniform distribution of all particles within the ice (Samyn, et al., 2005c). From the samples included in this study we found a large variety in the sizes of the entrained particles, although most of the particles were smaller than 30μm with a large volume to surface ratio of 67.45 $m^2/g$ (Aciego, unpublished)

The debris concentration in the Antarctic glacier samples varied considerably between sites, from 0.9% by volume in the Suess amber samples to 30 – 50% by volume in the Taylor debris-rich samples (Samyn, et al., 2005b, Mager, et al., 2009).

Basal ice from Camp Century has previously been reported to contain a debris concentration of 0.24% by weight (Herron and Langway 1979), which is almost the same average concentration found on the subset of samples in this study, measured to ~0.26% by weight. The basal samples were dominated by fine-grained material varying in particle size from 24.92 to 36.59 $m^2/g$, but did also contain a minor size fraction with a ratio of 14.01 - 15.66 $m^2/g$.

Permafrost samples from the Kap København formation were sampled from a sequence of unconsolidated marine sand and silt (Funder, 1989). The particle size in these samples varied from 0.04 – 0.24 $m^2/g$ with a total organic carbon content between 0.66 – 1.07%.

**Table 1**. List of Ice core samples from the Camp Century site, NW Greenland

| Sample Name/ #Tube no | Site Name | Location | Type | Extraction no. | Weight of silt (g) | Amount after scraping (ice + debris) (g) | Debris concentration (%) by weight | Depth (m) | Status |
|---|---|---|---|---|---|---|---|---|---|
| # TUBE 1049 A | Camp Century | Greenland | Basal Ice | 2-1 | 0.3 | 176.27 | 0.170 | 1352.75 - 1353.3 | contaminated |
| # TUBE 1049 A | Camp Century | Greenland | Basal Ice | 3-1 | 0.25 | 146.89 | 0.170 | 1352.75 - 1353.3 | + |
| # TUBE 1049 A | | | Basal Ice | 9-1 | 0.2 | 115.91 | 0.173 | 1352.75 - 1353.3 | + |
| # TUBE 1049 A | Camp Century | Greenland | Basal Ice | replication | 0.25 | 145.72 | 0.172 | 1352.75 - 1353.3 | + |
| # TUBE 1049 B | Camp Century | Greenland | Basal Ice | replication | 0.60 (incl drop of water) | 280.53 | 0.125 | 1353.3- 1353.85 | on hold |
| # TUBE 1050 A | Camp Century | Greenland | Basal Ice | 2-3 | 0.23 | 130.25 | 0.177 | 1354.25 - 1354.8 | + |
| # TUBE 1050 A | Camp Century | Greenland | Basal Ice | replication | 1 | 440.43 | 0.227 | 1354.25 - 1354.8 | on hold |
| # TUBE 1052 A | Camp Century | Greenland | Basal Ice | 2-5 | 0.55 | 157.74 | 0.349 | 1357.17 - 1357.72 | contaminated |
| # TUBE 1052 A | Camp Century | Greenland | Basal Ice | 3-2 | 0.24 | 70.43 | 0.341 | 1357.17 - 1357.72 | + |
| # TUBE 1052 A | Camp Century | Greenland | Basal Ice | replication | 0.35 (incl drop of water) | 100.44 | 0.199 | 1357.17 - 1357.72 | + |
| # TUBE 1053 A | Camp Century | Greenland | Basal Ice | 2-7 | 0.25 | 125.6 | 0.199 | 1358.59 - 1359.14 | contaminated |
| # TUBE 1053 B | Camp Century | Greenland | Basal Ice | 8-1 | 0.21 | 101.27 | 0.207 | 1359.14 - 1359.69 | + |
| # TUBE 1053 B | Camp Century | Greenland | Basal Ice | 2-7 | 0.25 | 120.71 | 0.207 | 1359.14 - 1359.69 | contaminated |
| # TUBE 1053 B | Camp Century | Greenland | Basal Ice | replication | 0.4 | 188.81 | 0.212 | 1359.14 - 1359.69 | on hold |
| # TUBE 1054 A | Camp Century | Greenland | Basal Ice | 6-1 | 0.35 | 154.23 | 0.227 | 1360.04 - 1360.79 | + |
| # TUBE 1054 B | Camp Century | Greenland | Basal Ice | 1-2 | 0.32 | 138.27 | 0.231 | 1360.79 - 1361.14 | + |
| # TUBE 1054 B | Camp Century | Greenland | Basal Ice | replication | 0.95 | 380.65 | 0.250 | 1360.79 - 1361.14 | + |
| # TUBE 1054 C | Camp Century | Greenland | Basal Ice | 6-2 | 1.01 | 112.3 | 0.899 | 1361.14 - 1361.5 | - |
| # TUBE 1054 C | Camp Century | Greenland | Basal Ice | replication | 0.37 | 80.86 | 0.458 | 1361.14 - 1361.5 | on hold |
| # TUBE 1055 B | Camp Century | Greenland | Basal Ice | 9-2 | 0.5 | 142.63 | 0.351 | 1362.05 - 1362.6 | + |
| # TUBE 1055 B | Camp Century | Greenland | Basal Ice | replication | 0.61 | 170.78 | 0.357 | 1362.05 - 1362.6 | on hold |
| # TUBE 1055 D | Camp Century | Greenland | Basal Ice | 6-4 | 0.42 | 115.99 | 0.362 | 1362.6 - 1362.98 | - |
| # TUBE 1056 B | Camp Century | Greenland | Basal Ice | 6-5 | 0.40 | 161.49 | 0.248 | 1362.98 - 1363.53 | + |
| # TUBE 1057 A | Camp Century | Greenland | Basal Ice | 1-4 | 0.64 (incl drop of water) | 149.4 | 0.335 | 1364.43 - 1364.98 | + |
| # TUBE 1057 A | Camp Century | Greenland | Basal Ice | | 0.3 | 121.23 | 0.247 | 1364.43 - 1364.98 | + |
| # TUBE 1023 A | Camp Century | Greenland | Clean ice | clean 1 | filtered | ~1 kg | ND | 1314.65 - 1315.2 | - |
| # TUBE 1033 B | Camp Century | Greenland | Clean ice | clean 2 | filtered | ~1.5 kg | ND | 1329.78 – 1330.33 | - |

Status refers to whether there was positive (+) or negative (-) amplification, or if the vector DNA had passed through the surface to the inner core e.g. through cracks (contaminated) in which case the sample was not used for further DNA analyses.
Samples marked in yellow were send for replication, sampled on hold were extracted but remains to be amplified.
Samples marked with blue were glacial clean ice and used as controls for dispersal based origin of the DNA.

**Table 2**. List of glacier ice samples from the Dry Valleys, Antarctica.

| Sample Name | Site Name | Location | Type | Extraction no. | Weight of silt (g) | Amount after scraping (ice + debris) (g) | Debris concentration (%) by weight | Status | |
|---|---|---|---|---|---|---|---|---|---|
| Suess 98/1 | Suess Glacier, Taylor Valley | Dry Valleys, Antarctica | | 4-3 | 0.1 | 151.11 | 0.066 | | - |
| Clean ice Suess 98#1 | Suess Glacier, Taylor Valley | Dry Valleys, Antarctica | Clear ice with no visible particles | clean 10-1 | | 120.34 | ND | | - |
| Suess Glacier: Su 98#3 | Suess Glacier | Dry Valleys, Antarctica | Amber ice | 2-9 | 0.31 | 156.27 | 0.198 | | + |
| Suess Glacier: Su 98#3 | Suess Glacier, Taylor Valley | Dry Valleys, Antarctica | Amber ice | 3-3 | 0.2 | 102.12 | 0.196 | | + |
| Clean ice Suess 98#3 | Suess Glacier, Taylor Valley | Dry Valleys, Antarctica | Amber ice | replication | 0.25 | 128.58 | 0.194 | | + |
| Taylor Glacier (tunnel): Taylor TG-6-B | Taylor Glacier, Taylor Valley | Dry Valleys, Antarctica | clear ice with black material layer | 2-11 | 18.39 | 54.43 | 33.787 | | - |
| Taylor Glacier (tunnel): Taylor TG-6-B | Taylor Glacier, Taylor Valley | Dry Valleys, Antarctica | clear ice with black material layer | 5-1 | 7 | 53 | 13.208 | | - |
| Taylor Glacier (tunnel): Taylor TG-6-B | Taylor Glacier, Taylor Valley | Dry Valleys, Antarctica | clear ice with black material layer | replication | 4.33 | 53 | 8.170 | | - |
| TGTA from Tunnel dug | Taylor Glacier, Taylor Valley | Dry Valleys, Antarctica | almost clear | 4-5 | 0.25 | 141.73 | 0.176 | | - |
| Clean ice TGTA from Tunnel dug | Taylor Glacier, Taylor Valley | Dry Valleys, Antarctica | Clear ice with no visible particles | clean 10-2 | | 135.71 | ND | | - |

Status refers to whether there was positive (+) or negative (-) amplification, or if the vector DNA had passed through the surface to the inner core e.g. through cracks (contaminated) in which case the sample was not used for further DNA analyses.
Samples marked in yellow were send for replication.
Samples marked with blue were glacial clean ice and used as controls for dispersal based origin of the DNA.

**Table 3**. Sample list from the Kap København formation in Northern Greenland.
Status refers to whether there was positive (+) or negative (-) amplification, or if the vector DNA had passed through the surface to the inner core e.g. through cracks (contaminated) in which case the sample was not used for further DNA analyses.

| Sample Name/ extraction no | Site Name | Location | Type | Weight of sample for DNA extraction | Status |
|---|---|---|---|---|---|
| KK221/EH1122 | Kap København | North Greenland | permafrost | 10g | + |
| KK239/EH1121 | Kap København | North Greenland | permafrost | 10g | + |
| KK244/EH1123 | Kap København | North Greenland | permafrost | 10g | + |
| KK271/EH1120 | Kap København | North Greenland | permafrost | 10g | + |
| KK2121/EH1118 | Kap København | North Greenland | permafrost | 10g | + |
| KK2108/EH1119 | Kap København | North Greenland | permafrost | 10g | - |
| 198B/2033 | Kap København, Loc 50 | North Greenland | permafrost | 10g | + |
| 201A/2038 | Kap København, Loc 75 | North Greenland | permafrost | 10g | + |
| 201B/2039 | Kap København, Loc 75 | North Greenland | permafrost | 10g | - |
| 203A/2042 | Kap København, Loc 75 | North Greenland | permafrost | 10g | - |
| 203B/2043 | Kap København, Loc 75 | North Greenland | permafrost | 10g | + |
| 204A/2044 | Kap København, Loc 75 | North Greenland | permafrost | 10g | - |

## LABORATORY WORK

### DECONTAMINATION

All of the equipment and reagents used for DNA extractions were cleaned before use by ultra-filtration (30-50 kilodalton (kDa) cutoff), bleach (5%) or DNAway, acid (2.5M HCl, 48h) and UV treatment at 254 nm for 20 minutes (UVP cross-linker, CL-1000 model). Moreover, the aDNA lab accords to the standards for such facilities by having positive air pressure, nightly λ=254nm UV- irradiation of surfaces, being regularly cleaned with bleach and separated physically from other modern biological laboratories. During lab work, full bodysuits, facemasks, and disposable gloves were used and daily movement was always from ancient to modern DNA laboratories (Willerslev, 2004a).

#### Ice Samples

Following the decontamination procedure with application of vector DNA and removal of outer surface layers, we performed PCR with vector-specific primers (T3/T7 universal primers, Invitrogen) to screen for the vector in both the outer layers and interior of the samples. PCR was performed in 25μL reactions containing 2μL of DNA extract, 0.2mM of each dNTP, 0.4μM of each primer, 5mM $MgSO_4$ and 1.0U Platinum® High Fidelity (HiFi) DNA polymerase (Invitrogen, Carlsbad, CA) with 1X High Fidelity buffer. Cycling conditions were 2' at 94$^{o}$C, 45 cycles of (20'' at 94$^{o}$C, 20'' at 53 $^{o}$C, 20'' at 68$^{o}$C) and a final extension at 72$^{o}$C for 10'.

### DNA EXTRACTION

#### Ice Samples

Camp Century, Greenland:

Following decontamination, the samples varied in weight from 70g to 440g. The silt content was weighed after the samples were melted and varied from 0.17g – 0.45g per 100g (see Table 1).

Dry Valleys, Antarctica:

Basal amber ice from Suess Glacier and basal debris-containing ice from Taylor Glacier were decontaminated and melted. The silt content varied from 33.7g per 100g in the Taylor sample to 0.2g per 100g in the Suess sample (see Table 2).

Melting procedure:

Basal ice

The samples were transferred to dedicated aDNA lab facilities and processed through stages of melting, centrifugation and filtering before extraction.

After melting, the samples were transferred to 50mL Nalgene® tubes, centrifuged at 14.000 g for 30 minutes, concentrating the silt and collecting the supernatant for filtering. This was repeated until the entire melted sample had been centrifuged.

Clean ice:

The clean ice samples were filtered on 30kDa Millipore Centricon filters and concentrated to 250 μL prior to extraction.

DNA Extraction Protocol:

A modified version of the protocol from (Bulat, et al., 2000) was used for DNA extraction. Up to 0.25g of silt or the concentrated clean ice was transferred to FastPrep® Lysing Matrix E tubes for DNA extraction and dissolved in 600μl enzymatic buffer. The enzymatic lysing buffer consisted of 50 mM Tris-HCl (pH 8), 50 mM EDTA, 150 mM NaCl, 2.5% N-lauroyl sarcosine, 500 mM 2-mercaptoethanol, 5mM DTT and 400 ug/ml proteinase K (Roche). The mixture of sample and lysis buffer was vigorously agitated through four runs on a FastPrep 120 (BIO 101) at speed 6.5 for 45 sec, with 2 minutes on ice between runs. The solution was supplemented with proteinase K (400 ug/ml) (Roche) and then left for gentle agitation at 55$^{o}$C overnight. 1M NaCl was added to the solution together with 1/2 of the total volume of chloroform/octanol (24:1) then left for slow agitation for 1 hour at room temperature. The water phase was isolated through centrifugation at 12,000g for 2 min. and transferred to a separate microtube for incubation at 2-3°C overnight. The precipitate was centrifuged at 12,000g for 2 min and the supernatant purified using silica spin columns and PB-buffer (Qiagen DNA Qiaquick PCR purification kit),

followed by washes in 0.5 ml Salton wash 1 and 2 (BIO 101), and 0.5 ml AW 1 and 2 (Qiagen tissue kit). The DNA was eluted with 60 µl Qiagen EB buffer after incubating 15 min at room temperature and stored at -20°C.

**Permafrost samples**

DNA Extraction Protocol:

Extraction of DNA from the permafrost samples (see Table 3) was carried out in the same aDNA lab facilities at the University of Copenhagen as above. Approximately 10 g of soil per sample was used for the extractions with the use of the PowerMax™ Soil PowerBead tubes (Cambio UK Ltd), and dissolved in 24 ml lysis buffer (Bulat et al. 2000). The tubes were then agitated vigorously for 1 min and left to incubate overnight at 65$^{o}$C under gentle agitation. Following extraction, the DNA was purified using the PowerMax™ Soil DNA Isolation Kit protocol according to manufacturer's instructions.

PCR

**Permafrost and Ice Samples**

To reduce the risk of contamination, PCRs were set up in aforementioned clean lab facilities in a UV-treated box. Reagents (except primers) were filtered through 30 kDa centrifugal filter units (Amicon Ultra-4, Millipore, Billerica, MA), before use and all plastic and pipettes were UV-treated before use.

Both generic and GS-FLX based fusion PCR reactions were performed with universal primers targeting short fragments of 50 – 120 base pair (bp) from conservative coding and non-coding regions within plants (rbcL, trnL), insects (CO1), mammals (16S), birds (12S) and nematodes (18S) (See Table 3 for primer and PCR information).

To allow for differentiation between the fusion amplicons after sequencing, the fusion PCR primers contained a 46 bp flanking sequence required for sequencing and a 8 bp Multiplex Identifiers tag sequence in the 5' end of the primer to be able to differentiate samples (Binladen, et al., 2007).

All PCR amplifications were performed in 25µL reactions containing 1-3µL of DNA extract, 0.2mM of each dNTP, 0.4µM of each primer, 5mM MgSO4, 1.0U Platinum®

High Fidelity (HiFi) DNA polymerase (Invitrogen, Carlsbad, CA) with 1X high fidelity buffer, in addition to 25µg of BSA (only in non-mammalian amplifications). Cycling conditions were 4' at 94°C, 55 – 65 cycles of (30'' – 1' at 94°C, 30'' – 1' at $T_a$, 30'' – 1' at 68°C) and a final extension at 72°C for 10'.
Each of the PCR amplifications was repeated in 2 - 5 independent reactions in house and replicated in two independent reactions in Murdoch University aDNA lab. This was to account for competition among molecules in the first rounds of PCR cycles whereby more abundant templates compete for reagents with less abundant ones such that the products from less abundant molecules are not detected (Schlichter & Bertioli, 1996).

The clean ice samples were tested for amplification products following the same PCR conditions as mentioned above. This was to test for any flora or fauna DNA in the clean glacier ice that would then indicate that the DNA was deposited and incorporated into the ice sheet from distant sources and transported to the ice by wind. If no amplification products were picked from the glacier ice we would expect the DNA from the basal ice to be of local origin adsorbed to the soil particles and preserved in the ice.

**Table 4**. Universal primers used for amplification and their details

| Primer name | Sequence 5' - 3' | Genomic Region, Organism | Annealing temp. (ºC) | Insert Fragment length | Reference |
|---|---|---|---|---|---|
| rbcL h1aF | GGCAGCATTCCGAGTAACTCCTC | chloroplast, Plants | 45-47 | 95 bp | (Poinar, et al., 1998) |
| rbcL h2aR | CGTCCTTTGTAACGATCAAG | | | | |
| trnL g | GGGCAATCCTGAGCCAA | chloroplast, Plants | 57 | 10 - 140 bp | (Taberlet, et al., 2007) |
| trnL h | CCATTGAGTCTCTGCACCTATC | | | | |
| 16SA&M F | CCCCGAAACCAGACGAGCTA | mtDNA, 16S, Mammals | 56 | 28 - 30 bp | (Rasmussen, et al., 2009) |
| 16SA&M R | TCACTATTTTGCNACATAGA | | | | |
| InsCOI F | TTATGCTATATTANCTATTGG | mtDNA, Cox1, insects | 50 | 55 bp | (Willerslev, et al., 2007) |
| InsCOI R | GTAAAGTAAGCTCGTGTATC | | | | |
| 12SHf | CCTTGACCTGTCTTGTTAGC | mtDNA, 12S, birds | 52 | 44 bp | (Cooper, et al., 2001) |
| 12SKr | CCTACATACCGCCGTCGCCAG | | | | |

**Replication of DNA extraction and amplification**:
Samples from both Greenland and Antarctica were sent to Murdoch University, Australia to be independently extracted and PCR amplified in the ancient DNA laboratory following the same protocols as described above. Replicated samples are listed in Tables 1-3 and PCR conditions in Table 4.

SEQUENCING

Regular, non-fusion PCR products were checked on 2% agarose gels and gel purified. The gel extraction was performed by mixing each of the PCR products with 5 μl loading dye, heating to 70°C for 5 minutes before cooling on ice and then run on a 3% gel. The size of the different amplicons was visualized under UV and gel slices of the correct lengths containing the DNA fragments were cut with sterile scalpels. The DNA fragments were gel-purified by Qiaquick Gel Extraction kit (Qiagen, Valencia, CA), according to manufacturer's instructions. Purified DNA was cloned (Topo TA cloning, Invitrogen) and conventional Sanger sequencing was performed at a commercial facility (Macrogen).

Fusion PCR products were also checked on gels to minimize sequencing of non-specific PCR amplicons following the above procedure. Dilution series ($10^{-1}$ – $10^{-6}$) of each PCR product were quantified by SYBR green-based quantitative real-time PCR (qPCR) against a known molecular standard following (Meyer, 2007). Real-time PCR reactions were performed in 25μl reactions that each contained 1μl amplicon DNA, 1X AmpliTaq Gold buffer (Applied Biosystems, Foster City, CA), 2mM MgCl2, 1μl Rox/SYBR mix (Invitrogen, Carlsbad, CA), 0.4 mM of each dNTP (Invitrogen), 0.4 μM of each primer, 1.0U AmpliTaq Gold (Applied Biosystems, Foster City, CA). The qPCRs were run on a Roche LightCycler 480, with following thermal profile: A pre-denaturing step, 10' at 95°C, followed by 45 cycles of (30'' at 95°C and 1' at 60°C), followed by a dissociation-curve step from 55°C to 95°C. Primers that target the FLX specific sequence were used: FLX_ampli_F: 5' GCCTCCCTCGCGCCATCAG, FLX_ampli_R: 5'GCCTTGCCAGCCCGCTCAG.

Based on the qPCR results, the diluted and tagged amplicons were pooled into 4 – 8 pools per sequencing run in an equimolar ratio, such that samples with the same tags were included in separate pools. The pools were then run on 4 ¼ or 8 1/8 pico-titer plate (PTP) regions in each sequence run.

All sequencing runs were performed on the Roche GS FLX platform (Roche-454, Branford, CT) according to manufacturers protocol for amplicon sequencing (Roche emPCR Method Manual Lib-A).

## DATA ANALYSIS

Biases in the sequence diversity of the dataset may be introduced at different stages in the processing pipeline. For instance Marguiles, et al. (2005) showed that the presence of homopolymers in sequence reads produce a higher error rate due to insertions, deletions or carry forward (insertions or substitutions) compared to reads without homopolymers (Huse, et al., 2007). Fragments with long and several homopolymers will therefore have a higher error score than fragments with few or none homopolymers due to the higher risk of sequence errors (Huse, et al., 2007). We attempted to correct for any false diversity introduced by amplification or sequencing errors by three denoising approaches: the Rasmussen approach, the AmpliconNoise and Orlando and Ginolhac approach, and, the OBITools approach. All these approaches were applied to the same dataset originating from several GS-FLX sequence runs and are described in detail in the following three sections. Additionally, I describe three different assignment approaches applied to the filtered and sorted sequence diversity after the denoising; SAP, OGA and EcoTag (all described below).

### DENOISING

#### Rasmussen approach

The raw sequence reads are generated through the “Signal Processing for Amplicons” in the GS FLX pipeline to decrease noise from ghost wells (Genome Sequencer FLX System Software Manual, version 2.3, 200) (Margulies, et al., 2005). The reads are

first processed using an in-house Perl script that implements the Needleman-Wunsch algorithm for global alignments and takes quality scores at each position into account. Then, the sequences are sorted according to primer and tag sequences and trimmed for these using the Smith-Waterman algorithm for local alignments (M. Rasmussen, personal communication).

A last step was included to remove diversity caused by amplification and sequencing errors due to the fact that a measured sequencing error rate of 0.2% per base is associated with the FLX platform (Rasmussen, personal communication), which is similar to what has been reported previously (Huse, et al., 2007). As the DNA barcode fragments used in this study contain homopolymers that are known to be prone to amplification and sequencing errors (Quince, et al., 2011) we followed a naïve strategy described below.

If any of the extraction controls or PCR controls had shown positive results they were also sequenced and the results were filtered for any overlap in sequence similarity. We applied a pipeline that extracts and excludes all data sequences that are 100% identical to the sequences in the controls within PCRs. Moreover, we also excluded sequences represented by a frequency of less than 1% of the most frequent haplotype and considered these as false diversity generated due to amplification or sequencing errors. If the most frequent haplotype was observed less than 100-fold, sequences represented by a unique sequencing read were discarded.

In addition, we have tested an alternative approach for filtering to be used in combination with SAP. This method was designed to remove false diversity relying on the assumption that errors occur randomly across the entire read length, and thus can be described by a binomial distribution. If two sequence haplotypes differ and one is much more common than the other, the less frequent would be considered false diversity due to error if the number of differences falls within the assumed error rate. The number of sequences filtered using this approach was similar to that obtained by the previously described Rasmussen approach (Table 5), except for rbcL which was less aggressively filtered – 43% reads were left after filtering compared to the previous 33%.

**Table 5**. Counts of amplicon reads for each of the four primer pairs at different stages of trimming in the process of removing erroneous diversity by the Rasmussen approach.

| | 16S | Cox1 | trnL | rbcL | Total reads |
|---|---|---|---|---|---|
| Primer and tag match from .fna files | 107349 | 86420 | 251136 | 286279 | 731184 |
| Cleaned for contaminants observed in controls | 25046 | 17357 | 207627 | 149932 | 456274 |
| Conservative removal for errors (if < 1% of most abundant type) | 18948 | 14224 | 162785 | 94566 | 290523 |

**AmpliconNoise and Orlando/Ginolhac (OG) approach:**

This approach deals with noise by making use of a combination of AmpliconNoise (Quince, et al., 2011) and a supplementary in-house Perl pipeline (Orlando and Ginolhac, unpublished).

GS FLX light intensity translated sff.txt sequence flow files were sorted per combination of primer and MID tags in separate files and trimmed accordingly while retaining information about the PCR, sample ID and primer. A cut-off of 40% mismatches within the primer sites was allowed while no mismatches were allowed within the tag sequences. The number of sequences passing this first primer and tag trimming and identification step represented a total of 1,040,460 reads, which is a higher number than in the Rasmussen approach due to the less conservative criteria for tag and primer match. However, as opposed to both the Rasmussen approach and the OBITools approach (described later) this method used the quality information of the reads included in the light intensity sff.txt files.

Trimmed sequence reads were used as input for AmpliconNoise in order to remove amplification and sequencing errors (Quince et al. 2011). AmpliconNoise is a program that deals with removing amplification and pyrosequencing errors using an approach where rare sequences with few mismatches to abundant ones are considered different due to error. AmpliconNoise uses a mixture model to fit noise distributions around proposed true sequences. The original procedure in AmpliconNoise that only keeps reads where the first noise flow occurred on or after flow number 360 for Titanium chemistry was relaxed to flow number 100, given the short length of the trimmed amplicons. In addition the program classifies signals for different nucleotides as noisy reads if they only have intensities between 0.5 – 0.7. Therefore only reads with signal intensities above 0.7 were accepted if occurring on or after flow 100 to reduce sequencing noise. AmpliconNoise clusters sequences into operational taxonomic units (OTUs) according to a model where a true sequence would have noisy sequences distributed around it and the relative weights for each cluster should correspond to the true relative frequency of the sequences. The expected magnitude of sequence noise is distributed as an exponentially decaying function where the distance between the true sequence and the sequence errors decrease accordingly.
The parameter for the expected magnitude of sequence error, $c_p$, was set at 0.01 and

that for determining cluster size, $\sigma_p$, was set at 1/60 and was based on hierarchical clustering from calculated edit distances of the flow grams according to a given cut-off, $\sigma_p$. Both were incorporated following the recommendations in Quince et al. (2011).

The magnitude of these parameters is important, as a cluster size that is too big will fail to capture the whole diversity present in terms of number of OTUs, and one that is too small might capture noise as OTUs. Another important consideration is to be sure to capture the true OTUs, because even if the number of different OTUs in a sample is correct they may not represent the real OTUs and could have missed some and kept noise as OTUs if these occur at a high frequency. It is shown that when sequences differ by less than 1.5% AmpliconNoise misses true OTUs but captures them when sequences differ by more than 1.5% (Quince, et al., 2011). Therefore, when short reads from trnL, 16S or COI are considered where most sequences are between 20 – 55 bp, one mismatch among sequences would yield a difference above the 1.5% level but for the rbcL reads of 95bp and the few longer reads of trnL (up to 140 bp) we might miss out some true OTUs by this denoising approach.

The different scripts involved in AmpliconNoise consisted of the following steps:
1. Splitkeys.pl uses the input of the sff.txt file from each FLX run, a file with primer sequences and a file with the MID tag sequences. This step searches for the reads with a match to each of the different primers and then for each primer searches for matches between the different tags and discards unmatched reads. 2. clean100TrimPrimerR.pl trims the reverse primer as well as sequences with a signal intensity too low to pass the quality criteria mentioned above.
3. PyroDist calculates the distances among the different trimmed sequences prior to the clustering step that is performed by FCluster. FCluster uses the parameter determining cluster size, $\sigma_p$, and also removes noisy sequences due to PCR errors. PyroNoise.pl uses the parameter, $c_s$, to remove further sequencing errors. These two steps are performed while the forward primer is still untrimmed to ensure that clustering only occurs within each primer and not among different fragments.
4. CleanTrimPrimerF.pl removes the forward primer. Both the primer-trimming steps use ClustalW to align sequences and calculate distances.
5. The Perseus script in the AmpliconNoise pipeline removes chimeric sequences but

only if the sequences are longer than 250 bp, hence, in this case this step did not detect any chimeric sequences. However, if such errors were present they would not have a blast hit and therefore be discarded further downstream in the assignment pipeline (see below). Sequences that passed the AmpliconNoise pipeline got an extension of .good.

6. The .good output files from AmpliconNoise were further processed through a step to check for any blast hit to the relevant databases using an in-house perl script (filterGoodBlastRevComp.pl), and only sequences with a match entered the next pre-assignment step and got the extension .ok.

7. Finally, sequences were further filtered for any contaminants that had been amplified from extraction or PCR controls through another perl script (prepDatasetHtAlready.pl) that provided the extension .fst for the filtered and cleaned sequences.

See Figure 3 for a flow chart representation of the above described steps and Table 6 for the number of sequences that passed the different steps with this denoising approach.

**Figure 3**. Illustration of the process of denoising the raw GS FLX light intensity sff.txt files. Boxes coloured in orange are part of the AmpliconNoise pipeline, while those in green are part of the in-house Perl scripts (Orlando & Ginolhac, unpublished)

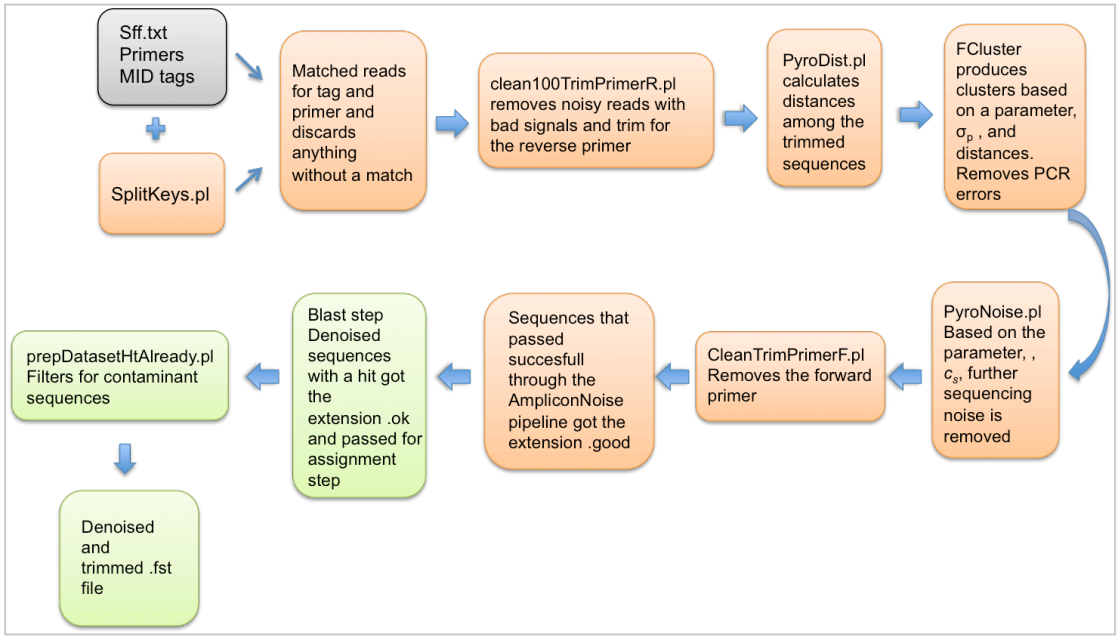


**Table 6**. Counts of amplicon reads for each of the four primer pairs at different stages of trimming in the process of removing erroneous diversity by a combination of AmpliconNoise and the Orlando and Ginolhac approach.

| | 16S | Cox1 | trnL | rbcL | Total |
|---|---|---|---|---|---|
| Primer match with splitKeys.pl and first cleaning of noisy reads (.fa) | 144499 | 385812 | 208288 | 301861 | 1040460 |
| Passing through AmpliconNoise (.good) | 85164 | 248822 | 147966 | 158589 | 640541 |
| Passing through filtering Blast step (.ok) | 37232 | 20300 | 147469 | 126303 | 329922 |
| Filtered for contaminants observed in controls (.fst) | 28524 | 8583 | 117248 | 87350 | 239032 |

**OBITools approach**

The raw sequence reads generated through the "Signal Processing for Amplicons" in the GS FLX pipeline, as mentioned in the Rasmussen approach, were also used to generate the input fasta files for downstream analyses in this approach. The unfiltered .fna sequence read files were analysed through OBITools software that is available from Rhône-Alpes Bioinformatics Center (www.prabi.grenoble.fr/trac/OBITools). This approach only accepts sequences that have an exact match to the 8bp tags and allows for two mismatches in the primer sites. Amplified sequences passing these criteria were trimmed for primers and tags and continued through to a clustering step that assembled strictly identical sequences together while retaining the information about sample ID, primers and occurrence frequency within the sequencing run. Sequences that were shorter than 10bp and only occurred once are likely to be PCR or sequencing errors and were therefore excluded from further analyses. Sequences passing this step got the extension tag.tab (see Table 7). Another filtering step was added that only allowed sequences with a best identity score of at least 95% to be reported in the results and which dealt with potential sequence errors in the following way. If several variants differing by only one nucleotide occurred in the same PCR, the variant being less frequent got a designation (i), indicating that it could be an erroneous representation of the most frequent variant (h). However, if rare variants occurred in low frequencies with no more frequent variants differing by only one

nucleotide to the rare variant, these were accepted as rare true diversity, designated as singletons (s).

Databases were generated for each of the four markers, rbcL, trnL, COI and 16S, based on the diversity present in GenBank (NCBI). These databases were filtered for redundancy and sequences without taxonomic information. The trnL marker has been extensively used as a barcode in the EcoChange project (http://www.ecochange-project.eu) and specific sequence databases based on this marker exist for the Arctic plant diversity (GenBank Accession number GQ244527 to GQ245667) and Boreal biodiversity database (not yet uploaded to GenBank) in addition to EMBL Nucleotide Sequence Database (EBI) containing all trnL sequences in genbank. Hence, three local databases were generated for trnL.

**Table 7**. Counts of amplicon reads for each of the four primer pairs at different stages of trimming in the process of removing erroneous diversity by the OBITools approach.

| | 16S | Cox1 | trnL | rbcL | Total reads |
|---|---|---|---|---|---|
| Primer and tag match when passed though the first step of OBITools (tag.tab) | 106426 | 328176 | 199355 | 226656 | 860613 |
| Removal of sequences with less than 95% identity score (filtered.tab) | 35417 | 9425 | 109769 | 159897 | 314508 |
| Sequences left when contaminants observed in controls are removed | 22394 | 9422 | 62591 | 98494 | 192901 |

## ASSIGNMENT OF SEQUENCES TO TAXONOMIC GROUPS

After trimming, the sequences were run through three different pipelines that assign the sequences from each genetic marker to the closest relative taxa groups at family, genera or species level. We chose three different approaches to look for possible differences in the diversity assigned, test for assignment bias and to be able to use a conservative interpretation of results found by more than one method. The three assignment pipelines are based on three different statistical and phylogenetic approaches. One is a similarity-based approach that assigns sequences based on similarity or best identity score (EcoTag). The other two methods follow a phylogenetics approach either by using Bayesian statistics (SAP) which provides posterior probabilities of assignments to a specific taxon, or, by using Maximum Likelihood statistics (OGA) which provides branch support scores of assignments to the smallest monophyletic clade of reference database sequences that a sample sequence clusters to. All three assignment methods are described below.

### Orlando and Ginolhac Assignment approach (OGA)

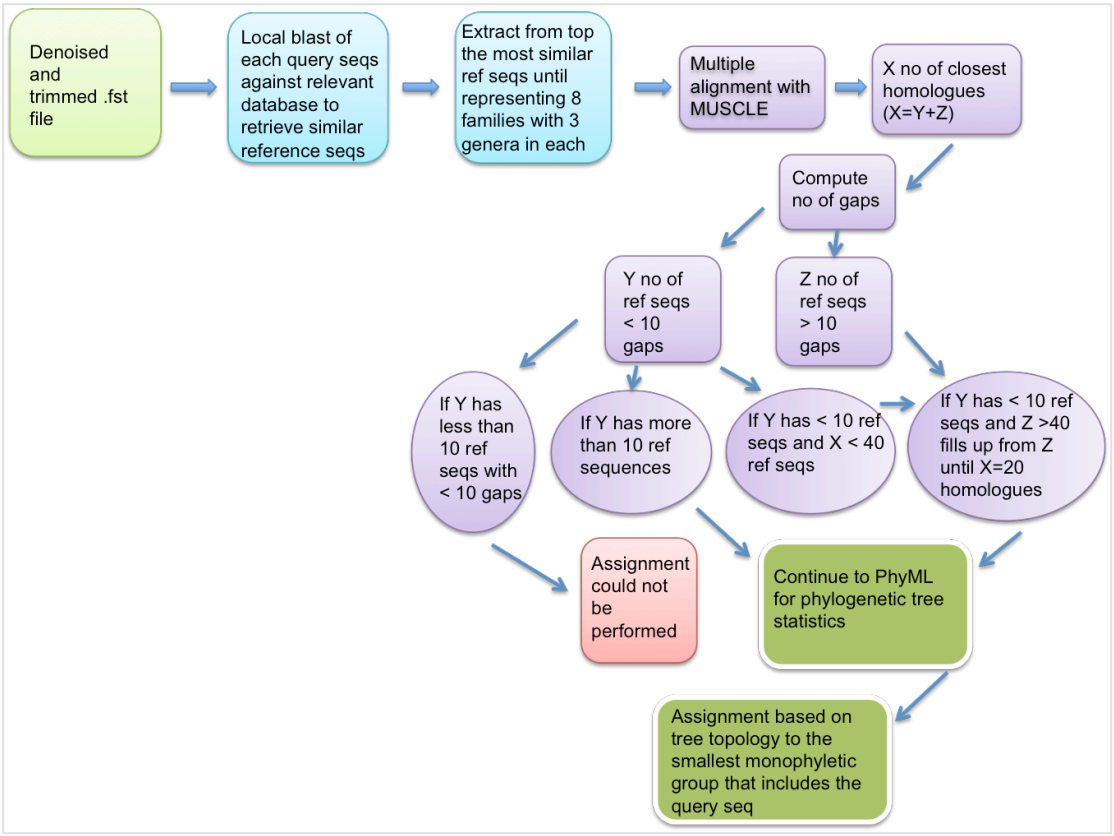
Denoised and trimmed fasta files were processed through a custom-made pipeline consisting of eight in-house Perl scripts. This pipeline was specifically designed for performing taxonomic assignments of short next-generation sequencing reads. Before assignments could be performed did we set up local databases generated from GenBank.

In the first step, a flat database file consisting of all trnL and rbcL sequences available for Embryophyta were generated by screening GenBank (release on 6th May 2011 and available at http://www.ncbi.nlm.nih.gov/nuccore/advanced) using first trnL and then rbcL as a keyword and Embryophyta as Organism (for a total of 111,695 and 49,580) downloaded sequences for trnL and rbcL, respectively). After conversion to fasta format, the GenBank flat database files were further parsed and sequences showing none of the trnL or rbcL barcoded PCR primers or a length superior to 200 nucleotides or inferior to 10 nucleotides within the primer sites were discarded. The sequences recovered were further trimmed for PCR primers and identical sequences were collapsed into haplotypes. Overall, the trnL and rbcL databases used for the

assignment pipeline consisted of a total of 13,427 and 4,432 unique sequences, respectively.

The assignment pipeline was set up according to the procedure outlined in Figure 4.

**Figure 4**. Illustration of the process of assigning query (.fst) sequences by the OGA approach (Orlando, Ginolhac *et al*., unpublished). Light green: input sequences from denoising and trimming step in AmpliconNoise;
Blue: BLAST step to retrieve most similar sequences from database covering 8 families with 3 genera in each to perform assignment against;
Purple: sorting, multiple alignments and computation of differences to retrieve similar sequences; Green: Maximum Likelihood-based phylogenetics assignment.




For each clustered, denoised and trimmed FLX sequence, a subset of most similar sequences were identified through a BLAST search using minimum blast score and word size of –e 10 and –W 7 against the relevant database. Then, the BLAST results were parsed for the highest e-values and associated taxonomic information was retrieved. The parsing is fulfilled when 8 families with at least 2 genera are recovered.

A supplementary step filtered out reference sequences showing more than 10 gaps to the query. This step is critical because while we want enough information from the reference sequences to make a successful assignment, it is also necessary to avoid too many reference sequences that would not add to the information about the monophyletic group a query sequence could be assigned to (see Figure 4, blue boxes). Global alignment of query sequences to reference sequences was done by MUSCLE (Edgar, 2004). Following the MUSCLE alignment indels were recoded as missing information (N) and the number of reference sequences were divided into two groups: (a) sequences containing up to 10 gaps (Y) and, (b) sequences containing more than 10 gaps (Z). Four different scenarios could occur (see Figure 4, purple ovals). If the reference sequences that fulfilled the diversity criteria had less than 10 gaps and constituted more than 10 different sequences, the assignment was continued to the phylogenetics step with all of these extracted sequences. On the other hand, if there were fewer than 10 reference sequences that had less than 10 gaps to the query sequence the pipeline would fail to perform assignment to very divergent taxa (red circle). If there were less than 10 similar sequences containing greater than 10 gaps and we obtained more than 40 reference sequences in the total extracted set (X), all of the sequences containing less than 10 gaps were included (Y) and sequences with more than 10 gaps (Z) were included until 20 reference sequences were obtained. If less than 40 reference sequences were obtained in (X), all references were used regardless of their number of gaps to the query.

Following this step of producing sample sets to base the phylogenetic computation of trees on, the next downstream step in the pipeline was phylogenetic analyses in a Maximum Likelihood framework (green boxes).

Maximum Likelihood phylogenetic inference was performed using PhyML v3.0 (Guindon, et al., 2010) assuming a HKY model of molecular evolution, where node support estimates were based on Approximate Likelihood Ratio tests (Anisimova & Gascuel, 2006). Tree topologies were parsed and the smallest monophyletic group showing support values superior to 0.80 that included the sequence to be assigned was recorded. Taxonomic levels common to all sequences present in the latter group provided a conservative taxonomic rank to assign the sequence to. Contrary to BLAST (Altschul, et al., 1997), this approach perform taxonomic assignments based on phylogenetic inference, rather than

sequence similarity, but without requiring extensive computational time. Hence, the approach is fully compatible with the high-throughput nature of next-generation massively parallel sequencing platforms, such as 454 GS-FLX. The taxon level a query sequence is assigned to then depends on what taxonomic level the query sequence will share the smallest monophyletic group with of the homologues from the reference sequences retrieved. If this is at the species level, we have a species assignment, however as the information in these short sequences is not always unique below family and genera level (or as the database is not complete) the smallest shared monophyletic group might be at a higher taxonomic level than species.

We tested the database against itself for both trnL and rbcL markers and found that a number of 8 families and a branch support of 0.80 were appropriate to achieve an assignment. With these parameters trnL provided assignment of 69% of the database sequences at the order level, of 64.2% at the family level, of 50.3% at the genus level and of 49.7% at the species level when testing the database against it self. rbcL marker provided assignment of 66.6% at the family level, of 46.7% at the genera level and of 44.1% at the species level. We also tested the effect of decreasing the branch support, which increased the number of assignments but because of the risk of false positives the 0.80 level was maintained.

**Table 8**. Percentages of unassigned sequences when the trnL database was tested against it self for different branch supports and different number of families in the reference dataset; for three, five and eight families, respectively, with 3 genera in each.

| trnL_test | Order | Family | Genus | Species |
|---|---|---|---|---|
| 3_Fam_50 | 31.1 | 36.9 | 51.1 | 51.6 |
| 3_Fam_70 | 31.8 | 37.0 | 51.7 | 52.2 |
| 3_Fam_80 | 32.7 | 38.5 | 52.9 | 53.4 |
| 5_Fam_50 | 30.3 | 34.6 | 50.0 | 50.5 |
| 5_Fam_70 | 31.0 | 35.5 | 50.0 | 51.0 |
| 5_Fam_80 | 31.9 | 36.6 | 51.7 | 52.2 |
| 8_Fam_50 | 29.5 | 33.0 | 47.2 | 47.8 |
| 8_Fam_70 | 30.1 | 33.2 | 48.0 | 48.3 |
| 8_Fam_80 | 31.0 | 35.2 | 49.7 | 50.3 |

Assignment success for the rbcL marker increased at the family, genera and species level when increasing from 5 families to 10 families for the homologue diversity (Table 9), however the computation time did as well, and therefore the analyses were continued with a number of 8 families for rbcL.

**Table 9**. Percentages of unassigned sequences when the rbcL database was tested against it self for different branch supports and different number of families in the reference dataset; for five, eight and ten families, respectively, with 3 genera in each.

| rbcL_test | Order | Family | Genus | Species |
|---|---|---|---|---|
| 5_Fam_50 | 37.6 | 38.2 | 45.3 | 45.3 |
| 5_Fam_70 | 37.8 | 38.4 | 45.4 | 45.4 |
| 5_Fam_80 | 38.0 | 39.0 | 45.7 | 45.7 |
| 8_Fam_50 | 34.7 | 34.7 | 42.3 | 42.6 |
| 8_Fam_70 | 34.7 | 34.7 | 42.4 | 42.7 |
| 8_Fam_80 | 34.9 | 34.9 | 42.2 | 42.4 |
| 10_Fam_50 | 33.1 | 33.4 | 39.7 | 39.7 |
| 10_Fam_70 | 33.2 | 33.5 | 39.3 | 39.3 |
| 10_Fam_80 | 33.3 | 33.5 | 39.9 | 39.9 |

After sequences had been run through the assignment pipeline, results were compiled in tables and compared among replicates of the same sample. Only results that were assigned with a confidence score from PhyML of >0.80 and which had an occurrence of > 1 were accepted, and passed on to the test comparison among assignment tests.

The number of sequences available for the assignment pipeline from each of the four markers for the 3 locations is listed in Table 10 (only trnL and rbcL amplicons were assigned by this approach).

**Table 10**. Number of sequences that passed through the filtering steps using a combination of the AmpliconNoise and Orlando/Ginolhac pipeline, from each sampling locality and for each of the four markers.

| | trnL | rbcL | 16SA&M | Cox1 |
|---|---|---|---|---|
| Camp Century | 83443 | 47847 | 13313 | 3890 |
| Antarctica | 20552 | 18732 | 0 | 4626 |
| Kap København | 10818 | 57 | 8855 | 0 |

**SAP**

The Statistical Assignment Package (SAP) (Munch, et al., 2008a) was another approach used for assigning all of the four markers. This program takes advantage of a Bayesian phylogenetic approach, providing a posterior probability of how well a query sequence belongs to a phylogenetic clade. The method relies on global alignments to a set of homologous sequences representing a number of families, genera and related species to each of the query sequences.

A Monte Carlo Markov Chain (MCMC) approach is used to evaluate all possible trees from the alignment step, which takes into account all the different substitution parameters (substitution rates, nucleotide frequencies, topology and branch lengths of the trees) that could provide different trees. Ten thousand unrooted trees sampled from the MCMC analysis were analyzed to obtain posterior probabilities of assignment to all taxa represented in the compiled set of homologues. The posterior probability was then summed over all the different phylogenetic trees where the MCMC algorithm samples trees, estimating a posterior probability of where the query sequence falls best within a taxonomic monophyletic group. This approach is advantageous to BLAST (Altschul, et al., 1997) as it takes into account information of the phylogenetic issue that is not included in a BLAST similarity score based on local alignments of homologues to a query sequence. Moreover, if more homologues are equally similar to a query sequence, the phylogenetic approach may be more optimal than BLAST (Altschul, et al., 1997) as it provides a statistical measure of the phylogenetic assignment to a specific taxon (Munch, et al., 2008a). However, the BLAST search is advantageous for the quick identification of highly similar homologues relevant to the retrieval of alignments and MCMC phylogenetic tree sampling steps. Homologous reference sequences are chosen based on the BLAST score of each query sequence to the NCBI where homologues with a BLAST score of at least half that of the query sequence are retrieved. Another step to decrease the number of sequences includes only one homologue for each species where up to 30 different species are included. However, if the cut-off has not been reached then more homologues are added, from a best score criteria until 10 different genera, 6 families, 5 orders, 3 classes and 2 phyla are included, summing to 50 different homologues for each sequence assignment. If the BLAST hits do not have more than 5 homologues with an e-value below 0.1, the assignment is cancelled.

The interpretation of the posterior probability is then a percentage of how correct the assignment is based on the available sequences in GenBank when a uniform prior is given on the tree topology.

The input sequences for SAP were trimmed for MID tags and primers using the above mentioned in-house perl scripts from the Rasmussen approach (Rasmussen, unpublished) and were cleaned for any contaminant sequences with a 100% match and were compiled into haplotypes if identical sequences were present within samples. Moreover, to avoid rare sequence variants that could be false diversity due to amplification or sequencing errors, only sequences with a frequency of more than 1% of the most common haplotype were included in the SAP assignment. Last, any sequences represented only once were discarded from the analysis.

Only sequences with a posterior probability of assignment to a taxonomic group superior to 0.8 were accepted, even though false assignments are only considered present when the posterior probability is < 0.5 (Munch, et al., 2008a). However, as the NCBI database has increased significantly since 2008, a conservative approach was chosen.

The numbers of sequences passing the above-mentioned criteria ready for assignment in SAP are listed in Table 11 for each of the four markers.

**Table 11**. Number of sequences that passed through the filtering steps using the Rasmussen approach pipeline, from each sampling locality and for each marker.

| | trnL | rbcL | 16SA&M | Cox1 |
|---|---|---|---|---|
| Camp Century | 118342 | 71813 | 16492 | 5079 |
| Antarctica | 20987 | 22654 | 0 | 18500 |
| Kap København | 21803 | 65 | 2456 | 107 |

**EcoTag**

The clustered and trimmed sequences that passed through the OBITools described previously were treated as molecular operational taxonomic units that were assigned against databases corresponding to the different markers used; rbcL, 16S, Cox1 and trnL. For the trnL marker, 3 different databases were used: the Arctic, the Boreal and the whole EMBL database. The results were gathered in a table format that allowed comparison among replicated sequences for each marker between different PCRs and different samples. Information was further provided about the frequency of the individual sequence to specify if the variant was a unique variant without any variants differing by only one nucleotide within the same PCR (designated as singleton (s)), a common variant (h) with a high frequency within the PCR, or a variant (i) with a low frequency in the PCR of another common (h) variant. Hence, both h and i variants differ by one nucleotide to another sequence variant within the same PCR where h is the most common and i the least common. If a sequence variant is not replicated and only found as an i variant, it might be false diversity caused by amplification or sequencing error of another more common variant. This kind of information is relevant for validating the biodiversity found and to test for reproducibility of the results as well as competition between PCRs.

When sequences had passed through the first filtering step of removing tags and primers and being represented with a frequency of above 1, the total number of unique sequence reads was 860.613.

When a second filtering step was added to the previous step that removed contaminants and sequences that could be attributed as errors and only allowed sequences assigned with a best identity score of > 0.95, the number of sequences that had a successful assignment was reduced to 186,106 (Table 12).

**Table 12**. Number of sequences that passed through the filtering steps in the OBITools pipeline, from each sampling locality and for each marker.

| | trnL | rbcL | 16SA&M | Cox1 |
|---|---|---|---|---|
| Camp Century | 44580 | 81867 | 17975 | 2123 |
| Antarctica | 16453 | 16580 | 0 | 7299 |
| Kap København | 1188 | 47 | 4419 | 0 |

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Anisimova M & Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology* **55**: 539-552.

Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R & Willerslev E (2007) The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing. *PLoS ONE* **2**: e197.

Bulat SA, Lubeck M, Alekhina IA, Jensen DF, Knudsen IMB & Lubeck PS (2000) Identification of a universally primed-PCR-derived sequence-characterized amplified region marker for an antagonistic strain of Clonostachys rosea and development of a strain-specific PCR detection assay. *Applied and Environmental Microbiology* **66**: 4758-4763.

Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J & Ward R (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature* **409**: 704-707.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Funder S (1989) Quaternary Geology of the Ice-Free Areas and Adjacent Shelves of Greenland. *Quaternary Geology of Canada and Greenland* **K-1**: 741-792.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W & Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**: 307-321.

Herron S & Langway CCJ (1978) The debris-laden ice at the bottom of the Greenland Ice Sheet. *Symposium on Glacier Beds: The Ice-Rock Interface, Proceedings of a Symposium held at Carleton University, Ottawa(Ontario) August 15-19, 1978. Journal of Glaciology,* **23**.

Huse SM, Huber JA, Morrison HG, Sogin ML & Mark Welch D (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**: -.

Mager S, Fitzsimons S, Frew R, Samyn D & Lorrain R (2009) Composition and origin of amber ice and its influence on the behaviour of cold glaciers in the McMurdo Dry Valleys, Antarctica. *Journal of Glaciology* **55**: 363-372.

Margulies M, Egholm M, Altman WE, et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**: 376-380.

Meyer M (2007) From micrograms to picograms: quantitative PCR reduces the material demands of high-throughput sequencing. *Nucleic Acids Research* **36**: e5.

Munch K, Boomsma W, Huelsenbeck J, Willerslev E & Nielsen R (2008a) Statistical Assignment of DNA Sequences Using Bayesian Phylogenetics. *Systematic Biology* **57**: 750-757.

Poinar HN, Hofreiter M, Spaulding WG, et al. (1998) Molecular coproscopy: Dung and diet of the extinct ground sloth Nothrotheriops shastensis. *Science* **281**: 402-406.

Quince C, Lanzen A, Davenport RJ & Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**: 38.

Rasmussen M, Cummings LS, Gilbert MTP, Bryant V, Smith C, Jenkins DL & Willerslev E (2009) Response to Comment by Goldberg et al. on ‚ÄúDNA from Pre-Clovis Human Coprolites in Oregon, North America‚Äù. *Science* **325**: 148.

Samyn D (2005a) Structural and Geochemical Analysis of
Basal Ice from Taylor Glacier, Antarctica:
On the Role and Behaviour of the Interstitial Fluid Phase. Phd
Thesis, Universit´e Libre de Bruxelles, Belgium, Bruxelles.

Samyn D, Fitzsimons SJ & Lorrain RD (2005b) Strain-induced phase changes within cold basal ice from Taylor Glacier, Antarctica, indicated by textural and gas analyses. *Journal of Glaciology* **51**: 611-619.

Samyn D, Svensson A, Fitzsimon SJ & Lorrain RD (2005c) Ice crystal properties of amber ice and strain enhancement at the base of cold Antarctic glaciers. *Annals of Glaciology, Vol 40, 2005* **40**: 185-190.

Schlichter UHA & Bertioli D (1996) Abundance and competition in PCR. *Nature Biotechnology* **14**: 1202-1202.

Taberlet P, Coissac E, Pompanon F, et al. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**: e14-e14.

Willerslev E (2004a) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution* **19**: 141-147.

Willerslev E, Cappellini E, Boomsma W, et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**: 111-114.

# Chapter IV

METHODS DISCUSSED

The increasing amount of data generated by high throughput sequencing technologies creates the need to save computational time when assigning large datasets. I explored three different assignment approaches and evaluated (i) if there is an assignment program that is faster than SAP (Munch, et al., 2008a) but equally robust (ii) and how a phylogenetic approach like that implemented in OGA compares to the similarity based one from EcoTag (http://www.grenoble.prabi.fr/trac/OBITools). In this section, I will discuss the assignment success of the three pipeline approaches (SAP, OGA and EcoTag) in the light of (i) the criteria used when assigning a taxonomic group (ii) how the three approaches sometimes provide different results (iii) how to handle possible biases in the assigned diversity (iv) the effect of an incomplete database on the assignment success between markers and sites.

The consensus results from the three approaches can also be used as conservative assignments for data coming from ancient/damaged samples. As OGA has not been published yet, we provide the correspondent assignment specificity and sensitivity results for the two markers analyzed.

## QUANTITATIVE EVALUATION OF NUMBER OF ASSIGNED SEQUENCES PER MARKER AND PER METHOD

In this section I am presenting the total number of sequences being assigned and not the collapsed number of haplotypes. This allows the assessment of the overall assignment success using the number of sequences that is removed through filtering and those that get an assignment of at least order level.

The results from the three assignment approaches at the three different sites are listed below showing the number of sequences assigned per marker. The assignment success shown is presented for each assignment method per location relative to the number of sequences that passed the initial filtering steps – the OBITools package does not provide the number of sequences in the pre-assignment step used per location so only the total assignment success is reported.

EcoTag and SAP were run for all four markers and the OGA pipeline was run for trnL and rbcL.

**EcoTag**

The sequence data analysed by the OBITools package were assigned by EcoTag (part of the OBITools package) and the number of sequences per marker and per site that had a successful assignment to any taxonomic level are shown in table 13 and summarized in table 14.

The overall total assignment success of OBITools showed large variance among markers (3 – 43%, Table 14). Although a difference in performance between mini barcodes is expected, here such large variance could be caused by an inadequate stringency of the filtering steps leading to the removal of healthy sequences for Cox1 markers (compare to the assignment success of SAP, below, where the Rasmussen approach was used for filtering). However, since the amount of sequences used in the assignment (post-filtering) is not disclosed by OBITools, this cannot be confirmed, and distinguished from an assignment problem.

**Table 13** Number of assigned sequences per marker and locality and assignment success of the three pipelines (assignment success is measured relative to the number of sequences obtained after the filtering step).

| | | **trnL** | | **rbcL** | | **16S** | | **Cox1** | |
|---|---|---|---|---|---|---|---|---|---|
| | | N.o Seqs | Assignment success | N.o Seqs | Assignment success | N.o Seqs | Assignment success | N.o Seqs | Assignment success |
| **EcoTag *** | Camp Century | 44580 | - | 81867 | - | 17975 | - | 2123 | - |
| | Antarctica | 16453 | - | 16580 | - | 0 | - | 7299 | - |
| | Kap København | 1188 | - | 47 | - | 4419 | - | 0 | - |
| **SAP** | Camp Century | 89996 | 0.760 | 62455 | 0.870 | 6516 | 0.395 | 2706 | 0.533 |
| | Antarctica | 18193 | 0.867 | 19260 | 0.850 | 0 | 0 | 6033 | 0.326 |
| | Kap København | 8965 | 0.411 | 41 | 0.631 | 1080 | 0.440 | 0 | 0 |
| **OGA** | Camp Century | 26838 | 0.322 | 15048 | 0.315 | - | - | - | - |
| | Antarctica | 3976 | 0.193 | 12672 | 0.676 | - | - | - | - |
| | Kap København | 1748 | 0.162 | 38 | 0.667 | - | - | - | - |

*Number of assigned sequences by locality and per marker passing through the EcoTag pipeline with an identity score > 0.95.

**Table 14**. Summary of the overall performance of all methods.

| | Method | 16S | Cox1 | TrnL | Rbcl | Total |
|---|---|---|---|---|---|---|
| **Number of sequences after primer/tag match** | OBItools | 106426 | 328176 | 199355 | 226656 | 860613 |
| | Rasmussen | 107349 | 86420 | 251136 | 286279 | 731184 |
| | OGA (AmpliconNoise/OG ) | 144499 | 385812 | 208288 | 301861 | 1040460 |
| **Number sequences after filtering** | OBItools | - | - | - | - | - |
| | Rasmussen | 18948 | 14224 | 162785 | 94566 | 290523 |
| | OGA (AmpliconNoise/OG ) | 28524 | 8583 | 117248 | 87350 | 239032 |
| **% Sequences after filtering** | OBItools | - | - | - | - | - |
| | Rasmussen | 17,7 | 16,5 | 64,8 | 33,0 | 39,7 |
| | OGA (AmpliconNoise/OG ) | 19,7 | 2,2 | 56,3 | 28,9 | 23,0 |
| **Total number of assigned sequences** | OBItools (EcoTag) | 6194 | 9422 | 62221 | 98494 | 176331 |
| | SAP | 7596 | 8739 | 117154 | 81756 | 215245 |
| | OGA | - | - | 32562 | 27758 | 60320 |
| **% Assigned sequences relative to post-filtering** | OBItools (EcoTag) | - | - | - | - | - |
| | SAP | 40 | 61 | 72 | 86 | 74 |
| | OGA | - | - | 28 | 32 | 30 |
| **% Assigned sequences relative to pre-filtering** | OBItools (EcoTag) | 6 | 3 | 31 | 43 | **20** |
| | SAP | 7 | 10 | 47 | 29 | **29** |
| | OGA | - | - | 16 | 9 | **12** |

## SAP

Denoising and filtering of sequence data by the Rasmussen approach and assignment in SAP to a taxonomic level provided results listed in table 13 and 14. For this approach we could deduce the assignment success per marker as well as per site of the sequences that underwent filtering and got an assignment.

The total assignment success after using the primer-matching step of Rasmussen approach followed by SAP is 29%, and unlike for EcoTag we can see that the

aggressive filtering of some of the markers probably contributed to the corresponding overall assignment success (Table 14).

**OGA**

Database tests were performed on the trnL and rbcL markers where each of the database sequences within the trnL or rbcL databases were assigned against the whole database for each marker. These tests showed that OGA assigns to at least order level with an assignment sensitivity or success of 69% for trnL and with 65% for rbcL when using 8 families with 3 genera in each for the diversity threshold (presented in table 8 and 9 in methods chapter). The assignment specificity for the two markers was lower than the sensitivity at the order and family levels with rbcL assigning 6% of the assigned sequences as false positives and trnL 14%, see figure 8 and 9 and table 21 and 22.

The analysis of the sequence data for COI and 16S markers is still pending and therefore is not presented. The results per site and marker are listed in table 13.

The overall average success of using the AmpliconNoise method combined with the assignment by OGA was 12% for the trnL and rbcL markers (Table 14).

**Total Assignment Success**

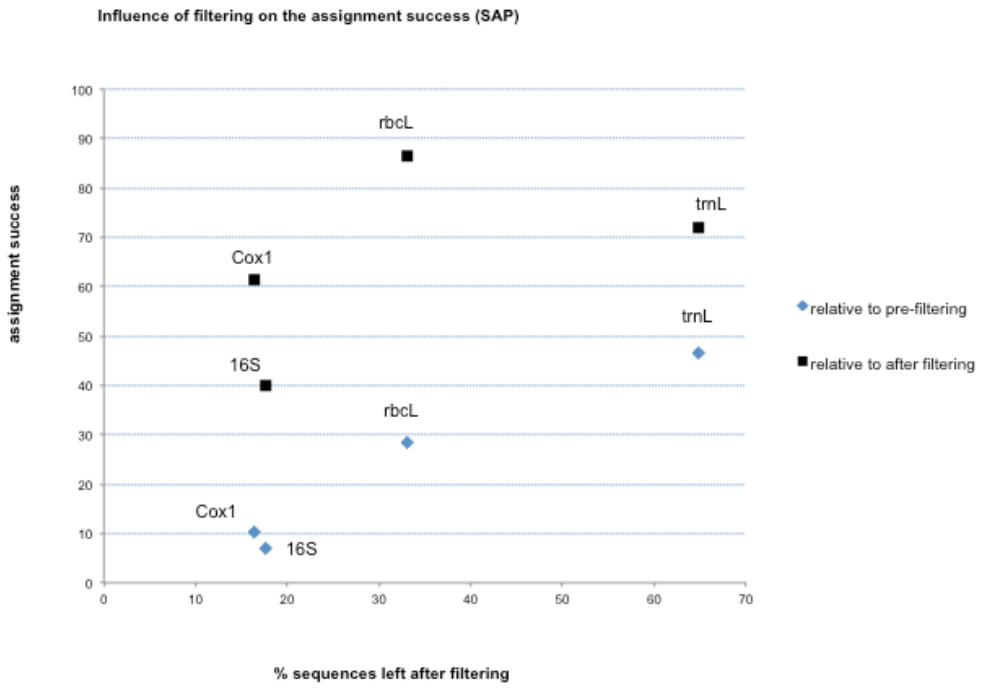
We would like to compare the total assignment success of the three methods by evaluating how many sequences are filtered and get a final assignment output. However (i) the OBITools package provides no information for the different filtering and assignment steps and (ii) we only have the results for the plant markers for OGA. We therefore present the results for all markers for SAP and EcoTag and trnL and rbcL results for OGA. We then compare the performance of all the methods using the two plant markers (Table 15).

**Table 15**. Summary of the assignment success for the plant markers for the three approaches.

| | Method | trnL | rbcL | Total |
|---|---|---|---|---|
| **% Assigned sequences relative to pre-filtering (plants)** | OBItools (EcoTag) | 31 | 43 | **37** |
| | SAP | 47 | 29 | **38** |
| | OGA | 16 | 9 | **12** |

Overall, when considering the four markers SAP has the highest total assignment success – an average of 29% success rate compared to EcoTag with 20%. When comparing all three methods for the two plant markers trnL and rbcL, we see that EcoTag and SAP performed almost equally well with total assignment success of 37% and 38%, respectively, whereas OGA presents a lower assignment success of 12% (see figure 5).

**Figure 5**. The overall trend in sequence assignment relative to the number of sequences filtered.




The relative number of assignments increases after filtering in a non-linear way considering the pre-filtered and filtered number of reads, indicating that the filtering steps have enriched some of the samples for relevant sequences – without necessarily increasing the overall amount of sequences being assigned, which means that "good" sequences are probably being removed at the filtering stage. Considering the results from SAP, we can say that if the filtering was equally efficient for all markers, we would expect to see a similar assignment success after filtering, but this is not what we find. Considering the higher assignment success rate for rbcL relative to the other markers one could say that:

1) If we compare to the trnL after filtering (Table 14, 86%/72% for SAP and 32%/28% for OGA), the higher assignment success seems to be the result of a very conservative filtering of rbcL (only 33% of the initial sequences passed the filtering steps before the SAP analysis, and 29% for OGA, compared to 65% and 56% for trnL) – the filtered reads could then be enriched for true sequences or even possibly contain some rare sequence variants (e.g. containing sequence errors/damage) that are kept after filtering inflating the false positive assignments of the rbcL sequences. This is supported by the results for COX and 16S, both with ~17% of reads left after filtering and assignment successes of 60% and 40% respectively for SAP, much lower than rbcl.
2) Since the trnL reference sequence database is much more diverse than the rbcL database, we would expect a higher assignment success for the trnL sequences from a phylogenetic point of view. However, if the similarity within families or genera is so high that the lineage sorting is not complete and the diversity larger within an order than between orders, the assignment will be unsuccessful. Some families are very diverse and hence more problematic to assign to, this could also provide a difference in the assignment success if they were better assigned by the rbcL than the trnL marker.
3) Also, the higher number of sequences being assigned to rbcL compared to trnL does not necessarily reflect a difference in the diversity being assigned. It might be that the diversity of the rbcL sequences is lower than for the trnL sequences and that some common haplotypes have a high assignment success. This would inflate the assignment success of the rbcL sequences for all methods. Examples of some families that do not assign below order level can be seen from figure 1.

Overall, the most successful pipelines are those involving the EcoTag (OBItools) and the SAP approaches, and the SAP takes the lead when considering all four markers (Table 14). Because the OBItools approach does not provide information on the amount of filtered reads, we cannot distinguish between assignment power post-filtering for this approach, and therefore cannot critically discuss why there is an inverse relationship between the number of assigned sequences to TrnL and Rbcl by the two approaches. More tests combining the different filtering approaches are

required for a more detailed assessment of the different pipelines.
The assignment success comparisons are evaluating the overall assignment of sequences examined by the different pipelines and not the diversity detected by the methods, which will be explored in the next section where I will evaluate how well the three methods detect biodiversity by the two markers trnL and rbcl.

## QUALITATIVE EVALUATION OF ASSIGNMENT SENSITIVITY AND DIVERSITY PER SITE

To evaluate sensitivity and diversity of the assigned diversity by the three methods we compared the results on a qualitative level by assessing the number of taxa that got assigned for the two plant markers. This implied looking for differences and overlaps on all the assigned diversity at the different levels – order, family and genera – for each of the methods. This qualitative approach aims at evaluating the results from a presence/absence point of view, ignoring the relative frequencies of a certain taxa after noise/error filtering, since multiple experimental factors could influence this number: (i) the denoising and filtering approaches might skew the number of sequences counted as assigned to a specific taxon and (ii) the amplification and sequencing steps might also influence the proportional success of a molecule to be amplified relative to its abundance in the sample. The representativity of a certain taxonomic group in a sample cannot be directly deduced from its overall assignment success so we present this result as presence/absence data.

### Diversity presented per Site and per Method

For each locality we generated tables counting the presence of all the different taxa found by each of the three assignment methods (see tables 15 – 17) to compare the degree of overlap between methods and the number of unique assignments by any of the methods. The criteria for being accepted as a positive result differed among the three methods as they are based upon different statistics:

1) The setup for EcoTag was based on similarity test statistics that compare the match of a query sequence to the reference database on the sequence level and searches for the closest match. Assignment by the EcoTag approach only accepted a result if the best identity score was above 0.95.

2) SAP present posterior probabilities for each assignment and the criteria for positive assignment to a taxon was accepted if the posterior probability was above 0.80.

3) OGA provides approximate Likelihood Ratio test scores for the branch support to the smallest monophyletic clade a query sequence would assign to. A minimum support threshold was set at 0.85 with a maximum of 8 nucleotide differences to the assigned taxon, which should served to avoid possible false positive assignments.

In addition to these method-specific criteria we applied a common rule for all assignments – to be accepted, an assignment had to be found at least twice by independent amplifications of DNA sequence material by all of the methods. This decreases the damage or amplification bias on the diversity assigned. In theory, DNA damage occurs randomly in any position of the sequence in double stranded DNA (Mitchell, et al., 2005). Although, it is more damage prone in 3' overhangs and in CT rich sites due to cytosine deamination and fragmentation by depurination (Orlando, et al., 2011). However, any chemical degradation process will occur at a very low rate in cold temperatures (Lindahl & Nyberg, 1972). Therefore the chance of having the same damage-caused mutation amplified in two independent reactions is very small.

Based on these criteria we generated tables for each locality, listing the number of taxon groups found by the three different assignment methods (table 15-17). This strategy is quite conservative as the different methods have a higher degree of overlap in the taxa found if we go down in assignment score. However, decreasing the assignment threshold will also increase the number of unique assignments, that could be incorrect assignments, and we therefore followed the approach described above. The following comparisons will all be relative to the trnL and rbcL markers to allow the comparison between all methods since these datasets were assigned by all three methods.

**Greenland, Camp Century Site**

From table 15 we observe that the diversity found at the Camp century is quite high considering the number of different taxa found. What is noteworthy is that we now see the opposite pattern in terms of assignment success at the qualitative level for the two markers for EcoTag and OGA where rbcL has the highest number of total assignments as compared to trnL. . Also, EcoTag and SAP find a much higher diversity using the trnL marker compared to rbcL, whereas OGA is the most successful in finding diversity using the rbcL, double as high as that found by the other methods. This could be because: (i) the assignment parameters in OGA are not as optimized for the rbcL marker as for the trnL (ii) higher amount of false positives since we know from the database tests that this value should be around 10% of the assignments being false positives (see figures 8 and 9), or (iii) that indels are affecting the assignment by OGA, which then performs better on rbcL compared to trnL that contain more indel sites. However, some of the extra assignments found by OGA might also be true and indicate that OGA is better at performing an assignment for this marker than the other two methods where it shows a sensitivity of 64% at the family level.

Furthermore, if we compare the number of taxa found by at least two methods to the total number we see that (i) there is a better overlap between the three methods for the trnL marker than for the rbcL and (ii) the proportion of shared taxa to the total decreases when we go from order level to genus level for both markers, indicating that the more specific the assignment is, the lower the overlap between the methods. These trends are generally seen for the three sites (figure 6-7).

**Table 15**. Listing the number of assigned taxa at the Camp Century site by each of the methods and by the two markers. Unique taxon assignments are only found by one method. All the numbers are counts of the presence of the different taxa.

**Camp Century**

| Taxon level | Marker | SAP unique | SAP total | EcoTag unique | EcoTag total | OGA unique | OGA total | total† | shared by at least two methods | ratio (shared/total) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Order** | **trnL** | 0 | 15 | 1 | 19 | 1 | 18 | 22 | 18 | **0.9** |
| | **rbcL** | 5 | 18 | 0 | 6 | 10 | 21 | 28 | 13 | **0.5** |
| **Family** | **trnL** | 1 | 31 | 4 | 36 | 10 | 36 | 48 | 32 | **0.7** |
| | **rbcL** | 6 | 27 | 12 | 36 | 30 | 53 | 76 | 23 | **0.3** |
| **Genus** | **trnL** | 13 | 37 | 17 | 47 | 26 | 48 | 91 | 24 | **0.3** |
| | **rbcL** | 10 | 20 | 14 | 22 | 45 | 54 | 81 | 10 | **0.1** |
| **unique/ total*** | **trnL** | 0.13 | | 0.18 | | 0.29 | | | | |
| | **rbcL** | 0.33 | | 0.32 | | 0.63 | | | | |

*unique/total=[(unique order/total order) + (unique family/total family) + (unique genus/total genus)]/3
† Number of different taxa found by all methods: includes both common and unique assignments among the methods for each of the markers at the order, family and genus level.

## Antarctica, Suess Site

At the Suess site in Antarctica we find less diversity than at Camp Century and we see that the assignment success of the two markers differs in a different way than before (table 16). Here EcoTag finds a similar number by the two markers, whereas SAP and OGA finds more taxa by the rbcL marker and fewer for the trnL.

**Table 16**. Listing the number of assigned taxa at the Antarctic, Suess site by each of the methods and by the two markers. Unique taxon assignments are only found by one method. All the numbers are counts of the presence of the different taxa.

**Suess**

| Taxon level | Marker | SAP unique | SAP total | EcoTag unique | EcoTag total | OGA unique | OGA total | total† | shared by at least two methods | ratio (shared/total) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Order** | **trnL** | 0 | 6 | 3 | 10 | 0 | 7 | 10 | 7 | **0.7** |
| | **rbcL** | 4 | 12 | 1 | 8 | 9 | 18 | 22 | 9 | **0.4** |
| **Family** | **trnL** | 0 | 8 | 3 | 12 | 0 | 12 | 17 | 9 | **0.5** |
| | **rbcL** | 5 | 14 | 1 | 14 | 12 | 20 | 27 | 13 | **0.5** |
| **Genus** | **trnL** | 4 | 8 | 7 | 8 | 2 | 5 | 19 | 4 | **0.2** |
| | **rbcL** | 8 | 12 | 1 | 6 | 12 | 15 | 26 | 5 | **0.2** |
| **unique/ total*** | **trnL** | 0.17 | | 0.48 | | 0.13 | | | | |
| | **rbcL** | 0.45 | | 0.12 | | 0.63 | | | | |

*unique/total=[(unique order/total order) + (unique family/total family) + (unique genus/total genus)]/3
† Number of different taxa found by all methods, is including both common and unique assignments among the methods for each of the markers at the order, family and genus level.

## Greenland, Kap København Site

The Kap København site shows the lowest diversity and is the site that has been sequenced and replicated to a lower degree than the other two sites, as this is still in progress. However, the numbers from table 17 show that when the diversity is low the difference between the methods is less evident but that the overall trend from the other two sites is the same, that we have less overlap between methods the more specific we are at the assignment level. Furthermore, all three methods perform equally well in terms of the number of assigned taxa and with trnL assigning more sequences than rbcL. However the difference in number of sequences available for assignment was very low for the rbcL, which is the main reason for the difference in diversity.

**Table 17**. Listing the number of assigned taxa at the Kap København site by each of the methods and by the two markers. Unique taxon assignments are only found by one method. All the numbers are counts of the presence of the different taxa.

**Kap København**

| Taxon level | Marker | SAP unique | SAP total | EcoTag unique | EcoTag total | OGA unique | OGA total | total† | shared by at least two methods | ratio (shared/total) |
|---|---|---|---|---|---|---|---|---|---|---|
| **Order** | **trnL** | 1 | 8 | 1 | 7 | 2 | 7 | 11 | 7 | **0.6** |
| | **rbcL** | 0 | 2 | 1 | 3 | 1 | 3 | 4 | 2 | **0.5** |
| **Family** | **trnL** | 1 | 8 | 1 | 8 | 0 | 5 | 10 | 7 | **0.7** |
| | **rbcL** | 0 | 2 | 1 | 3 | 1 | 2 | 4 | 2 | **0.4** |
| **Genus** | **trnL** | 3 | 7 | 3 | 8 | 2 | 4 | 13 | 5 | **0.4** |
| | **rbcL** | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | **1** |
| **unique/ total*** | **trnL** | 0.23 | | 0.21 | | 0.26 | | | | |
| | **rbcL** | 0 | | 0.22 | | 0.28 | | | | |

*unique/total=[(unique order/total order) + (unique family/total family) + (unique genus/total genus)]/3
†Number of different taxa found by all methods, is including both common and unique assignments among the methods for each of the markers at the order, family and genus level.

ALL METHODS AND SITES COMPARED

The number of assigned taxa differed among the three assignment methods at the Camp Century site where the highest diversity is found. OGA seemed to find consistently more unique assignments than EcoTag and SAP on the family and genera level. This signal was not as evident from the Suess site where OGA only found most unique assignments for the rbcL marker and performed similar to SAP for the trnL marker. EcoTag finds the most shared taxa with either of the other methods following for the Suess site. At the Kap København site the diversity of taxa was small and all three methods performed very similar in assignment.

**Unique Assignments**

When considering the unique assignments for all sites, we observe that the number of unique assignments increases when we go down in taxonomic level from order to family to genus level for the Camp Century and Suess site. This either shows that (i) each method is able to find some specific groups that are not picked up by the other methods (ii) there is an increase in the false positive rate. Since we do not have the specificity numbers on all methods, we chose to discard unique assignments, as they might be false positives.
Further, it was not observed that any method was systematically better at assigning genera within some families different to other methods. In some cases the number of unique assignments will decrease for the genus level because there are overall fewer sequences passing through to genus level compared to family and order level (this is the case with the Kap København site where the number of taxa assigned for the two markers are lower than at the other two sites and the number of unique assignments low as well).

OGA shows a very high number of unique assignments, particularly for rbcL (Table 18). This should be interpreted carefully, since this number is very dependent on the sequence diversity in the database. OGA can assign a high score to a sequence even when there is a high number of mismatches, as long as the matched database sequence is more similar to the query than to the other sequences in the database. This means that in order to use the OGA pipeline one needs to optimize the number of mismatches allowed between query and reference sequences for different databases – for the trnL/rbcL example, the differences in unique assignments by OGA could be the result of a bigger more diverse database for trnL. Noteworthy, is that for the site with the most assigned taxa, Camp Century, the number of uniquely assigned sequences for rbcL is always higher for all methods when compared to trnL.

The overall aim of this work was to be able to describe all of the taxa present in a sample which is crucial for biodiversity studies, and that is reflected in the number of taxa found by the different methods. However, one should be always aware of any bias that can result from sequence/amplification errors and method implementation.

**Table 18**. Ratios of unique assignments per marker, per site and per method.

| | | Unique assignment ratios | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Site** | **Marker** | **SAP** | **EcoTag** | **OGA** | **average** | **Order** | **Family** | **Genus** |
| **Camp Century** | **trnL** | 0.13 | 0.18 | 0.29 | 0.20 | 0.2 | 0.3 | 0.7 |
| | **rbcL** | 0.33 | 0.32 | 0.63 | 0.43 | 0.5 | 0.7 | 0.9 |
| **Suess** | **trnL** | 0.17 | 0.48 | 0.13 | 0.26 | 0.3 | 0.5 | 0.8 |
| | **rbcL** | 0.45 | 0.12 | 0.63 | 0.40 | 0.6 | 0.5 | 0.8 |
| **Kap København** | **trnL** | 0.23 | 0.21 | 0.26 | 0.23 | 0.4 | 0.3 | 0.6 |
| | **rbcL** | 0 | 0.22 | 0.28 | 0.17 | 0.5 | 0.5 | 0.0 |
| **average** | **trnL** | **0.18** | **0.29** | **0.23** | **0.23** | 0.3 | 0.4 | 0.7 |
| | **rbcL** | **0.26** | **0.22** | **0.51** | **0.33** | 0.5 | 0.6 | 0.6 |

When we compare the unique assignments between sites, we observe that the numbers differ for all taxonomic levels and across sites (Figure 6 and 7 and table 18).

**Figure 6.** Summary of Table 18: unique assignment ratios per marker, site, and taxonomy level.

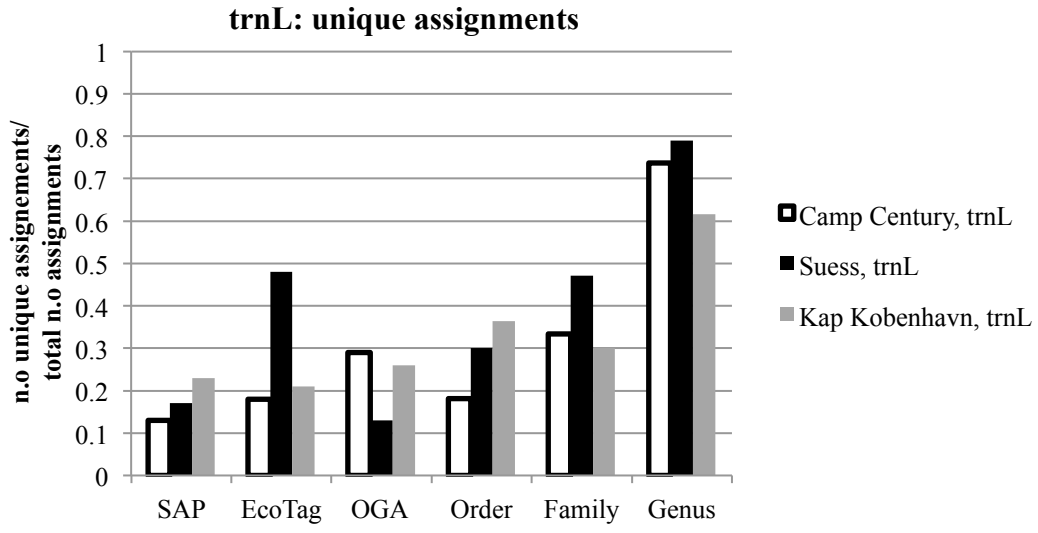
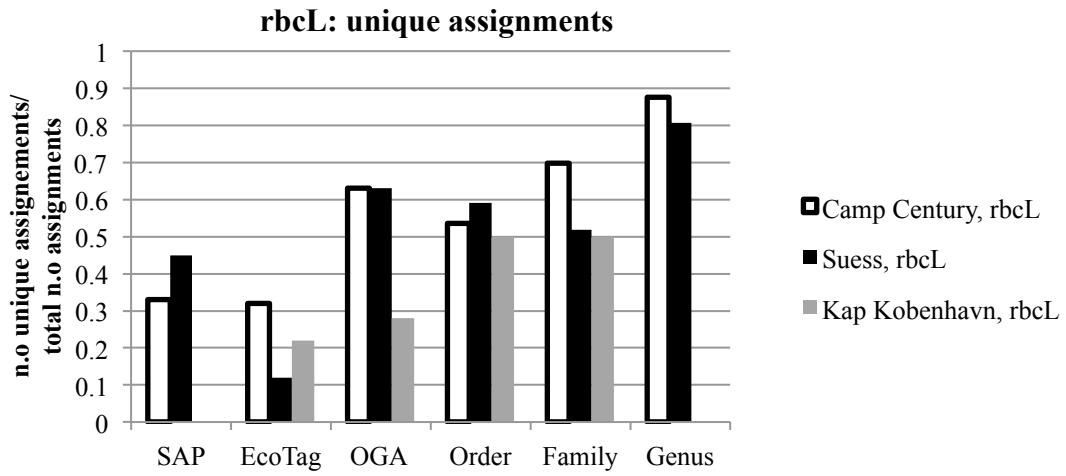


**Figure 7.** Summary of Table 18: unique assignment ratios per marker, site, and taxonomy level




There are various reasons for this:

1. The difference in the total number of assigned taxa at the three localities, with Kap København showing null results for many categories.
2. Differences across sites can also result from differences in the database coverage of the corresponding diversity making the assignment success more specific in some sites than others.
3. There is the possibility that for a site with a high biodiversity the difference in assignment between methods will be more obvious than for one with a low biodiversity where there are less taxa to assign too (compare Kap København with the other two sites for all methods, Table 18) – higher biodiversity might mean more rare variants, and the way the different approaches deal with this can be reflected in the uniquely assigned sequences number (although the overall average ratio is similar for the Camp Century and Suess sites and the two markers, see the differences in trend between trnL and rbcL for EcoTag and the other two approaches.

**Shared Assignments**

When evaluating the shared assignment success from table 19 and 20 and compare the three methods at the family level – where there is higher diversity than at the order level but fewer unique assignments than at the genus level – we find that EcoTag and

SAP had a higher shared assignment success than OGA for trnL, whereas for rbcL EcoTag has the highest number of shared taxa with OGA, and OGA and SAP sharing the same number. It therefore seems that EcoTag is a good support to either of the two methods for evaluating the assigned taxa from both a phylogenetic and similarity based point of view.

The two markers differed overall in the number of shared taxa but for each of the taxonomic levels, trnL assigned more taxa than the rbcL if we sum over all three sites, see table 19 and 20.

**Table 19**. Number of taxa found by more than one approach.

| Taxon level | Site | Marker | SAP | EcoTag | OGA |
|---|---|---|---|---|---|
| **Order** | **Camp Century** | trnL | 15 | 18 | 17 |
| | | rbcL | 13 | 6 | 11 |
| | **Suess** | trnL | 6 | 7 | 7 |
| | | rbcL | 8 | 7 | 9 |
| | **Kap København** | trnL | 7 | 6 | 5 |
| | | rbcL | 2 | 2 | 2 |
| **Family** | **Camp Century** | trnL | 30 | 32 | 26 |
| | | rbcL | 21 | 24 | 23 |
| | **Suess** | trnL | 8 | 9 | 12 |
| | | rbcL | 9 | 13 | 8 |
| | **Kap København** | trnL | 7 | 7 | 5 |
| | | rbcL | 2 | 2 | 1 |
| **Genus** | **Camp Century** | trnL | 24 | 30 | 22 |
| | | rbcL | 10 | 8 | 9 |
| | **Suess** | trnL | 4 | 1 | 3 |
| | | rbcL | 4 | 5 | 3 |
| | **Kap København** | trnL | 4 | 5 | 2 |
| | | rbcL | 1 | 1 | 0 |

**Table 20**. Listing the number of shared taxa across the three sites for the three levels, order, family and genus for rbcL and trnL.

| Taxon level | Marker | Assignments shared by at least two methods |
|---|---|---|
| **Order** | trnL | 32 |
| | rbcL | 24 |
| **Family** | trnL | **48** |
| | rbcL | **38** |
| **Genus** | trnL | 33 |
| | rbcL | 16 |

## DATABASES AND ASSIGNMENT CRITERIA

### Assignment to Databases

There is a general concern when using any of the methods which is related to how they find the most closely related reference sequences when assigning a sample sequences. The variance between methods in number of unique assignments suggests that a substantial difference exists in how well the three assignment methods find suitable reference sequences for a successful assignment.

For the phylogeny-based approaches (SAP and OGA), the initial set of reference sequences retrieved from the database by any of the programs when trying to assign one single query sequence have to contain both (i) a diversity of families, genera and orders, and (ii) the closest related sequences with the fewest differences to the query. Given the different ways that the assignment approaches use for the reference sequence retrieval, more pronounced differences in assignment success for the same site are expected when trying to assign sequences that only have distant groups represented in the database. Also, the assignment success could differ between sites depending on the degree of diversity between the samples.

EcoTag is only similarity-based and does not (i) account for reciprocal monophyly or reciprocal paraphyly or (ii) model the diversity present in the database and use this information for improving accuracy. EcoTag simply searches for best match at the sequence level. This means that given a query, if more than one target sequence in the database is equally alike, the assignment goes to the shared taxonomic level of these (e.g. if for one query there are two sequences in the database that match it equally well at the species level, the assignment is done at the genera level).

Both phylogenetic and similarity-only-based approaches have downsides. A similarity-based approach does not provide any information about other very similar sequences that might be informative for the assigned sequence. If the assignment match is 99% and the other sequences have 98% match but come from different genera it could be informative for how well supported the relatedness between the query and the reference sequence is. On the other hand, a phylogenetic-based approach can provide a high assignment score to a sequence even when it does not

have a high similarity with the query, as long as the other related database sequences are even less similar to the query.
Another general drawback is specific to reference sequence retrieval through blast searches – which are used by all three approaches. BLAST (Altschul, et al., 1997) only evaluates similarity through local alignment and to eliminate this error source it would be necessary to re-rank all possible reference sequences for each query sequence based on global alignment scores for each individual sequence. To compensate for this effect the Blast searches were corrected for sequence size.
Further, the two phylogenetic assignment methods use multiple global alignments to re-align the retrieved reference sequences, which are performed by ClustalW (Larkin, et al., 2007) in SAP and in OGA by MUSCLE (Edgar, 2004). In SAP the global alignments are used to base the Markov chain Monte Carlo (MCMC) tree sampling approach on (Huelsenbeck & Ronquist, 2001), which thereby accommodates an evaluation of all possible trees from the alignment and also models species not present in the database (Zhang, et al., 2011). In OGA multiple alignments are performed prior to the Maximum Likelihood tree sampling approach that evaluated a sub-set of all possible trees from the alignment by the use of PhyML (Guindon, et al., 2010).

Given all the differences between the methods, it is a good idea to test their assignment performances using known sequences to better understand the differences in the results. One straightforward way is to assign a subset of known database sequences against the database to see how often they are correctly assigned. This exercise can be done in two ways: (i) the database is tested against itself, (ii) the sequence being assigned is excluded from the database (to test the performance when assigning sequences to an incomplete database).
SAP was tested for specificity on a subset of the NCBI database in Munch et al. (2008a) with the level of correct assignment for trnL at family, genus and species level are 100%, 90% and 51% respectively for the full Bayesian mode when tested in 2008. For Cox1 the numbers were 99%, 99%, 90% for family, genus, and species. SAP has been made to deal with incomplete databases when being run in the intensive full Bayesian approach instead of the constrained Neighbour Joining approach (Munch, et al., 2008b).
We assigned the trnL and rbcL database against itself for the OGA method and the performance for trnL and rbcL the assignment success of 64.1% and 63.7% for the

two markers respectively (tables 21 and 22, figures 8 and 9). This test further showed that some orders contain families with too high sequence diversity to be assigned to anything but order level using this assignment approach.

Although it would be relevant to compare the performance of SAP and EcoTag on the present day full databases for all markers, these tests are still pending.

**Figure 8.** Assignment success of trnL database at the family level showing the names of most common represented families in the database. The largest green piece shows the proportion of unassigned families and they are presented at the order level below. Here the most common orders are named with the majority still unassigned at the order level.

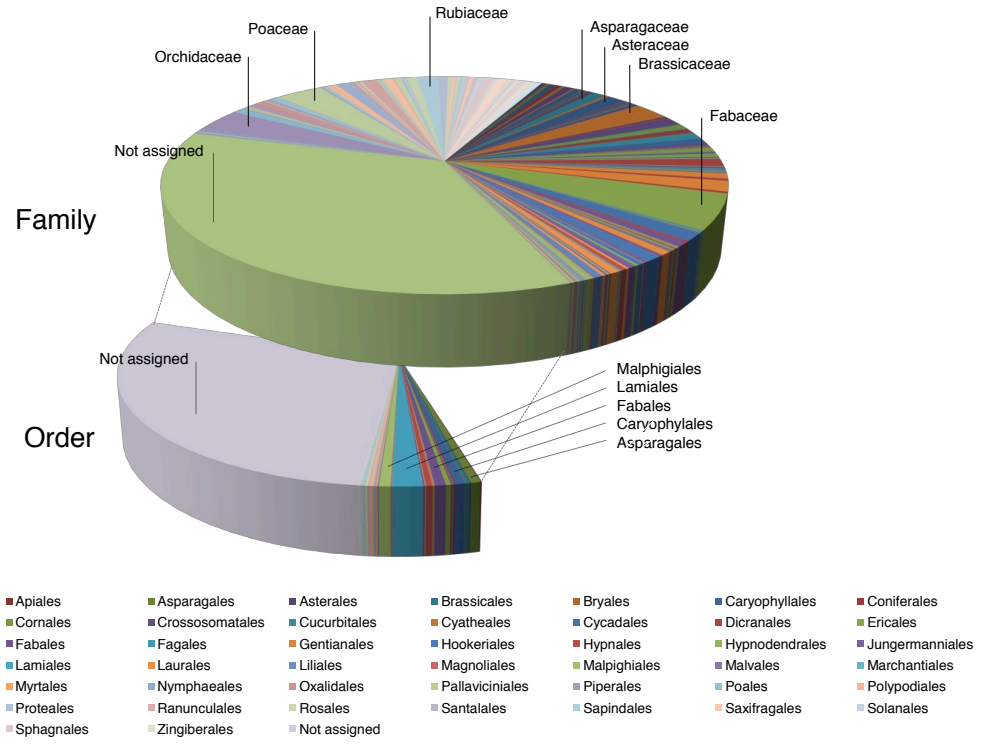



**Table 21**. Assignment sensitivity and specificity of the trnL marker at the four different taxonomic levels with the OGA method (on May 2011 database).

| Taxonomic level | Assignment Sensitivity (%) | Assignment Specificity (%) | False positives (%) |
|---|---|---|---|
| Order | 69 | 74 | 26 |
| Family | 64 | 75 | 25 |
| Genus | 41 | 97 | 3 |
| Species | 25 | 97 | 3 |

Assignment sensitivity is calculated as the proportion of assigned sequences to the total number of sequences in the database
Assignment specificity is the proportion of the assigned sequences that were correctly assigned
False positives is the proportion of incorrect assigned sequences

**Figure 9**. Assignment success of rbcL database at the family level that showing the names of most common represented families in the database. The largest purple piece shows the proportion of unassigned families and they are presented at the order level below. Here the most common orders are named with the majority still unassigned at the order level.

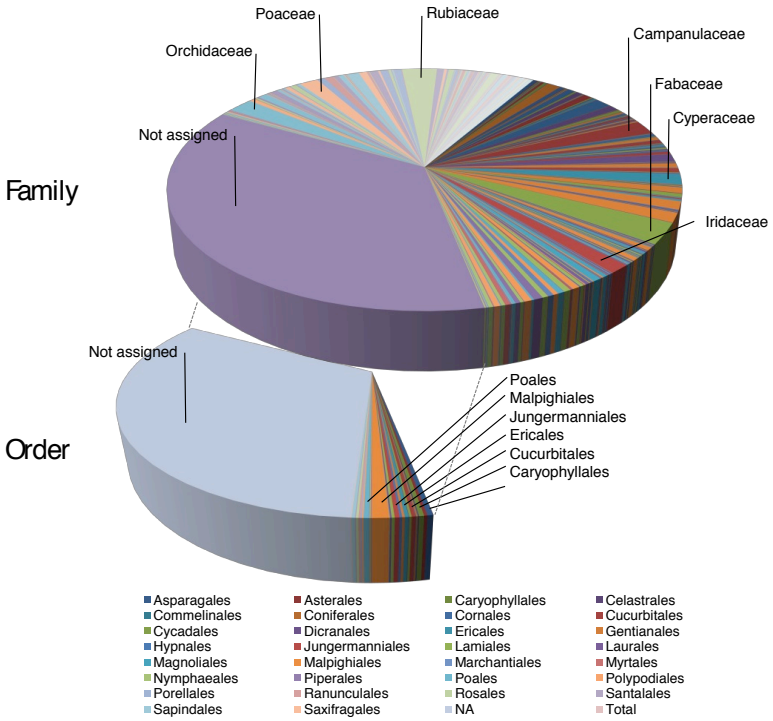



**Table 22**. Assignment sensitivity and specificity of the rbcL marker at the four different taxonomic levels with the OGA method (on May 2011 database)..

| Taxonomic level | Assignment Sensitivity (%) | Assignment Specificity (%) | False positives (%) |
|---|---|---|---|
| Order | 65 | 89 | 11 |
| Family | 64 | 87 | 13 |
| Genus | 45 | 99 | 1 |
| Species | 33 | 99 | 1 |

Assignment sensitivity is calculated as the proportion of assigned sequences to the total number of sequences in the database
Assignment specificity is the proportion of the assigned sequences that were correctly assigned
False positives is the proportion of incorrect assigned sequences

Above I had mentioned that in order to use OGA, one needs to optimize the number of mismatches allowed between query and reference sequences for different databases, since sequences can be wrongly assigned when the database does not contain a lot of diversity (in practice we see this from the high number of uniquely assigned sequences for rbcL relative to trnL). Given the results of the database against database exercise, we find that there is also room for an overall optimization of OGA at two levels: i) the sensitivity should be increased at least at the three higher taxonomic levels, even though we expect the sensitivity to decrease for genus and species levels as these mini-barcode fragments will not have enough information to differentiate all taxa and hence cannot provide a 100% sensitivity success, and ii) the specificity should be improved for the order and family level as the overall assignment success indicated by the sensitivity number is inflated by false assignments.

The level of correct assignments at the species level is low for both OGA and SAP (30% and 50%) and therefore we suggest not to interpret ecological biodiversity conclusions solely on species information when assigning unknown sequences against incomplete databases as a 100% species match might be equally likely with a variant not represented in the database and hence cause erroneous conclusions. Thus, we recommend reporting the genus level of a species assignment. For this study, we chose a conservative approach and only relied on assignments above species level.

Finally, I would like to discuss the assignment protocols when assigning short DNA fragments. Should one consider the phylogeny of short DNA fragments that do not necessarily reflect each species' evolutionary history, or instead evaluate the similarity of this fragment between taxa? The phylogeny approach is weakened by the fact that different genetic regions have different evolutionary rates in different taxa resulting sometimes in incomplete lineage sorting introducing a bias in the assignment process. On the other hand, a similarity-based approach might miss out on information that could distinguish between very similar but taxonomically different sequences and provide a statistical measure of confidence for a well or poor supported assignment. In summary, the combination of the phylogenetic-based and similarity-based methods is well-reasoned, since these provide independent estimates of the assigned biodiversity. Either method chosen should be tested for its sensitivity and

specificity to capture the diversity in the dataset.

**Criteria for Accepting Assignment Results**

We have decided to follow a conservative approach to reduce the risk of false positive assignments, and only accepted assignments for any downstream interpretation if they were found by at least two of the three assignment methods. This decreased the number of taxa found at the order, family and genus level to a larger degree for the rbcL marker than for the trnL marker. By following such a conservative approach the cost for removing the unique assignments and possible false positives is that we might remove some true diversity that is found better by one of the methods. However, simulations have not yet been performed that test for the impact of how much of the informative biodiversity is removed by such a conservative approach. We therefore chose the strategy that reduced the level of false positives, although at the cost of removing true diversity.

**Inevitable Bias from Incomplete Databases**

In this work, there is a common problem to the assignment performed by all approaches – all assignments were done against reference sequence databases that do not represent the complete tree of life, and therefore do not cover all of the biodiversity present in the sampled environments. When a taxa group is absent in the database, it will either (i) produce a false assignment of a query if neither the correct or closely related taxon is represented in the database, (ii) produce an assignment to a less specific or general taxonomic order or (iii) will not produce an assignment at all (Munch, et al., 2008a). The representation of the tree of life in GenBank (http://www.ncbi.nlm.nih.gov/) is increasing by the day, and with time this should decrease the problems related to the incomplete diversity. The trnL database have been sampled thoroughly for arctic and boreal species as part of a larger barcoding study (Taberlet, et al., 2007) but has a large global representation too. The rbcL marker has also been used as a barcode but sampled less thoroughly. This implies that the results obtained in this study for trnL should be more representative of the assignment approaches true potential relative to those of rbcL.
There is yet another limitation, specific to the successful assignment of ancient DNA

sequences, which is the fact that all sequence information is contained in short DNA fragments. Even though the markers used here are known to target areas of high variation between taxa, there is a difference in the degree of sequence divergence within some families or orders. Low divergence can complicate the assignment, which is illustrated by the fact that OGA is not able to assign at family level for orders as e.g. Asparagales, Malphigiales, Lamiales and Fabales, containing families with low divergence (Figures 8 and 9). However, for these particular orders the other two methods were able to assign to families level, again reinforcing the idea that one should use more than one method.

## CONCLUSIONS AND PERSPECTIVES

We have presented results from three different denoising and assignment approaches of short DNA sequence amplicons. By denoising high-throughput sequence data from the 454 FLX platform, we removed a large part of the sequences that was produced by amplification, sequencing error as well as damage and contamination. The three different denoising approaches differ in how they remove sequences. The filtering and assignment approaches for OGA and SAP could be compared at the post-filtering step and showed that we are removing some assignable diversity by this conservative approach.

The accepted diversity was then assigned by three methods: two phylogeny-based ones, SAP and OGA, and the similarity-based EcoTag. The methods showed differences but also overlap in the total level of assignment success. SAP and EcoTag and their respective denoising pipelines provided more conservative results than the OGA approach as seen for the proportion of unique assignments (this difference was most evident for the rbcL marker, see Table 18).

By following three independent pipelines of denoising and assignment we get a thorough evaluation of the data. In order to compensate for possible amplification, sequencing and damage errors that would inflate the biodiversity present in samples used for meta-barcoding studies we have combined three different denoising and assignment pipelines and only accepted results confirmed by two of three pipelines. This strategy involves a risk of removing true variants and does not guarantee the

removal of erroneous diversity. Hence, to avoid a signal from errors that would cause false positives we only kept results that were found by two of the three methods. The use of this conservative approach in this study is even more relevant as the material found in these samples is of very old age.

By combining a phylogenetic and similarity-based assignment approach the biodiversity is verified by two independent methods that strengthens the support for the conclusion. Additionally, we obtain both an estimate of the similarity score but also the support score from the phylogenetic method.

Assigning sequences against a reference database will always be influenced by the incompleteness of that database as not all diversity is represented. Also, the sequence data present in the database is of present day taxa, and this might influence the assignments of ancient sequences, although this influence can be reduced by using markers with low mutation rates as we did in this work.

To evaluate the methods more thoroughly all combinations of each of the denoising approaches with the three different assignment approaches need to be tested for consistency.

Considering the phylogenetic-based methods, SAP is a lot more computationally demanding than OGA, taking around 50% more time to assign the same dataset. However, in order to substitute SAP by OGA, the latter needs further optimization for achieving the robustness of SAP.

## FINAL REMARKS

This chapter presented the criteria used when accepting a sequence assignment from the different computational approaches, EcoTag, SAP and OGA, to characterize the ancient molecular diversity from the ecosystems once present in the three localities, Camp Century, Antarctica and Kap København. A combination of the results from all three was considered as a conservative approach, but a necessary one to restrict the number of false positives that can occur given the limitations of datasets, databases and assignment programs. Also, combining the results from phylogeny-based

approaches with similarity-based ones should strengthen the assignments results, given the difference in sequence assignment implementation. Optimization of the filtering steps should also be accounted for, given that true diversity might be lost at this step. The real risk of erroneous assignment of filtered sequences should be assessed in order to make decisions on the filtering strength. The taxonomic results will be discussed in the following three chapters for each site.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**: 3389-3402.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792-1797.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W & Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* **59**: 307-321.

Huelsenbeck JP & Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754-755.

Larkin MA, Blackshields G, Brown NP, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Lindahl T & Nyberg B (1972) Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **11**: 3610-3618.

Munch K, Boomsma W, Willerslev E & Nielsen R (2008b) Fast phylogenetic DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363**: 3997-4002.

Munch K, Boomsma W, Huelsenbeck J, Willerslev E & Nielsen R (2008a) Statistical Assignment of DNA Sequences Using Bayesian Phylogenetics. *Systematic Biology* **57**: 750-757.

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

Taberlet P, Coissac E, Pompanon F, et al. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research* **35**: e14-e14.

Zhang AB, Muster C, Liang HB, et al. (2011) A fuzzy-set-theory-based approach to analyse species membership in DNA barcoding. *Molecular Ecology*.

# INTRODUCTION TO CHAPTERS ON PALAEO-ENVIRONMENTS IN GREENLAND AND ANTARCTICA

The common feature about the palaeo-environments of Greenland and Antarctica prior to ice-coverage is that there are uncertainties about when they last existed, how abrupt the ecosystems were changed into the polar ice covered areas we see today and what kind of ecosystems they represented prior to ice coverage. The fossil record dates back to the time prior to ice coverage, which differs substantially between the two places and have been found at the margins of the ice sheets. In Antarctica the majority of fossils represent time periods with tropical to boreal climates from the Permian to the Eocene (Francis, et al., 2007b) and only a single site has fossils indicating a tundra-like environment (Francis, et al., 2007a). In Greenland the fossil record is even more limited with fossils dating back to the late-Pliocene/early-Pleistocene where the Greenland ice sheets is believed to form (Funder, 1989). The time period when Greenland and Antarctica changed from forest-covered areas with diverse ecosystems to the ice covered polar ecosystems we find today is debated for both areas. So even though Greenland and Antarctica differ in the amount of fossil material available they share the lack of fossils representing a more recent time period from the late Pliocene and through the Pleistocene that would aid the understanding of climate fluctuations, dynamics of ice sheet coverage and the ability of vegetation to re-establish after glacier-retreat. In Antarctica it is generally believed that the formation of The Antarctic Ice Sheet (AIS) began with a drastic cooling around 35MY (Zachos, et al., 2001) but how stable the ice sheet has been since the Miocene and how much it has fluctuated in size and distribution through warm periods has been the source of a heated debate (e.g. (Webb & Harwood, 1987, Sugden, 1996).

We do not know exactly how the change occurred from diverse ecosystems to ice sheet covered landscapes and if it was abrupt or gradual, which is relevant for climate models and for getting a better understanding of how changes might occur in the future. Access to fossils representing a wider geographical scale is difficult due to the large ice sheets. Several ice cores have been drilled from ice caps in both hemispheres and of these some reaching all the way to bedrock (figure 1) (Gow & Meese, 1996, Bentley & Koci, 2007). Such cores contain basal ice that holds debris from the land below the ice that could potentially hold biomolecular information from the past vegetation existing in the areas the ice have slided across through time. Such information is relevant for understanding more of the dynamics of the once ice free landscape and how some areas could have been ice free while the ice caps retreated in warmer periods.

By using ancient DNA preserved in the soil from basal ice in glaciers or ice cores we get access to areas not investigated previously. If the molecules are preserved from the past terrestrial ecosystems in the soil we can add further knowledge and data to the biological record from these areas.

This study focuses on three different polar sites (see figure 1). Ice covered sites presenting basal debris rich ice from a Greenland ice core, Camp Century and from two Antarctic glaciers in the Dry Valleys, and one ice-free but permafrozen area from the Kap København formation of Northern Greenland.

From these three sites we test the idea that the basal sections hold frozen ancient DNA from the past flora and fauna and thereby act as a frozen archive of molecular fossils. Each of these sites will be presented individually in the following sections.

**Figure 1**. Maps showing locations of deep core drilling sites in A, Antarctica and B, Greenland. Red stars show sample sites for this study. (Figure modified from from Bentley and Koci (2007)

A.


B.

References

Bentley CR & Koci BR (2007) Drilling to the beds of the Greenland and Antarctic ice sheets: a review. *Annals of glaciology* **47**: 1-9.

Francis J, Ashworth A, Cantrill D, et al. (2007b) 100 million years of Antarctic climate evolution: Evidence from fossil plants. *Antarctica, a keystone in a changing world. Proceedings of the 10th International Symposium on Antarctic Earth Sciences, Santa Barbara, California* 19–28.

Francis JE, Haywood AM, Ashworth AC & Valdes PJ (2007a) Tundra environments in the Neogene Sirius Group, Antarctica: evidence from the geological record and coupled atmosphere-vegetation models. *Journal of the Geological Society* **164**: 317-322.

Funder S (1989) Quaternary Geology of the Ice-Free Areas and Adjacent Shelves of Greenland. *Quaternary Geology of Canada and Greenland* **K-1**: 741-792.

Gow A & Meese D (1996) Nature of basal debris in the GISP2 and Byrd ice cores and its relevance to bed processes. *Annals of glaciology* **22**: 134-140.

Sugden DE (1996) The East Antarctic Ice Sheet: unstable ice or unstable ideas? *Transactions of the Institute of British Geographers* 443-454.

Webb PN & Harwood DM (1987) The Sirius Formation of the Bearmore Glacier region. *Antarctic journal of the United States* **22**: 8-12.

Zachos J, Pagani M, Sloan L, Thomas E & Billups K (2001) Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* **292**: 686-693.

# Chapter V

TITLE:

A forest-meadow palaeo-ecosystem in northwestern Greenland recovered by ancient metagenomics of the Camp Century Ice Core

ABSTRACT

Ancient DNA has previously been detected in the basal ice from the Dye 3 deep ice core from southern Greenland and used to reconstruct a rich boreal ecosystem based on a limited number of cloned DNA sequences from plants and insects. Here we present an extended study of deep sequenced ancient DNA amplified from basal ice in the Camp Century ice core from northwestern Greenland. Camp Century was drilled on the margin of the supposedly reduced Greenland Ice Sheet during the last interglacial period. Our results show that the area remained covered by ice during the previous interglacial, the Eemian, 130-116KY. In addition, we find that the area was last free of ice in a period that likely extends further back in time than MIS 11. The aDNA assigns to a relatively diverse biota with boreal to northern temperate taxa, suggesting an ecosystem with forest-meadow landscape with mountains. To be able to provide a climate warm enough to support such an ecosystem we believe this time period coincided with the Dye 3 location and possibly extends back to approximately one million years ago. The biodiversity from Camp Century share plant taxa with both the fossil findings from the Kap København formation and with taxa from the Dye 3 site, like e.g. *Taxus*. Temperature constraints for the Camp Century palaeo-climate can be deduced from indicator taxa found among the genetic biodiversity. These suggest average summer temperatures above 10°C and a growth season of four months to support the existence of yew, cypress and oak. Last, we open up for the possibility that the period where forest occurred in both southern and north western Greenland could have co-existed with the Kap København formation prior to the full coverage of the Greenland Ice Sheet sometime around 0.9-1MY.

# INTRODUCTION

Traditionally, palaeo-ecosystems are described by fossil findings that represent the past biodiversity of a site. However, in ice-covered areas access to fossils is severely limited by the presence of ice sheets or glaciers. In North Greenland north of the distribution area of the Greenland Ice Sheet (GIS), fossil findings from the Kap København, formation have supported a hypothesis of a forest covered Greenland prior to the origin of the Greenland Ice Sheet (GIS) (Funder, 1989, Bennike & Bôcher, 1990). However, fossil evidence from other regions of Greenland to support and expand upon such a hypothesis has been very limited, until the advent of molecular studies on the basal ice from the Dye 3 ice core in southern Greenland (Willerslev, et al., 2007).

The time when the GIS originated is debated and have been suggested to coincide with the cooling event that started in late-Pliocene/early-Pleistocene around 1.8 – 2.5 million years ago (MY)(Shackleton, et al., 1984, Funder, 1989), while others suggest it to become stable around 0.9-1MY ago when the glaciation cycles changed from 40,000 year (KY) intervals to the more inert 100KY intervals (Bintanja & van de Wal, 2008, Pollard & DeConto, 2009).

Isotopic studies of the deep glacier ice and basal debris-rich ice from the GRIP and Dye 3 cores have demonstrated that the GIS contains ice from very different time periods that suggests that the ice sheet has fluctuated considerably in size through time (Souchez, et al., 1994, Willerslev, et al., 2007). During such dynamic fluctuations, an advancing ice sheet can incorporate relict non-glacial ice already present in the landscape at its base (Souchez, 2000). The main reason why the age difference between glacial and basal ice is sometimes substantial is due to the processes occurring at the base of the ice sheet. At the GRIP deep drill site, it was shown from isotopic measurements that the basal ice originated from a previous glacier or ice sheet that was overridden and incorporated into the new GIS where the overlying glacial ice dates back to the Eemian, 130 thousand years ago (KY) (Souchez, et al., 1994). However, it is a general trend that basal ice is very difficult to date as it is mixed with debris and often has a non-continuous profile with the overlying ice (Aciego, et al., 2011).

Entrainment of debris in the basal ice makes it interesting from a palaeo-ecological and palaeo-climatic point of view, as the debris or soil can contain preserved ancient DNA from past vegetation if temperature conditions have remained below freezing. The Dye 3 study found evidence of a diverse forested ecosystem in Southern Greenland from the short fragments of preserved ancient terrestrial DNA in the basal ice (Willerslev, et al., 2007). By applying four novel dating methods to the basal ice, an age of both the basal ice and the ancient biomolecules could be determined to span a range of 450 – 800 KY, which is older than the late Eemian age of the overlying glacial ice (Willerslev, et al., 2007). Therefore, genetic assessment of preserved biomolecules in the basal ice can, in combination with dating methods, supplement models for possible scenarios of ice coverage and ice sheet dynamics, and, redefine the past climate from the flora and fauna detected. Moreover, it will add information about the distribution and composition of taxa groups in high latitude areas at a time that they were previously ice-free and supported a much warmer environment than what is known today. Such information is relevant for inferring adaptation potentials and biodiversity dynamics of flora and fauna from ice-covered areas that have not been studied previously in high latitude locations.

If temperatures and physical conditions have allowed the incorporation of degraded ancient DNA (aDNA) from past ecosystems in other deep ice cores from Greenland, aDNA methods can be employed to investigate the past ecosystems in areas that are today covered by thick ice sheets.

In this study, we applied ancient genetic methods on basal ice from the Camp Century ice core (see Figure 1) to test if ancient terrestrial biomolecules are preserved in the basal ice of Camp Century and determine the kind of ecosystem they represent. Another aim of the study was to test if the ancient genetic data could reveal more information about when the Greenland Ice Sheet (GIS) built up in this region and how it might have fluctuated in size since its original formation.

METHODS:

**Dating methods:**

Basal ice is generally challenging to date due to basal processes (like e.g. sliding across land surfaces, layer folding and entrainment of debris) mixing up the layers, and no well-established dating methods exist for such an undertaking. We, therefore, applied three different approaches to determine an age range that spans the age of the ice and the time when the soil and DNA got incorporated into the ice. To be able to cover such an age span of the ice we chose the methods for their ability to cover both a minimum age (of dust or debris) and a maximum age (time of build-up of precipitation into glacier). Age estimates of the ice and the entrainment of silt material is considered to represent the minimum age of the DNA, as the DNA got adsorbed into the soil before the environment cooled and the soil was entrained into the ice sheet. Finally, these methods were chosen as they had all been tested on ice material previous to this study (Willerslev, et al., 2007, Aciego, et al., 2009). Methods that providing minimum ages of the DNA that were used in this study were: Optical Stimulated Luminescence dating (OSL), Uranium ($^{234}U/^{238}U$) recoil dating and cosmogenic 10Be/36Cl dating. See Supplementary Material for detailed descriptions of the methods.

**Sampling and DNA extractions**:

We sampled 10 different samples in the basal ice sequence of the Camp Century ice core and 2 glacial clean ice samples further of the ice core sequence. The basal ice samples were extracted in multiple replicates if enough sample material was available. See chapter 3 for sample details, decontamination and extraction protocol.

The sampling and DNA extractions were carried out in accordance to the protocols described in Chapter 3.

**DNA Amplification and Sequencing**:

Short DNA fragments from plants (trnL/rbcL, cpDNA), insect), insects (Cox1, mtDNA) and mammals (16S, mtDNA) were PCR amplified and sequenced from the basal ice DNA extractions as described in Chapter 3.

12S bird PCR amplicons of the basal ice DNA extractions were generated as 2-3 replicates per sample. The PCR products representing each sample were pooled, cloned using TOPO® TA Cloning® Kit following manufacturers protocol and sequenced by the Macrogen sequencing facility

Clean glacial ice was PCR amplified with the universal plant trnL and rbcL primers to investigate if the plant DNA could be of windborne origin and hence not represent an indigenous environment from the period prior ice-coverage. See chapter 3 for sample details.

**Assignment Analyses:**

Three pipelines described in detail in Chapter xx were used to filter and assign the plant sequences amplified for the trnL and rbcL markers.

The Cox1 (insect) and 16S mammalian sequences were filtered and assigned according to the same criteria as described for the plant markers in Chapter 3 and 4, although only by two programs, EcoTag, available at www.prabi.grenoble.fr/trac/OBITools), and SAP (Munch, et al., 2008a).

The 12S bird cloned sequences were analyzed by the SAP assignment program following the full Bayesian approach (Munch, et al., 2008a).

**Assigned Presence-Absence Data**:

We generated taxonomic diversity tables for each of the taxonomic levels (order, family and genus) for the two plant markers, rbcL and trnL. Tables were designed to illustrate the diversity found per sample and method. The taxonomic results are presented as presence-absence data and the colours represent the markers that

identified the different taxa. For the plant markers, green colour indicates where there is an overlap between the trnL and rbcL markers and grey and black colors where trnL and rbcL identify different taxa, respectively. The taxa amplified in a subset of samples (1049A, 1052A, 1054B, 1057A), which were sent for independent replication at the clean lab facilities in Murdoch University, Perth, Australia, are marked with asterisks (*). Furthermore, asterisks were only applied to taxa if they were found in both the Murdoch and Copenhagen clean labs. The aforementioned presence-absence data are illustrated in Tables 1-3.

EcoTag and SAP assignment programs were used to assign the 16S and Cox1 markers and results are discussed below.

**Pollen Analysis**:

A basal ice sample from tube #1052A corresponding to depth 1357m was melted for pollen analysis and performed in dedicated lab facilities at the Geological Survey of Canada. Analyses, results and methodological details are outlined in Appendix C (J. Bourgeois, unpublished).

## RESULTS AND DISCUSSION

The first aim of testing for amplifiable aDNA was positive and we therefore have three possible contamination based scenarios to explain the aDNA findings;

If the results were not representing a local environment the contamination could be **a)** sample based, **b)** laboratory based, or **c)** dispersal based from windborne material.

Three approaches were followed to test for the three contamination sources.

**(a)** Sample based contamination could arise by penetration of modern DNA through cracks in the core to the inner parts used for DNA extraction. This was tested for by applying known bacterial vector DNA to the surface of all ice samples and after incubation removing *ca* 1.5 cm of the surface layers. The vector was then amplified in both the surface layers and extractions from the inner core. Samples that proved positive for vector DNA in surface layers and negative in inner core were accepted as free of sample based contamination by modern DNA and used for amplifications by flora and fauna primers.

**(b)** The laboratory-based contamination was tested for by having parallel amplifications performed of a sub-set of samples in Murdoch University clean lab facilities, Perth. These were able to reproduce findings of northern hemisphere indicator taxa of boreal to northern temperate climate zone, indicative of a sample specific and not laboratory dependent signal. Moreover, all samples were reproduced by multiple amplifications where arctic and northern temperate taxa not commonly detected in Denmark were found.

**(c)** Dispersal based contamination from pollen source or other windblown sources was tested for by first extracting DNA from large amounts of clean glacial ice (see chapter 3). Then we attempted PCR amplification with the plant markers to test for results of the windborne pollen present in the clean glacial ice. No positive amplification results were found following the same procedures as for the basal ice samples.

**Plant diversity in Presence-Absence Data**

Initial inspection of the presence-absence data of the plant markers showed that there was a large discrepancy in some of the taxa groups picked up by the two different markers (see Appendix A). The rbcL marker is assigned to a much smaller database thant the trnL sequences and do not have as high a representation of arctic and boreal species as the trnL marker that have specifically used in a the EcoChange project sequencing almost a full coverage of all arctic and boreal species existing today. Therefore it is not surprising that we get many more assignments by this marker to sub-tropical to tropical indicator taxa as opposed to the trnL marker that only picked up a few of the warm climate indicator taxa. If we do not have a good global representation of taxa in the rbcL database and do not have a good match to the unknown more cold adapted diversity in the samples, we might get assignments to distant related sequences because they are the closest match. If a sample sequence has just a few mismatches to a tropical taxon in the database we would get a good but false positive assignment to such warm adapted taxa. This would influence the interpretation of the diversity significantly. We ascribe this to be due to both an assignment-based problem and database problem when rbcL sequences are assigned to very different taxa than the trnL sequences. We would expect the markers to find the same taxa overall, or at least taxa within the same climate range given that the rbcL database contains only about a third of the diversity representative in the trnL database. Therefore, we accepted results where both the trnL and rbcL markers found the same taxa and where trnL taxa were assigned by more than one method in one or several samples.

The results from the Murdoch lab were used for validating the results and climatic signal in the diversity assigned from the Copenhagen results. This way we are not presenting results from the Murdoch lab that do not have an overlap with the Copenhagen results. Common food contaminants like soya (*Glycine*), *Cannabaceae* (beer production) and nightshade (potato/tomato) (*Solanum*) were also excluded as representative of the true diversity in the samples and not presented in the presence-absence tables but can be found in Appendix A. The presence-absence diversity data is listed in Tables 1-3 and were used for the ecological and climatic interpretations.

The diversity of plant taxa in the reduced set shows that arctic flora is represented by *Oxyria* and *Dryas*, and cold temperate flora by the taxa *Betula* and *Bistorta*. Furthermore, we have indicator taxa like *Picea*, *Pinus*, *Euphorbiaceae*, *Populus*, *Urticaceae* and *Persicaria* that can occur in boreal zone but do not tolerate a climate colder than "cold temperate". *Chamaecyparis*, *Anacardiaceae*, *Sapindaceae*, *Taxus*, *Quercus*, *Juglandaceae*, *Oleaceae* and *Passifloracea*e represent taxa that occur a climate zone that is at least temperate or boreo-nemeral forest. We only find one indicator taxon for a tropical to sub-tropical climate, namely *Dalbergia*. *Dalbergia* belongs to the *Fabaceae* family and has once been found in Greenland during the Cretaceous where the climate in Greenland was much warmer. However, as *Dalbergia* was only found in one sample and with a low sequence occurrence, we chose not to place a high emphasis on this taxon warm climate indicator.

**Flora and Past Climate**

The diversity of taxa (all listed in table 1-3) representing the different climate regions is as follows:

Arctic - *Oxyria* and *Dryas*
Arctic and cold temperate - *Betula* and *Bistorta*
Cold temperate boreal (not colder than subarctic) - *Picea*, *Pinus*, *Euphorbiaceae*, *Populus* and *Urticaceae*
Temperate and boreal - *Chamaecyparis*, *Taxus*, *Quercus*, *Fagaceae*, *Juglandaceae*, *Oleaceae*, *Passifloraceae*, *Anacardiaceae* and *Sapindaceae*

The Köppens Climate Classification Map in Figure 2 illustrates the different climate zones with their geographical distribution. The majority of above-mentioned taxa are today found in areas that are not colder than northern temperate climate (Zone Cfb in Figure 2). These taxa are typical boreal representatives and we therefore infer the DNA to be preserved from a past-forested environment. Moreover, with a strong representation of genera like *Betula*, *Chamaecyparis*, *Picea*, *Pinus*, *Quercu*s and families like *Juglandaceae* (that contains *Juglans* and *Pterocarya*, where *Juglans* did not pass the assignment thresholds), Oleaceae (that contains the genera *Fraxinus*),

*Anacardiacea*e, and, *Sapindaceae* (that contains *Acer* and *Hippocastanaceae*, which were found but did not pass the assignment thresholds), we infer further support for the presence of a forested ecosystem.

A few of these taxa can occur in the boreal/taiga zone that is characterized by the Walter Heinrich classification scheme (Walter & Breckle, 2002) and by the Whittaker diagram (Whittaker, 1975) as having annual average temperatures below 5°C and annual average precipitation of 40-100 cm/year (see Figure 3). Nevertheless, the majority of the taxa we detect in our samples are more typical in the temperate-nemoral zone that is found in eastern North America, Europe and eastern Asia and characterized by large seasonal changes between summer and winter with a temperature range of 31°C, annual average precipitation around 81cm and positioned in the zone where polar and tropical air masses meet that can produce large daily temperature changes (Rubel & Kottek, 2010). *Chamaecyparis* has a widespread temperate to boreal distribution and can be found in areas such as south east Alaska and in Hokkaido (Japan) with annual frost and snow fall in winter and spring but is otherwise more often found in the temperate zone. Both trnL and rbcL markers assign to this genus with zero mismatch for the trnL fragment and with one or zero mismatch for the longer rbcL fragment. For the rbcL fragment, *Chamaecyparis* assigned sequences differ by either one or two mismatches to *Juniperus* sequences. In addition, *Chamaecyparis* is found today at its most northern distribution in Japan and at latitudes up to 60°North in south-east Alaska (C. Nootkaense) so is extending its distribution into northern temperate/boreal climate zone and tolerates winter freezing with high snow fall (Carrara, et al., 2007). Some of the *Cupressaceae* sequences are not assigned below the family level and could be from e.g. the *Juniperus* or *Thuja* genera. The *Thuja* sequences has 2 mismatches to both the *Chamaecyparis* assigned sequences originating from the rbcL and trnL fragments and was previously found in the Kap København formation as a macrofossil. *Thuja* is another indicator taxa for a boreal/northern temperate climate, although more cold tolerant than *Chamaecyparis* (Funder, et al., 1985).

Overall, the presence of *Cupressaceae* together with *Taxus* and *Quercus* suggests a humid cool climate in winter with snow and frost and with a growth season of at least 120 days in the summer with temperatures above 10°C, suitable for deciduous tree growth (Funder, et al., 1985, Walter & Breckle, 2002). A northern temperate/nemoral climate zone is characterized by a moderate climate with freezing and frost resistant

deciduous temperate forest (Walter & Breckle, 2002). The nemoral zone in eastern North America today contains a higher biodiversity than what is found in Europe and consists of both deciduous and needle-leaved forests with an understory of shrubs and herbs (Ricklefs, 2008), very similar to the composition of taxa we find. Nevertheless, there is a paradox in that the arctic taxa findings of *Dryas* and *Oxyria* cannot co-occur with the temperate-nemoral taxa unless these were from a higher elevation area than the temperate taxa. Such a scenario is possible if glaciers overrode a varied landscape surface from mountain slopes over lowland areas and incorporated sediment particles containing the aDNA into the ice during such movements. An example of a present day environment with temperate-nemoral forest and alpine flora on the higher altitude mountain slopes is the northern Appalachian Mountains in eastern North America (Barbour & Billings, 2000).

Groups of taxa that have not been mentioned as indicator taxa but with representatives in the arctic flora are *Asteraceae*, *Brassicaceae*, *Caryophyllaceae*, *Cyperaceae*, *Crassulaceae*, *Equisataceae*, *Fabaceae*, *Juncaceae*, *Papaveraceae* and *Ranunculaceae*. However, these taxa have more representatives in the temperate zone compared to the arctic or boreal zone (Hartvig, personal communication). Most of these flowering plants are light sensitive and normally found in open patches of grasslands, which suggests that besides a mixed deciduous and coniferous forest the environment has also provided open areas where vascular plants like grasses, herbs and flowers could thrive. This is similar to what is found in alpine areas with meadow forested ecosystem like the northern Rocky Mountains and the northern parts of the Appalachian Mountains in North America or the northern temperate zone of Europe from the Alps to central Sweden (Barbour & Billings, 2000, Walter & Breckle, 2002). If we evaluate the consensus signal of where these taxa occur today, they fall within the nemoral ecozone, characterized by a mean annual precipitation that varies between 100mm and 200 cm. Average annual temperature can fall between 5 – 15°C as the summer and winter temperatures are highly seasonal and depending on a continental or coastal locality can differ substantially (Whittaker, 1975).

**Faunal DNA**

The insect and mammalian fauna was in much lower abundance than the plant sequence data, probably because of a lower initial biomass compared to the plants. Three samples (1052A, 1054A, 1057A) contained insect sequences but only 1052A and 1054A passed the assignment thresholds. The sequence diversity that was assigned covered two orders Diptera and Hemiptera, with the families *Culicidae* and *Aphidida*e, respectively. *Aphidida*e is a family with thousands of species known as plant pests and is widely distributed. Therefore, it does not provide a lot of climatic information, but instead supports the existence of a vegetated ecosystem. *Culicidae* is another large family that contains mosquitoes, and, if these represent the bloodsucking species that occur today in the nemoral zone with moist climate, then we would expect to find larger animals in such a habitat.

The mammalian 16S sequences were highly contamination-prone to human DNA and after filtering for these we found the following assignments to the family level – the *Bovidae*, *Muridae* and *Suidae*. However, only mouse and pig were assigned to genus level. All these three mammal families are known to occur as common laboratory contaminants from the deoxynucleoside triphosphates (dNTPs) used in the PCR setup. The dNTPs are obtained from the production of deoxynucleoside monophosphates by hydrolysis of animal DNA and then phosphorylated chemically to produce triphosphates. Therefore, incomplete hydrolysis of animal DNA during dNTP production may lead to dNTP contamination (Leonard, et al., 2007). All the mammalian groups were therefore not interpreted as an *in-situ* sample signal but ascribed as contamination.

Lastly, it should be mentioned that the 12S bird DNA from sample 1054B was assigned by SAP to the *Anatidae* family that contains geese, swans and ducks. All three of these taxa occur in boreal and temperate zones and Canadian geese are even found seasonally in the high arctic in the Thule area of Northern Greenland today (Bennike, personal communication).

**Pollen and DNA**

The pollen record found in the basal ice sample overlaps by 41% with the DNA findings, see Table 4. Analyses of the glacial clean ice further up in the core (Fredskild & Wagner, 1974) showed a 50% lesser pollen concentration compared to the basal ice, yet some of the taxa groups identified were the same. *Graminae*, *Acer* and *Artemisia* were only found in the glacial ice. *Ambrosia* and *Spaghnum* were found in both the glacial ice and in basal ice core, prior to decontamination, suggesting that this was windborne contamination. The rest of the common taxa found in both the decontaminated basal ice and glacier ice represented *Alnus*, *Betula*, *Castanea* (*Betulaceae*), *Juglandaceae*, *Pinus*, *Quercus*, *Tsuga, Ulmus*, *Cyperaceae* and *Urticaceae*. Taxa that were additionally found in the basal ice but not in the clean ice were *Fraxinus*, *Juglans*, *Carpinus, Larix*, *Picea*, *Salix*, *Caryophyllaceae*, *Dryas*, *Ericaceae*, *Oxyria, Papaver*, *Poaceae*, *Ranunculaceae*, *Asteraceae* and *Brassicaceae*.

aDNA from all of the samples overlaps with the pollen flora but exceeds the pollen floral diversity, even within the same sample tested for both (1052A). Some of the indicator species for a northern temperate environment only found by aDNA include *Taxus, Cupressaceae, Fabaceae* and *Passifloraceae*. Furthermore we see from table 4 that DNA picks up more diversity than what could only have been identified by pollen and still shows the climate signal inferred above. Nevertheless, we cannot rule out that some of the pollen originated from distant pollen sources, as they were found in the glacial ice, although we still do not expect a strong DNA signal from the pollen due to the following reason; We did not amplify plant DNA from the large volumes of glacial ice studied that should contain on average between 40-100 pollen per sample (Fredskild & Wagner, 1974) as compared to the basal samples that on average contain around 20 pollen per sample (due to smaller sample size). Another recent study investigated the local signal of sedimentary aDNA from frozen lake cores with regards to pollen and found that the DNA signal was local, and not originating from the pollen record (Jørgensen, et al., 2011). Hence, we overall considered the aDNA to be representative of what existed as active biomass in the soil from the ancient ecosystems that existed around the Camp Century site.

Furthermore, we consider the biodiversity detected as non-contaminant DNA because we detect taxa that are not common species in either Denmark or West Australia, and

were reproduced in both clean labs. These involve e.g. taxa like *Oxyria*, *Dryas* and *Taxus*.

In the current study, we would expect the signal to be more regional than local due to the basal ice processes occurring during ice sheet mass movements across the land surface (Knight, 1997). Here, soil or debris is entrained at the base and subsequent folding of the basal layers in the ice sheet can cause material from different regions or areas to be in the same ice core sequence (Souchez, 2000). Hence, incorporation occurs in a very different manner to how sedimentary cores are layered.

**DNA preservation**

DNA preserved *post-mortem* in basal ice is expected to undergo the same kind of degradation as seen from other studies on coprolites and permafrost (Poinar, et al., 1996, Lydolph, et al., 2005, Hansen, et al., 2006). However, the rate and to which degree this DNA decay occurs, is expected to be reduced when temperatures are cold and environment dry (Lindahl & Nyberg, 1972, Willerslev, 2004a), especially in materials like dry cold basal ice. Although the chemical and physical properties of the environment the DNA is preserved in has an influence on the decay process, general damage patterns are hydrolytic or oxidative modifications, which principally cause deamination of cytosine and produce error containing sequences (Pääbo, 1989) Other damage related modifications of the DNA involves fragmentation of the DNA due to spontaneous events of depurination and inter-strand crosslinks, which prevent amplification of the DNA (Mitchell, et al., 2005, Hansen, et al., 2006). Although, DNA preservation in permafrost and glacial ice share the similar features of a cold and dry environment, they can differ substantially in regards to temperature history, chemical and physical properties. DNA in permafrost is exposed to a higher degree of radiation and chemical reactions from sediment particles (Hansen, et al., 2006, Bidle, et al., 2007) than DNA preserved in basal ice where the sediment concentration is more likely of a few percent (Herron & Langway, 1978). Preservation conditions within cold-based ice sheets are considered optimal for long term DNA preservation (Willerslev, et al., 2004) due to the very stable and dry conditions that would slow down the rate of spontaneous chemical decay reactions, according to the Arrhenius equation (see review, chapter 1).

If we compare the expected survival time of DNA at cold temperatures only from an extrapolation from the Arrhenius equation at the average temperature expected from Camp Century (-15C) (Dahl-Jensen, unpublished), we get a survival time of just below one million years. However, this extrapolation is based on DNA in solution and we would expect longer preservation in dry environments where hydrolytic damage is expected lower than when in solution. Moreover, temperature has been shown to play a minor role in the preservation of ground sloth coprolites from a warm but very dry environment of samples dating to 11-31KY (Poinar, et al., 2007).

Above all, we know that a variety of chemical and physical factors, besides temperature have shown to affect the decay rate and the complexity of how these factors interact in a synergistic or antagonistic way is not resolved (Willerslev, 2004a, Mitchell, et al., 2005, Cai, et al., 2006). We therefore do not know the time limit of how long DNA can be preserved.

There are many uncertainties in regards to DNA survival and very different results can be produced within a sample site and even between microclimates within a sample where different molecular preservation niches exist (Ginolhac, et al., 2011, Orlando, et al., 2011). However, so far the oldest DNA findings, that are independently reproduced and verified in terms of damage signals, are done in cold and dry environments where temperatures have remained frozen for hundreds of thousands of years in permafrost and cold based basal ice (Willerslev, 2003, Johnson, et al., 2007, Willerslev, et al., 2007).

The basal ice DNA sequences in Camp Century show an ancient signal in terms of being fragmented and only amplifiable in short amplicons. It is recommendable to authenticate the aDNA content by quantitative PCR to test the concentration of starting material available in the samples and if the DNA accord to a fragmented nature of the DNA that would indicate a ancient signal (Cooper & Poinar, 2000). However, as the DNA in the ice samples contain a variety of environmental DNA in low concentration it was not possible to construct reliable standards to perform such measurements. Another criteria to assess the ancient nature of old samples is to use amino acid racemization to verify that the DNA is preserved and show an old signal. One assumption taken when evaluating the amino acid D/L ratio in environmental samples is that it performs like a closed system. The multiple replicates of several

samples from this study illustrated that such an assumption could not be taken and the amino acid racemization were not applicable for assessing the degradation or preservation of environmental DNA or as an independent dating method (see Supplementary Material). Another study found a similar outcome from preservation conditions in bone that the amino acid racemization signal was not indicative of the aDNA preservation within bone material (Collins, et al., 2009).

Moreover the insect Cox 1 sequences are being implemented as an independent dating method by the use of relative rate test, although results are still pending.

**Dating to find a climatic time slice**

Interglacial periods in the past have been compared to the present interglacial to better understand the climate changes occurring now and in the future. When attempting comparisons to the current interglacial it is crucial for climate models to identify the most similar interglacial from the Pleistocene. Two candidates have been used - the Eemian (130-113KY) and Marine Isotope Stage 11 (MIS11) (420-360KY) - both considered warmer than the previous interglacials within the Pleistocene like MIS13, MIS15 and MIS17 (Lambert, et al., 2008).

The climate record provides information on the timing of interglacial and glacial periods in the past that is necessary when determining a time slice that supported the palaeo-ecosystem we detect. The three dating methods we applied compiled data from atmosphere and precipitation (cosmogenic radionuclide decay, $^{10}Be/^{36}Cl$ - decay ratio), geology and physical processes (OSL) and physical decay processes ($^{238}U/^{234}U$ recoil dating) to give a minimum age range of when the environment was from. Figure 4 shows the consensus minimum age of the ice by the three methods that can also be interpreted as the minimum age of the DNA and the time the environment was last ice-free.

The $^{234}U/^{238}U$ recoil dating method provided the oldest minimum ages between 281-384KY, OSL provided a mean age of 260KY (the oldest sample gave a signal of 700KY) and $^{10}Be/^{36}Cl$ provided an age range between 150 – 400 KY. The consensus between methods is that we identify a period older than the Eemian (130-116KY). MIS11 is the interglacial previous to the Eemian and occurred 360-420KY. This time

frame overlaps with the minimum consensus age we find by the three dating methods. Furthermore, MIS11 is considered to be warmer than the Eemian and is subject to the discussion of whether or not it was warmer than or comparable to the current interglacial we are in (Poore & Dowsett, 2001, Alley, et al., 2010). However, two cases point to an age older than MIS11 for Camp Century. First, boreal forest was not detected at the Dye 3 site in southern Greenland in MIS11 but previous to this period, and second a boreal signal from south western Greenland was detected during MIS11 (Willerslev, et al., 2007, de Vernal & Hillaire-Marcel, 2008), Therefore, we would not expect to find a warmer climate thousands of kilometers more north and it would contradict some models describing the climate fluctuations from this period (Alley, et al., 2010). Moreover, the general consensus so far is that MIS11 was not much warmer than the interglacial we are in now, but with a different orbital configuration and of a long duration of approximately 100KY that is believed to have contributed to extensive retreat of the GIS (Steig & Wolfe, 2008, Alley, et al., 2010).

To test how the ice sheet and temperature profile around Camp Century would have responded if MIS11 was ice-free, a temperature model was run with two scenarios. One; where the temperature was around 5°C degrees warmer than today, and another that follows the scenario from existing temperature models used today (Lambert, et al., 2008, Steffensen, et al., 2008, de Boer, et al., 2010). The temperature extrapolations were run for a timeline between the present and back to 2.5MY. If the area was free of ice during MIS11, the model showed that Camp Century would have been free of ice for several hundred thousand years. The interglacials that followed MIS11 would also have had warm temperatures and the ice sheet would have retreated or experienced melting. Melting and warm temperatures of just below freezing in long and repeated time intervals, does not support DNA preservation well (Willerslev, 2004a). If temperatures have fluctuated as described above we would not expect to find preserved DNA from the MIS11 but from a more recent period. When the temperature model is run by the second approach where Camp Century is not ice-free during MIS11, we find that the area has remained ice covered since that period and stayed cold (see appendix B). The model suggests that the previous period supporting a warm ecosystem in Camp Century would then extend far back in time, but the model is not accurate enough to suggest a time period (Dahl-Jensen, unpublished).

Overall, we find reason to expect a glaciated area around Camp Century during the MIS11 because the northern-temperate signal we detect, would be outcompeted by a more cold-tolerant and recent flora from any following interglacials. It seems, therefore, that we are limited by the dating methods because we were not able to set a maximum age of the ice.

All of the three dating methods are in development and proved more efficient on the more central ice cores of GRIP and Dye 3 on the GIS where the influence of coastal climate and bedrock was less significant. In the Camp Century site we were not able to distinguish the signal of the ice from inputs of the sediment that might skew the age estimates significantly. If we compare the results from the OSL and $^{234}U/^{238}U$ dating methods of Dye3 and Camp Century the measures suggest an older signal for Camp Century compared to Dye 3, which could indicate that the ages of these two sites overlap.

We were therefore left to evaluate our results with regards to the three warm periods providing evidence of boreal forest different places in Greenland within the Pleistocene; (i) The MIS11 interglacial that supported forest in south-west Greenland detected from the marine core south west of Greenland (de Vernal & Hillaire-Marcel, 2008), (ii) the Dye 3 ice free period between 450-800KY (Willerslev, et al., 2007) and (iii) in the early Pleistocene around 900-1MY when the northern glaciations caused the climate to cool and coincided with the change of glacial orbitals from durations of 40KY to 100KY (Bintanja & van de Wal, 2008). An alternative explanation that is not supported by the expected range of DNA survival would be to compare the ecosystem to be of a similar age as the more northern located boreal Kap København (1.8-2.5MY) (Funder, 1989).

(i) If Camp Century was ice free during the MIS 11 and we extrapolate winter and summer temperatures from an Eemian estimate of 5°C warmer climate than today (Andersen, et al., 2004) we get an additional temperature increase of 7°C due to a lower elevation of the surface of the site when the ice sheet is absent. This estimate incorporates the effect of isostatic rebound and adds up to in total 12°C warmer than today, corresponding to a mean annual temperature of -11°C, mean summer temperature of 9°C and mean winter temperatures of -33°C (Dahl-Jensen, personal

communication). The orbital configuration at the time of MIS11 was different from now and would provide an additional energy input during summer at this northern location but would also have a significant lowering of the winter temperatures (Berger, 1978). Therefore, if the environment were ice-free there would still have been permafrost, which would exclude the growth of oak, yew and cypress. The warm summer temperatures could have allowed for tree growth but not supportive of the northern temperate trees we detect. It is therefore more likely that the area was ice-covered during MIS11 (to allow for the survival of aDNA) and a warmer interglacial period supportive of a permafrost-free area with summer temperatures above 10°C must be responsible for the DNA remains found in the ice.

(ii) If a northern temperate ecosystem have occurred at the same time as the boreal climate at Dye 3 we would need to have had mean annual temperatures above 5°C at the 77°N latitude, co-existing with the mean annual temperatures supporting a boreal ecosystem found at the 65°N latitude at Dye 3 more than 1300km south of Camp Centtury. The boreal ecosystem existed roughly at an elevation of 1000m at Dye 3 whereas the northern temperate at Camp Century existed at an elevation of 900m thereby corresponding to a 1°C temperature increase at Camp Century if only elevation is considered. Moreover, the climate at Camp Century might have been more continental than at the Dye 3 site providing warmer summers and colder winters. Additionally, with increasing latitude the full day sunlight period would be longer at 77°N and could have increased the summer temperatures favorable for tree growth. Therefore, we cannot exclude a co-existence with the Dye 3 ecosystem.

(iii) Alternatively, we might be evaluating an interglacial prior to the origin of the stable GIS, which could be the period of around 0.9-1MY when the glacial cycle pacing changed from 40KY to 100KY at the time when the North American and Greenland Ice Sheet survived the insolation maxima and reached continental size (Bintanja & van de Wal, 2008). Moreover, the period around 1.07MY ago equivalent to MIS 31 is modelled to be a period when the West Antarctic Ice Sheet suffered significant reduction or collapse. This period could represent a warm phase prior to the glacial

cycle change producing stable ice sheets in both Greenland and West Antarctica around 1MY (Pollard & DeConto, 2009). An age of around 1MY would also be possible in terms of DNA preservation.

If we compare the findings of biodiversity and age with the Dye 3 findings and with the fossils from Kap København, one could argue that they fit together in terms of describing a past-forested Greenland as described by Funder et al., (1989)(Funder, et al., 2001). However, the ages should then be adjusted making Dye 3 slightly older from the maximum range of 800KY to 900KY-1MY and by reducing the Kap Købehanhavn age by at least 1MY from the current 1.8-2.5MY age (Funder, 1989). Another alternative would be to extend the age of the Camp Century to co-occur with Kap København. We do not find the latter reasonable in terms of DNA preservation. However, the two sites do have some shared taxa although Camp Century DNA represents a warmer climate than the vegetation reconstructed at Kap København (Bennike & Bôcher, 1990). The distance to Kap København around 82°N is roughly corresponding to 800km further north. If the Camp Century site was forested around the same time as the boreal Kap København formation it would make sense to find an ecosystem containing similar to more warm tolerant taxa further south, which is the case with the genera *Chamaecyparis, Quercus, and Taxus* from Camp Century.

## CONCLUSIONS

We find aDNA from an ecosystem that has a biodiversity similar to what is found today in northern temperate and nemoral climate zones, as e.g. in eastern North America with both mountain and lowland taxa within proximity. Cool winters with snow and frost and summers with a growth season of around 4 months where summer temperatures rose above 10°C, for supporting the growth of yew, cypress and oak, chracterize such an ecotone. The Camp Century ecotone is characterized by forest, meadow areas and elevated mountain slopes to support taxa like *Dryas* and *Oxyria*. The mean annual temperature for such a zone varies but does not go below a mean annual average temperature of 5°C.

Attempting to define the interglacial period that supported the environment around Camp Century when it was last warm and ice-free is of interest for setting the climate and ecosystem into perspective and compare it to what has been found by other climate and terrestrial records. Additionally, it is of interest for describing the dynamics and fluctuations of the GIS through time. So far, the age range estimated for the basal ice by the three applied methods provided a consensus minimum estimate around the MIS11 *ca*. 300,000 – 400,000 years ago. This indicates that the ice sheet at Camp Century did not retreat during the Eemian, which is reducing uncertainty for models describing the fluctuations of the GIS at this site during the last interglacial. However, we know that the dating methods are limited by their ability to detect a maximum age and we do not consider MIS11 as a likely candidate to have supported a nemoral-forested northwestern Greenland. We therefore believe that the ecosystem with mixed nemoral forest, meadows and mountain slopes around Camp Century co-existed with the boreal ecosystem at the Dye 3 site that was dated to the 450-800KY-age window. Moreover, we speculate that the two locations could have co-existed with the Kap København formation supporting the picture of a forested Greenland similar to the description extrapolated from the Kap København taxa (Funder, et al., 2001). However, such a hypothesis would change the ages for all sites to be at least the age of when the GIS became stable in the early to mid-Pleistocene, perhaps around 1MY.

**Acknowledgements**

A great thank you to Tina Brand (Centre for GeoGenetics) for sharing long hours in the freezer helping me processing the samples. Thank you to Mike Bunce and Morten Allentoft for processing and replicating the samples in Murdoch University clean lab, Perth, Australia. Within Centre for Ice and Climate I thank Henrik Clausen, Bo Vinter, Anders Svensson for valuable discussions on ice sheet dynamics and basal processes and the local history of the samples. Moreover, I would like to thank Svend Funder for discussions on the palaeo-environmental history of Greenland and for helpful comments. Thank you to Morten Rasmussen, Aurélien Ginolhac and Ludovic Orlando for developing pipelines to sort, filter and assign the data. Thank you Rute Fonseca and Tobias Mourier for helpful scripts and discussions on data. Moreover, I would like to add a special thank you to Per Hartvig for providing valuable considerations on the plant history and ecosystem niches. Thank you Dorthe Dahl-Jensen for providing temperature models on the site. A huge effort was invested in dating the material so a special thank you to the team of people involved in dating and last thank you to my supervisors for valuable discussions on ideas and brainstorms and for making this whole study possible.

**Figure 1**. Map showing the location of the Camp Century ice core and other deep drill sites on the Greenland Ice Sheet.

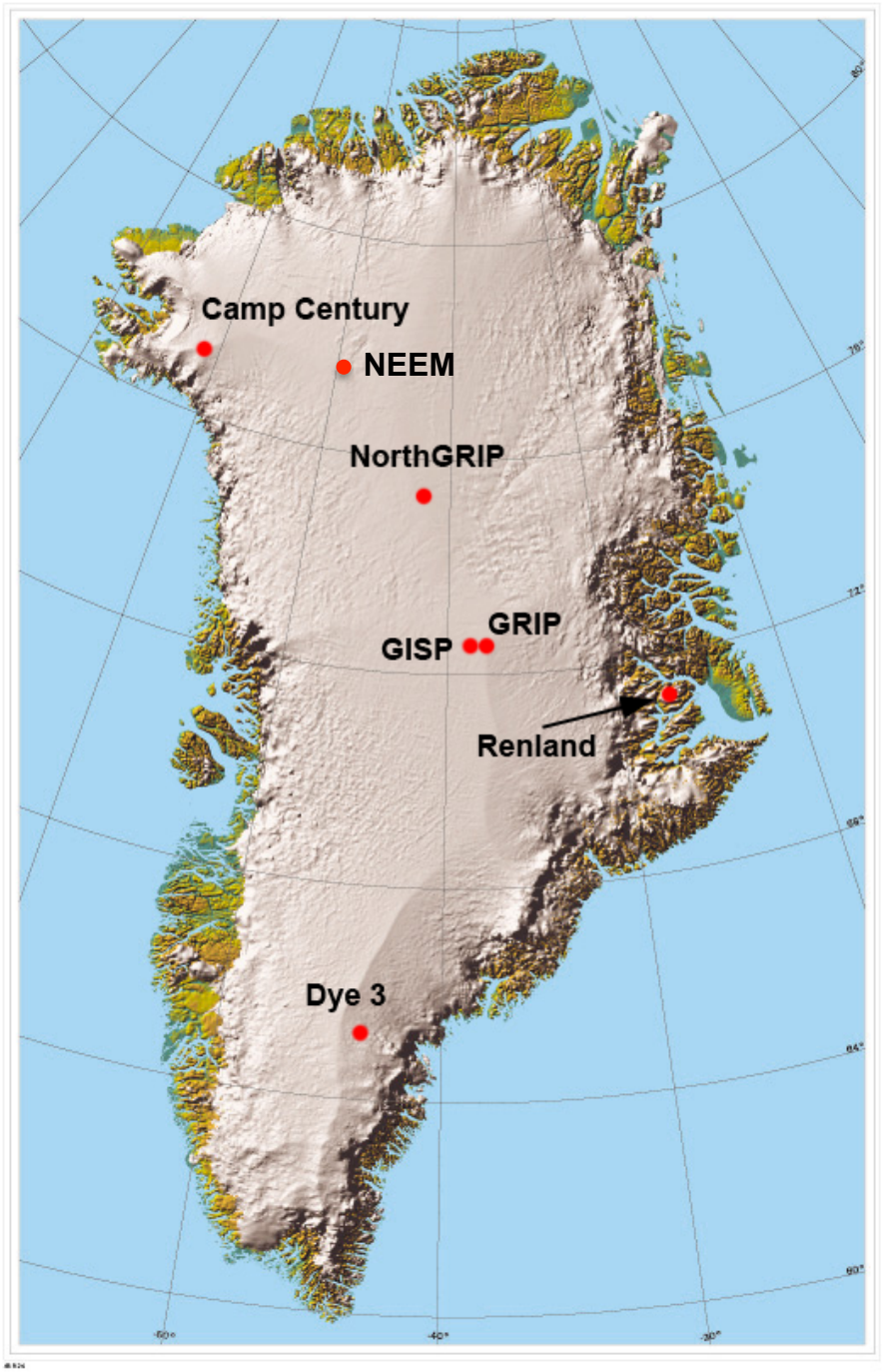


**Figure 2**. World Map of Köppen-Geiger Climate Classification from Rubel & Kottek (2010). The zone Cfb represents the nemoral, northern temperate zone with moist climate, no dry season and warmest temperatures below 22°C.

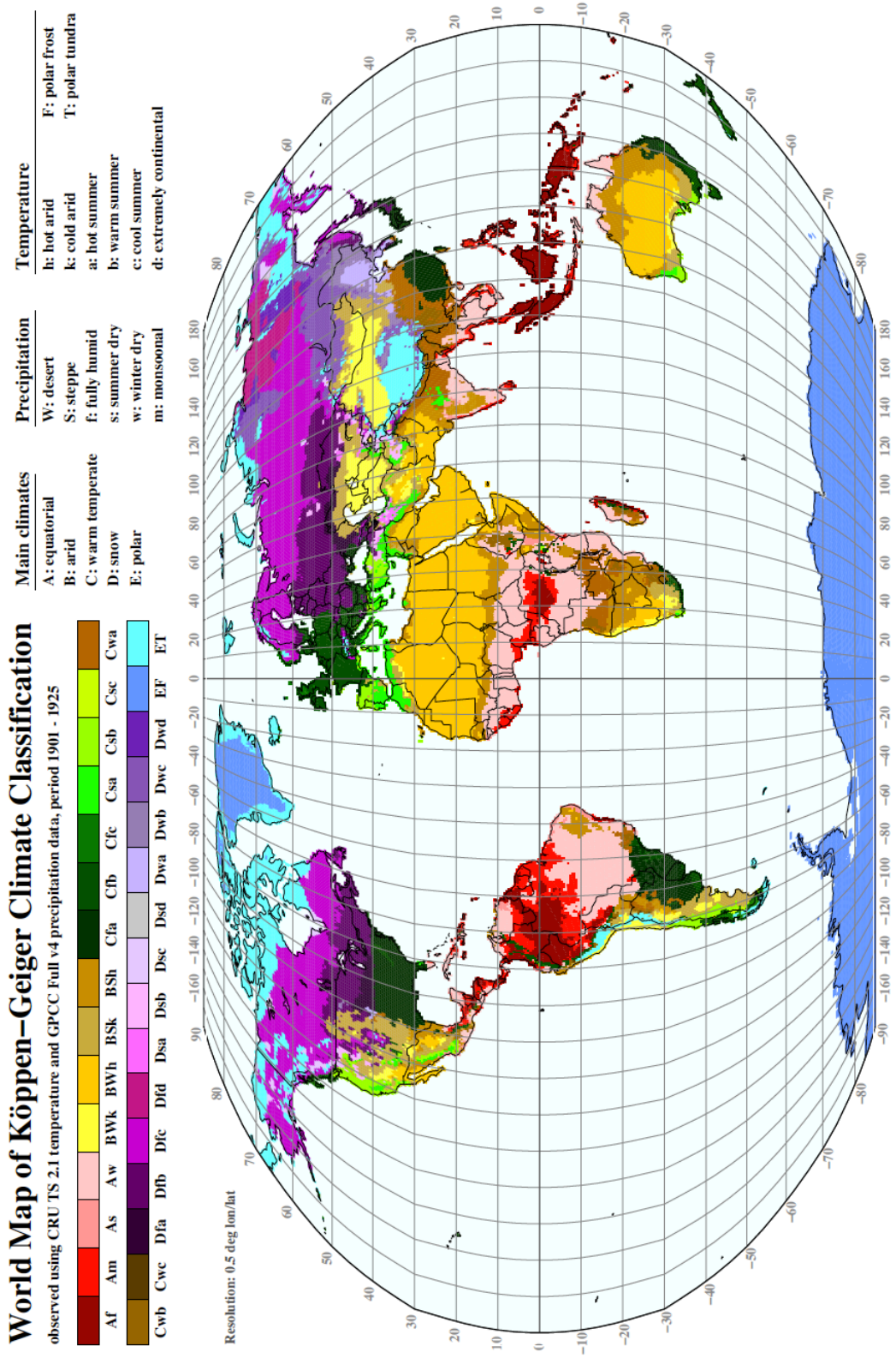


**Figure 3.** Biome diagram modified from (Whittaker, 1975) illustrating the temperature and precipitation constraints on the different biome zones (Ricklefs, 2008).

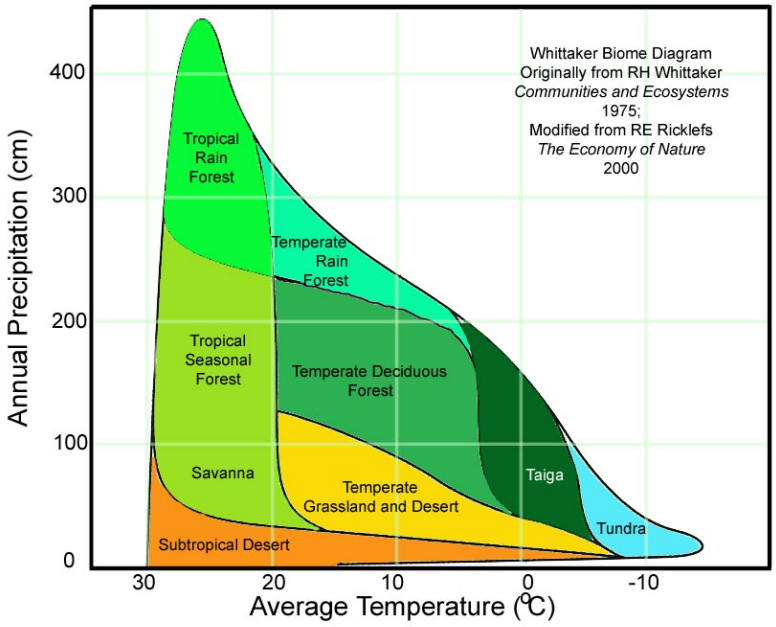


**Figure 4.** Illustrating the minimum age estimates provided by three dating methods; Uranium ($^{234}U/^{238}U$), OSL and Be/Cl ($^{10}BE/^{36}Cl$). The thick bars are the minimum age intervals for each method and the thin bars illustrate the uncertainty of the methods where no maximum ages can be applied and we therefore extended them to the late-Pliocene/Pleistocene transition.

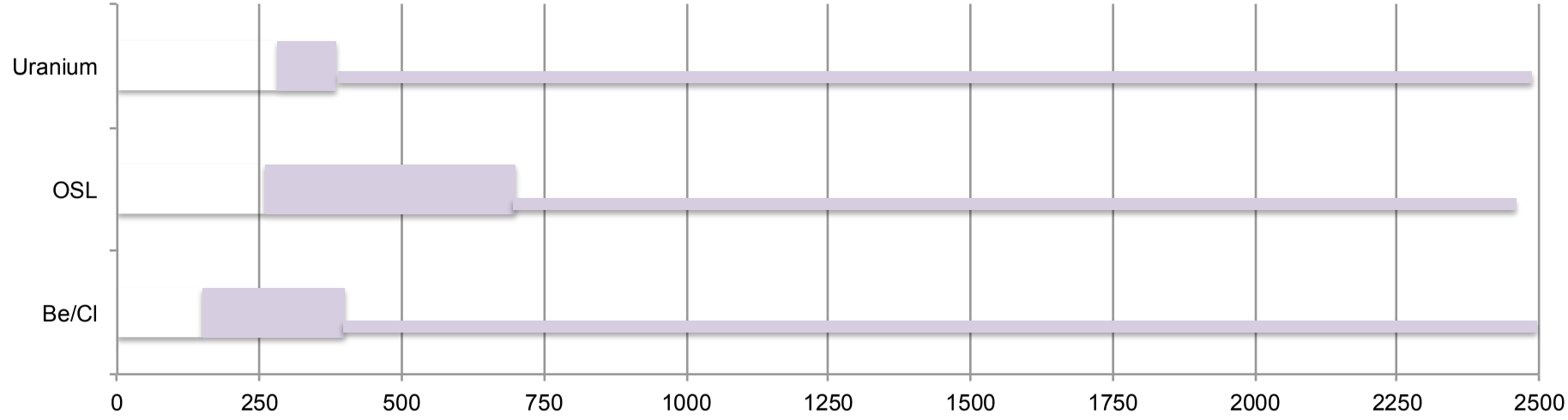
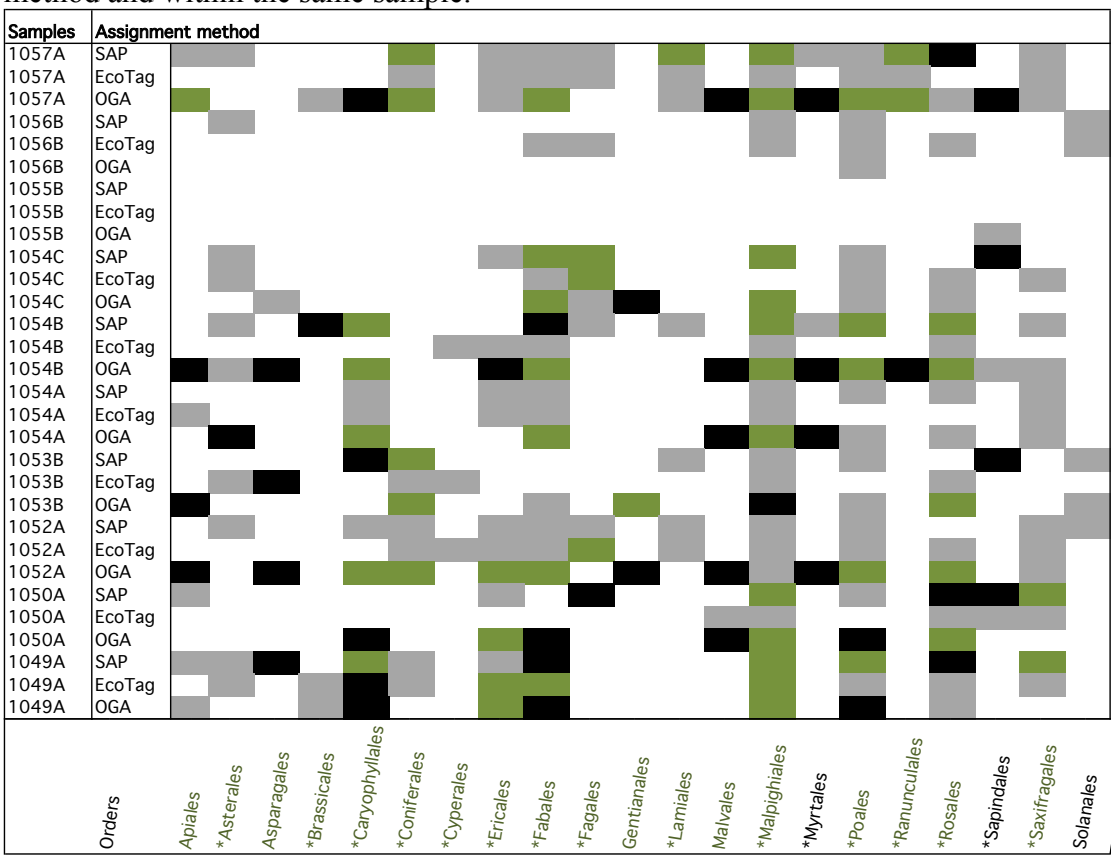
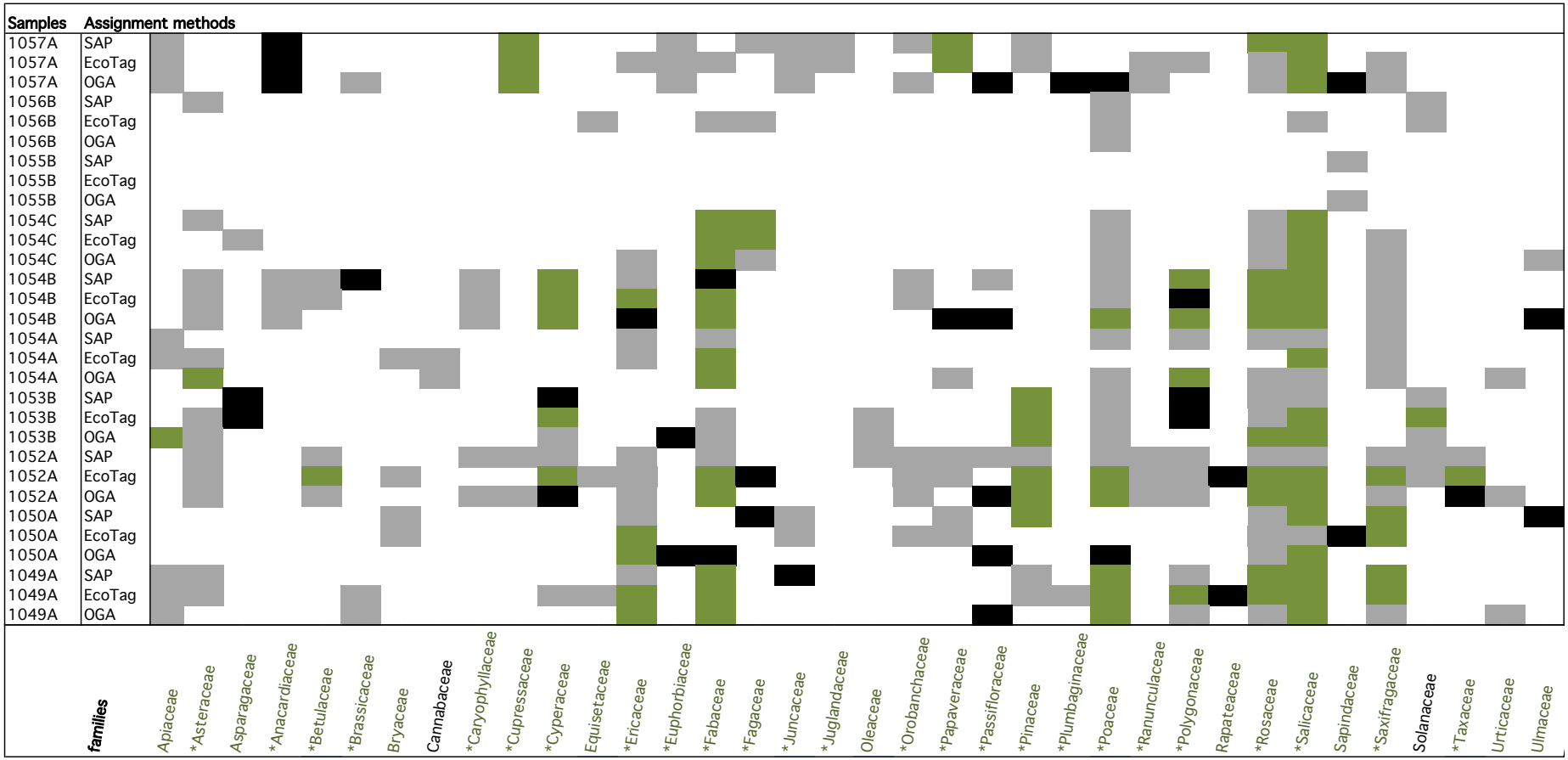


**Table 1**. Presence-absence data at the order level of assigned taxa as determined by each of the three assignment methods. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, black by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.



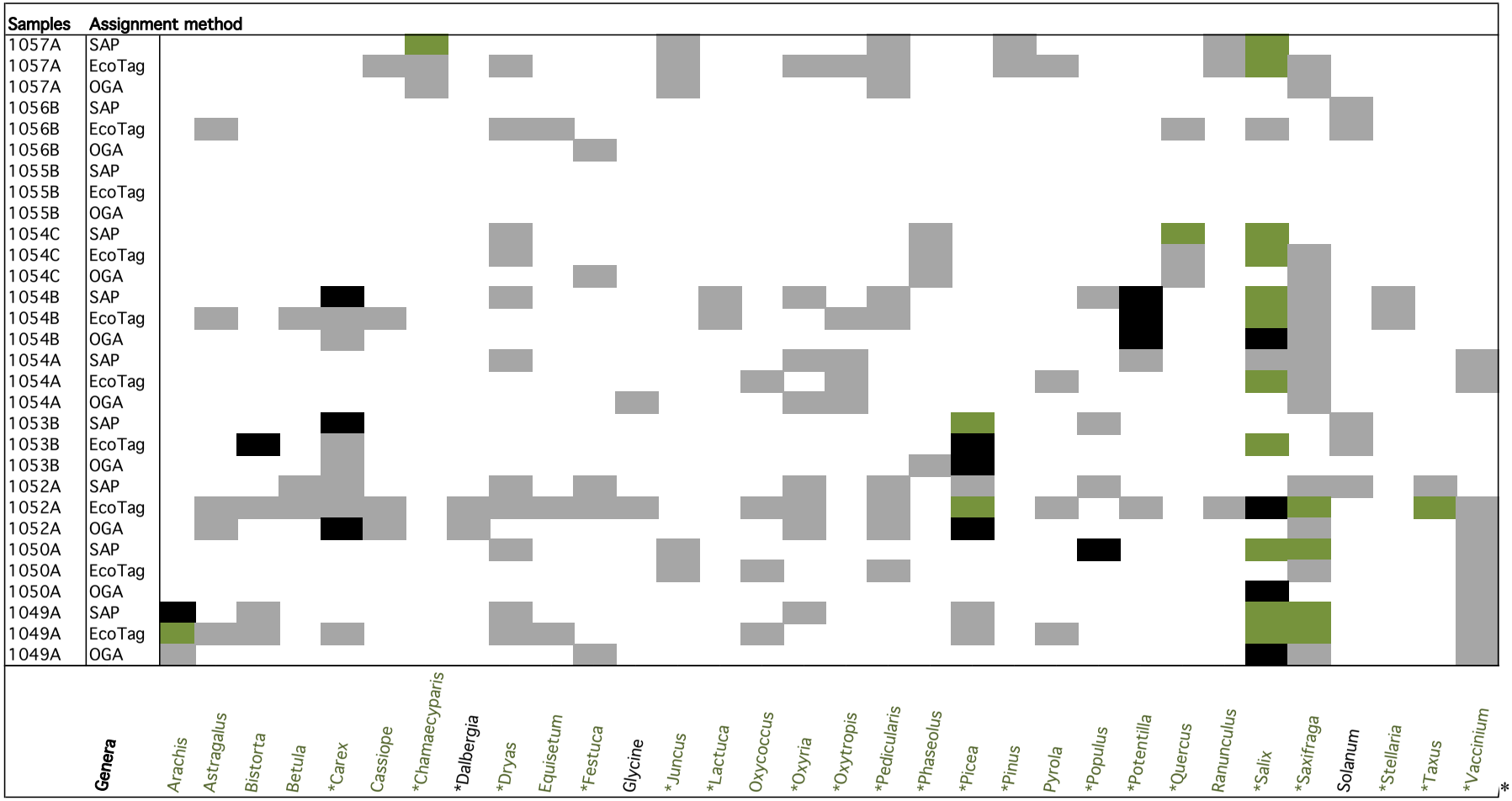

* mark where taxa were replicated on the subset of samples send to the Murdoch University clean lab, Perth, Australia.
Green writing indicates the taxa that were found by more than one method in one or several samples.
Black writing indicates the taxa that are possible contaminants from exotic food plants or misassignments.

**Table 2**. Presence-absence data at the family level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL, black by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.




* mark where taxa were replicated on the subset of samples send to the Murdoch University clean lab, Perth, Australia.
Green writing indicates the taxa that were found by more than one method in one or several samples.
Black writing indicates the taxa that are possible contaminants from exotic food plants or misassignments.

**Table 3**. Presence-absence data at the genus level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL, black by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.




* mark where taxa were replicated on the subset of samples send to the Murdoch University clean lab, Perth, Australia.
Green writing indicates the taxa that were found by more than one method in one or several samples.
Black writing indicates the taxa that are possible contaminants from exotic food plants or misassignments.

**Table 4**. Listing the diversity at family and genera level where blue colors shows overlap with the pollen results from the glacial ice, and pink where there is overlap with pollen from the basal ice.

| Families | pollen | Genera | pollen |
|---|---|---|---|
| Apiaceae | | Arachis | |
| *Asteraceae | | Astragalus | |
| Asparagaceae | | Bistorta | |
| *Anacardiaceae | | Betula | |
| *Betulaceae | | *Carex | |
| *Brassicaceae | | Cassiope | |
| Bryaceae | | *Chamaecyparis | |
| Cannabaceae | | *Dalbergia | |
| *Caryophyllaceae | | *Dryas | |
| *Casuarinaceae | | Equisetum | |
| *Cupressaceae | | *Festuca | |
| *Cyperaceae | | Glycine | |
| Equisetaceae | | *Juncus | |
| *Ericaceae | | *Lactuca | |
| *Euphorbiaceae | | Oxycoccus | |
| *Fabaceae | | *Oxyria | |
| *Fagaceae | | *Oxytropis | |
| *Juncaceae | | *Pedicularis | |
| *Juglandaceae | | *Phaseolus | |
| Oleaceae | | *Picea | |
| *Orobanchaceae | | *Pinus | |
| *Papaveraceae | | Pyrola | |
| *Passifloraceae | | *Populus | |
| *Pinaceae | | *Potentilla | |
| *Plumbaginaceae | | *Quercus | |
| *Poaceae | | Ranunculus | |
| *Ranunculaceae | | *Salix | |
| *Polygonaceae | | *Saxifraga | |
| Rapateaceae | | Solanum | |
| *Rosaceae | | *Stellaria | |
| *Salicaceae | | *Taxus | |
| Sapindaceae | | *Vaccinium | |
| *Saxifragaceae | | | |
| Solanaceae | | | |
| *Taxaceae | | | |
| Urticaceae | | | |
| Ulmaceae | | | |

References

Aciego S, Bourdon B, Schwander J, Baur H & Forieri A (2011) Toward a radiometric ice clock: uranium ages of the Dome C ice core. *Quaternary Science Reviews* **30**: 2389-2397.

Aciego SM, Bourdon B, Lupker M & Rickli J (2009) A new procedure for separating and measuring radiogenic isotopes (U, Th, Pa, Ra, Sr, Nd, Hf) in ice cores. *Chemical Geology* **266**: 194-204.

Alley RB, Andrews JT, Brigham-Grette J, et al. (2010) History of the Greenland Ice Sheet: paleoclimatic insights. *Quaternary Science Reviews* **29**: 1728-1756.

Andersen KK, Azuma N, Barnola JM, et al. (2004) High-resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**: 147-151.

Barbour MG & Billings WD (2000) *North American terrestrial vegetation*. Cambridge University Press, Cambridge, U.K. ; New York, NY.

Bennike O & Bôcher J (1990) Forest-Tundra Neighboring the North-Pole - Plant and Insect Remains from the Plio-Pleistocene Kap København Formation, North Greenland *Arctic* **43**: 331-338.

Berger AL (1978) Long-term variations of daily insolation and Quaternary climatic changes. *Journal of Atmospheric Science* **95**: 5.

Bidle KD, Lee S, Marchant DR & Falkowski PG (2007) Fossil genes and microbes in the oldest ice on Earth. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 13455-13460.

Bintanja R & van de Wal RS (2008) North American ice-sheet dynamics and the onset of 100,000-year glacial cycles. *Nature* **454**: 869-872.

Cai P, Huang QY & Zhang XW (2006) Interactions of DNA with clay minerals and soil colloidal particles and protection against degradation by DNase. *Environmental Science & Technology* **40**: 2971-2976.

Carrara PE, Ager TA & Baichtal JF (2007) Possible refugia in the Alexander Archipelago of southeastern Alaska during the late Wisconsin glaciation. *Canadian Journal of Earth Sciences* **44**: 229-244.

Collins MJ, Penkman KEH, Rohland N, Shapiro B, Dobberstein RC, Ritz-Timme S & Hofreiter M (2009) Is amino acid racemization a useful tool for screening for ancient DNA in bone? *Proceedings of the Royal Society B: Biological Sciences* **276**: 2971-2977.

Cooper A & Poinar HN (2000) Ancient DNA: Do It Right or Not at All. *Science* **289**: 1139.

de Boer B, van de Wal RSW, Bintanja R, Lourens LJ & Tuenter E (2010) Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic delta(18)O records. *Annals of glaciology* **51**: 23-33.

de Vernal A & Hillaire-Marcel C (2008) Natural Variability of Greenland Climate, Vegetation, and Ice Volume During the Past Million Years. *Science* **320**: 1622-1625.

Fredskild B & Wagner P (1974) Pollen and fragments of plant tissue in core samples from the Greenland Ice Cap. *Boreas* **3**: 105-108.

Funder S (1989) Quaternary Geology of the Ice-Free Areas and Adjacent Shelves of Greenland. *Quaternary Geology of Canada and Greenland* **K-1**: 741-792.

Funder S, Abrahamsen N, Bennike O & Feylinghanssen RW (1985) Forested Arctic - Evidence from North Greenland. *Geology* **13**: 542-546.

Funder S, Bennike O, Bocher J, Israelson C, Petersen KS & Simonarson LA (2001) Late Pliocene Greenland - The Kap Kobenhavn Formation in North Greenland. *Bulletin of the Geological Society of Denmark* **48**: 117-134.

Ginolhac A, Vilstrup J, Stenderup JT, et al. (2011) Improving the performance of true-Single Molecule Sequencing for ancient DNA. *BMC Bioinformatics* **Submitted**.

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.

Herron S & Langway CCJ (1978) The debris-laden ice at the bottom of the Greenland Ice Sheet. *Symposium on Glacier Beds: The Ice-Rock Interface, Proceedings of a Symposium held at Carleton University, Ottawa(Ontario) August 15-19, 1978. Journal of Glaciology,* **23**.

Johnson SS, Hebsgaard MB, Christensen TR, et al. (2007) Ancient bacteria show evidence of DNA repair. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 14401-14405.

Jørgensen T, Kjaer KH, Haile J, et al. (2011) Islands in the ice: detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Molecular Ecology*.

Knight PG (1997) The basal ice layer of glaciers and ice sheets. *Quaternary Science Reviews* **16**: 975-993.

Lambert F, Delmonte B, Petit JR, et al. (2008) Dust-climate couplings over the past 800,000 years from the EPICA Dome C ice core. *Nature* **452**: 616-619.

Leonard J, Shanks O, Hofreiter M, et al. (2007) Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science* **In Press, Corrected Proof**.

Lindahl T & Nyberg B (1972) Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **11**: 3610-3618.

Lydolph MC, Jacobsen J, Arctander P, et al. (2005) Beringian paleoecology inferred from permafrost-preserved fungal DNA. *Applied and Environmental Microbiology* **71**: 1012-1017.

Mitchell D, Willerslev E & Hansen A (2005) Damage and repair of ancient DNA. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **571**: 265-276.

Munch K, Boomsma W, Huelsenbeck J, Willerslev E & Nielsen R (2008a) Statistical Assignment of DNA Sequences Using Bayesian Phylogenetics. *Systematic Biology* **57**: 750-757.

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

Pääbo S (1989) Ancient DNA - Extraction, Characterization, Molecular-Cloning, and Enzymatic Amplification. *Proceedings of the National Academy of Sciences of the United States of America* **86**: 1939-1943.

Poinar H, Kuch M, Grocke DR, et al. (2007) A preliminary analysis of the DNA and diet of the extinct Beothuk: A systematic approach to ancient human DNA. *American Journal of Physical Anthropology* **132**: 594-604.

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

Pollard D & DeConto RM (2009) Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature* **458**: 329-U389.

Poore RZ & Dowsett HJ (2001) Pleistocene reduction of polar ice caps: Evidence from Cariaco Basin marine sediments. *Geology* **29**: 71-74.

Ricklefs RE (2008) *The economy of nature*. W.H. Freeman, New York.

Rubel F & Kottek M (2010) Observed and projected climate shifts 1901-2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorologische Zeitschrift* **19**: 135-141.

Shackleton NJ, Backman J, Zimmerman H, et al. (1984) Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* **307**: 620-623.

Souchez R (2000) Basal ice formation and deformation in central Greenland: a review of existing and new ice core data. *Special publication - Geological Society of London* **176**: 13.

Souchez R, Tison JL, Lorrain R, et al. (1994) Stable Isotopes in the Basal Silty Ice Preserved in the Greenland Ice-Sheet at Summit - Environmental Implications. *Geophysical research letters* **21**: 693-696.

Steffensen JP, Andersen KK, Bigler M, et al. (2008) High-resolution Greenland Ice Core data show abrupt climate change happens in few years. *Science* **321**: 680-684.

Steig EJ & Wolfe AP (2008) Atmospheric science - Sprucing up Greenland. *Science* **320**: 1595-1596.

Walter H & Breckle S-W (2002) *Walter's Vegetation of the earth : the ecological systems of the geo-biosphere*. Springer, Berlin ; New York.

Whittaker RH (1975) *Communities and ecosystems*. Macmillan, New York.

Willerslev E (2003) Diverse Plant and Animal Genetic Records from Holocene and Pleistocene Sediments. *Science* **300**: 791-795.

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

Willerslev E, Cappellini E, Boomsma W, et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**: 111-114.

# Chapter VI

TITLE:

Revelation of a vegetated palaeo-ecosystem from the basal ice in the Dry Valleys of Antarctica: An ancient metagenomics approach

ABSTRACT

Antarctica provides the most optimal conditions for DNA preservation, being one of the coldest and most stable environments on earth. Here, we present findings of the oldest, preserved ancient DNA reported thus far in a dry environment with mean annual temperatures of –20°C, which pushes back the limits of survival time of DNA. The ancient DNA results support previous fossil findings of a warm mid-late Pliocene climate in Antarctica. A forested boreal ecosystem in the area around Taylor Valley, Victoria Land was detected by ancient DNA sequence analyses and points to a co-existence with a tundra environment in the Transantarctic Mountains of Antarctica. The composition of taxa found describes a boreal biome dominated by conifer trees with grasses, herbs and shrubs that today exist in continental boreal climate zones where summer temperatures allows for tree growth with mean annual temperatures around -10°C.

Temperatures that constrain the existence of a boreal ecosystem today fall within the model predictions of this time period approximately 3 million years ago. The results have implications for predicting both ancient DNA survival and for models describing the responses to a warming climate in a period with global mean annual surface temperatures that were 2-3°C warmer than today with $CO_2$ concentrations similar or slightly higher. Moreover, the results are in contrast to model predictions of a stable East Antarctic Ice Sheet during the last 14 million years and instead support a dynamic behavior of the East Antarctic Ice Sheet in the Pliocene period with an absence of the West Antarctic Ice Sheet. This study adds to the understanding of the ecological history and diversity in Antarctica by new tools that can be applied on a broader geographical scale to further explore the polar areas.

BACKGROUND

**History**

Antarctica is divided by the Transantarctic Mountains that forms a transcontinental arc of more than 4000 km and are varying in width from 100-300 km. The mountain range defines the boundary between East Antarctica and West Antarctica in both a geological and morphological setting (Denton, et al., 1991).

The Antarctic Ice Sheet covers around 98% of the continent and only reveals evidence of its rock base through visible mountain ranges, nunataks or coastal outcrops (Tingey, 1991). Below the ice sheet, Antarctica is known to show a high degree of variability in physical, chemical and biological characteristics (Samyn, 2005a). The West Antarctic Ice Sheet (WAIS) only comprises around 9.1% of the entire ice sheet in Antarctica whereas the East Antarctic Ice Sheet (EAIS) comprises 88.7% of the Ice Sheet and is based on a bedrock predominantly above sea-level (Denton, et al., 1991). In Antarctica the largest ice free area is found in the Ross sea region that occupies around half of the ice free land in all of Antarctica (Samyn, 2005a). The Transantarctic Mountains provide a natural barrier against ice expansion from the East Antarctic Ice Sheet towards the Ross Sea. However, there are three glaciers that cross the mountain range, making pathways for discharge of continental ice from East Antarctic Ice Sheet (EAIS) into the Dry Valleys in the Ross Sea region (Fitzsimons, 1996, Souchez, et al., 2004). These three valleys are Taylor Valley, Victoria Valley and Wright Valley (see Figure 1).

The growth of the Antarctic Ice Sheet is suggested to have started by an abrupt cooling in the late Eocene (ca 34 MY) (Zachos, et al., 2001) at which time there was a sudden decrease in atmospheric $CO_2$ concentrations (Pearson, et al., 2009). The ice sheets have fluctuated in size and advanced beyond their current levels since the Miocene (23-5.3MY), with an increased frequency of ice sheet dynamics since the Pliocene-Pleistocene (Anderson, et al., 2011). In the warm period during the mid-Pliocene (ca 3MY), the stability of the EAIS has been significantly debated (Webb & Harwood, 1987, Marchant, et al., 1993, Sugden, 1996, Francis, et al., 2007a). The WAIS is thought to have experienced great fluctuations in both thickness and extent throughout the Pliocene and Pleistocene (Convey, et al., 2009, Pollard & DeConto, 2009). "Stabilists" believe that the EAIS has been stable since the last 14 MY and

that the warm period during the mid-Pliocene had little impact on the extent of the EAIS (Marchant, et al., 1993). On the other hand, the "dynamists" suggest that the EAIS has responded in a dynamic way to the warming in the late-Pliocene where temperatures were estimated to be 2 – 3°C warmer than today (Salzmann, et al., 2011).

**Palaeo-Ecosystems and Climate**

The cooling trend that had started during the early Cretaceous was interrupted by warm periods that supported the existence of conifer and deciduous forests in Antarctica since the last 100 million years (Poole & Cantrill, 2006). The mean average temperatures in Antarctica were 7-15°C around the early Eocene (59-50 MY), when such high latitude polar forests thrived (Francis & Poole, 2002) before the drastic cooling in the Late Eocene that caused cold seasonal climates and the onset of the Cenozoic ice sheet build up (Zachos, et al., 2001). These polar forests have no present day analogues as they were adapted to summers with continuous sunlight and dark winters (Francis & Poole, 2002). There has been no evidence suggesting that the diverse forest ecosystems were replaced by more cold tolerant vegetation gradually. Instead, the youngest fossil findings from the Transantarctic Mountains from the late Pliocene period (~3 MY) show that a tundra vegetation was able to re-colonize areas after the ice sheets and glaciers retreated with mean annual temperatures of -12°C (Webb & Harwood, 1987), although these findings have been disputed for their age (Ashworth & Cantrill, 2004). Another study supports the hypothesis of a warm Pliocene climate around 3 MY by micro fossil findings in a sediment core north-northeast of Cape Adare off the coast Victoria Land (see Figure 1) (Fleming & Barron, 1996). If the late Pliocene date is correct, this period is the latest known to support vascular plants in Antarctica.

The fossil findings in the Sirius Group strata in the Transantarctic Mountains (TAM) (Webb & Harwood, 1987, M. Haywood, et al., 2002), indicate a tundra environment in this area during the warm Pliocene period. The age of the Sirius Group strata is disputed, which is one of the reasons why the discussion about the stability of the EAIS is still not resolved (Sugden, 1996, Francis, et al., 2007a).

The Sirius group strata fossil findings from TAM and the palynomorphs found in sediment cores from the Ross Sea, both show evidence of a terrestrial environment

with a diverse flora in the mid-Pliocene period (Lewis & Marchant, 2008, Anderson, et al., 2011). The Sirius Group vegetation from a high elevated area was composed of *Nothofagus* shrubs, moss, liverworts, seeds of several vascular plants, insects and beetles and is suggestive of a tundra environment with short summers with temperatures up to 16°C and long winters with freezing temperatures (Francis, et al., 2007b). These biological findings were used to constrain the mean annual temperatures to a minimum of -5°C, below which *Nothofagus* cannot exist (Marchant, et al., 1993).

Coupled atmosphere-vegetation models support that temperatures during the late-Pliocene would support a tundra-like ecosystem in East Antarctica around the Sirius group area in the Transantarctic Mountains with short summer seasons with temperatures up to 5°C and mean annual temperatures of -12°C (Francis, et al., 2007a). Other biological models support a hypothesis of a dynamic ice sheet providing ice-free refugias such as nunataks or low altitude areas as the Dry Valleys and coastal habitats where relict biota could survive during successive glacial cycles from the Neogene and Quaternary until the late-Pleistocene (Convey, et al., 2009). However, very limited glaciological and geological evidence exist to support such a hypothesis and there is a need to identify such potential refugia by new methods.

**This Study Site**

The Dry Valleys of Antarctica are known for their stable cold and arid climate, with very limited precipitation, a high degree of sublimation and continuous winds that remove any falling snow (Samyn, 2005a). Taylor Valley has several glaciers of different types; outlet glaciers with a continental flow from the EAIS towards the Ross Ice Shelf (Taylor Glacier), smaller alpine glaciers that descends from the valley slopes (e.g. Suess, Meserve and Rhone Glaciers) and piedmont glaciers that are found in low coastal promontories (Fitzsimons, 1996). Such a polar environment with dry and stable cold temperatures (Sugden, 1996) offers optimal conditions for DNA preservation, as both the dry and extremely cold temperatures slow down the kinetics of DNA hydrolytic reactions and thereby degradation (Willerslev, et al., 2004). Winter temperatures in the Dry Valleys go down to -45°C and summer temperatures can occasionally reach a few degrees above freezing but the mean annual temperature borders -20°C (Samyn, 2005a).

Taylor glacier (terminal zone) and Suess glacier have been sampled for this study and are both considered cold and dry without basal melting for the sampled locations (Samyn, et al., 2010). The source for these two areas differ in that Taylor Glacier is an outlet glacier with inflow from the EAIS whereas Suess is a much smaller alpine glacier; see Figure 2 for sample locations.

## METHODS

### Dating methods:

Basal ice is generally challenging to date due to basal processes mixing up the layers, and no well-established dating methods exist. Therefore, we applied two different approaches to find a minimum age range that represents the time that precipitation and dust/sediment particles got incorporated into the glaciers. Furthermore, these methods were chosen as they had all been tested on ice material previous to this study on Dye 3 and Dome C ice cores (Willerslev, et al., 2007, Aciego, et al., 2009). Age estimates of the ice and the entrainment of silt material is considered to be a proxy of the minimum ages of the DNA as the entrainment would have occurred at some time later than when the DNA from degraded dead organisms got adsorbed by the soil. Methods providing minimum ages of the ice were uranium ($^{234}U/^{238}U$) recoil dating and cosmogenic $^{10}Be/^{36}Cl$ dating. See Supplementary Material for detailed descriptions of the methods.

### Sampling and Sample Handling and DNA extractions:

Samples from Antarctica originated from two different sites in the Dry Valleys of Antarctica: Suess Glacier and Taylor Glacier (see Figure 2 and Chapter 3 for sample details). The two glaciers have very different flow characteristics but are both cold-based glaciers with basal ice temperatures of -17°C.

We sub-sampled three times from the Suess Glacier basal ice sample (Suess 98/3), with fine silt particle content (0.19%) providing the reddish coloured “amber ice” and three times from the Taylor Glacier basal ice (TG-6B) with the high debris content (see Chapter 3). Sample Su98/1 was sampled above the amber ice and contained a very limited content of silt particles (<0.07%), almost similar to the samples termed

clean ice. Basal ice sample from the Taylor Glacier was similarly sub-sampled three times from the sample TG-6-B that contained a heavy load of debris (8-33%) and sample TGTA that contained limited debris content and was only sampled once (0.19%)

Clean glacial ice was sampled to test for non-local sources of DNA and was taken at the top of the ice blocks Suess 98#1 and TGTA, from Suess and Taylor respectively, representing younger ice than the basal ice below it. See Chapter 3 for sample details, decontamination and extraction protocol.

**DNA Amplification and Sequencing**:
Plant (trnL/rbcL, cpDNA), insect (Cox1, mtDNA) and mammalian (16S, mtDNA) short DNA fragments were PCR-amplified and sequenced from the basal ice DNA extractions as described in Chapter 3.

Clean glacial ice was PCR-amplified with the universal plant trnL and rbcL primers to investigate if the plant DNA could be of windborne origin and hence not represent an indigenous environment from the period prior to the ice-coverage. See Chapter 3 for amplification protocols.

RESULTS AND DISCUSSION

The first aim of testing for amplifiable aDNA was positive and we therefore tested as in the Camp Century chapter for the three possible contamination based scenarios to explain the aDNA findings;
If the results were not representing a local environment the contamination could be [a)] sample based, [b)] laboratory based, or [c)] dispersal based from windborne material. Three approaches were followed to test for the three contamination sources.

[(a)] Sample based contamination could arise by penetration of modern DNA through cracks in the core to the inner parts used for DNA extraction. This was tested for by applying known bacterial vector DNA to the surface of all ice samples and after incubation removing *ca* 1.5 cm of the surface layers. The vector was then amplified in both the surface layers and extractions from the inner core. Samples that proved positive for vector DNA in surface layers and negative in inner core were accepted as free of sample based contamination by modern DNA and used for amplifications by flora and fauna primers.

[(b)] The laboratory-based contamination was tested for by having parallel amplifications performed of a sub-set of samples in Murdoch University clean lab facilities, Perth. These were able to reproduce the findings and moreover found the same result in the Taylor sample that no amplification product came out and that Suess provided amplifications. This also suggests that the Suess results could not have been false positives due to cross contamination from the Camp Century results, as we would also expect to get a signal in the Taylor samples. Moreover, another reason against cross contamination is that all assignments on the order and family levels were detected on sequences that were not identical to sequences from the Camp Century samples and on the genus level there was only one assignment that overlapped with Camp Century on the sequence level (*Festuca*). Last, the decontamination and DNA extractions on the Camp Century and Antarctic samples were not performed the same days.

[(c)] Dispersal based contamination from pollen source or other windblown sources was tested for by first extracting DNA from large amounts of clean glacial ice (see chapter 3). Then we attempted PCR amplification with the plant markers to test for results of the windborne pollen present in the clean glacier ice. No positive

amplification results were found following the same procedures as for the basal ice samples.

**Amplification Success**

The amplification was only successful for one of the Antarctic samples and no results were obtained from the Taylor Glacier. This was at first a surprising result given the high debris content in the TG-6-B samples. However, we believe that this is due to coarse textured and low organic content type of the debris (Barrett, et al., 2006) and the near absence of fine particle material that would be more typical of an ancient terrestrial soil type with a higher organic content. The preservation conditions in the cold basal ice are otherwise optimal for ancient DNA but if the debris content has not sustained a vegetated environment prior to entrainment in the glacier, it would not yield terrestrial DNA as in this study. No results were obtained from the low debris content TGTA sample and the glacial clean ice TGTA sample, both from Taylor Glacier as well.

From the Suess Glacier, we obtained plant and insect sequences for the Suess98/3 sample in all of the three sub-samples. The fine-grained silt material in the amber ice contained silt with a high organic content, which was more similar to the Camp Century basal ice (see Chapter 2), but was overall finer grained. No amplifications were successful from the clean glacial ice from Suess Glacier.

**Assignment Analyses:**

Three pipelines described in detail in Chapter 3 and 4 were used to filter and assign the plant sequences amplified for trnL and rbcL markers.

The Cox1 insect and 16S mammalian sequences were filtered and assigned according to the same criteria as described for the plant markers in Chapter xx, although only by two programs, EcoTag, available at www.prabi.grenoble.fr/trac/OBITools, and, SAP (Munch, et al., 2008a).

The 12S bird cloned sequences were analyzed by the SAP assignment program following the full Bayesian approach (Munch, et al., 2008a)

**Assigned Presence-Absence Data**:

We generated taxonomic diversity tables for each of the taxonomic levels, from order, family to genus, for the two plant markers, rbcL and trnL (Tables 1-3). Tables were designed to show the diversity found per sample and method. The taxonomic results are presented as presence-absence data and colours indicate which of the markers detected the different taxa. For the plant markers, green colour indicates where there is an overlap between the trnL and rbcL markers and grey and blue colours where trnL and rbcL find different taxa, respectively. The light colors indicate if the taxa has been replicated once and found by each of the three assignment methods, and dark color if the taxa has been replicated several times.

Samples Suess98/3 and TG-6-B were send for replication to Murdoch University Clean laboratory, Perth, Australia to test the amplification success for both samples and for overlap between replications in the two laboratories. The taxa amplified in both the Copenhagen and Murdoch labs are marked with asterisks (*).

Amplification of mammalian 16S was not successful but Cox1 provided sequence results for Suess98/3, which were assigned by the EcoTag and SAP assignment programs and results are discussed below.

**Plant diversity in Presence-Absence Data**

Assessment of the presence-absence data for the Suess#98/3 sample from the two plant markers illustrated a difference between the trnL and rbcL markers in the biodiversity they picked up (Appendix D). This trend is similar to the results from the previous study of the Camp Century site in that rbcL assigns sequences to more sub-tropical to tropical taxa compared to trnL, which only picks up a few of these tropical indicator taxa (see Chapter 5). Although the past climate in Antarctica at the time of DNA degradation is largely unknown, we follow the same procedure as for the Camp Century study and accept results when both the trnL and rbcL markers find the same taxa, and where taxa are assigned by more than one method in at least two replicates. The sample material was very limited for the Suess98/3 and provided two different extraction replicates in the Copenhagen lab and one extraction in the Murdoch lab. Therefore, the presence-absence data is presented per sub-sample in order to evaluate how results had been replicated. Results were accepted when they had been replicated

within or between labs and those found between both labs are marked with an asterisk. This reduced diversity is presented in Tables 1-3 and was used for the ecological and climatic interpretations described below.

The diversity listed in Tables 1-3 was used to assess the biodiversity in terms of their distribution in present day climate zones. We find that the assigned taxa groups in general have a widespread distribution. No distinct indicator taxa strictly confined to arctic and boreal boreal climate zones were found but taxa that are more widespread from arctic to temperate were detected in greater majority compared to taxa found in the temperate to tropical climate zones. Figure 3 provides an overview of the distribution of taxa found in the different climate zones. In addition to climate specific taxa, we also found some globally distributed taxa like *Brassicaceae*, *Poaceae*, *Salicaceae*, *Salix*, Rosales and Caryophyllales. Of these, *Caryophyllaceae*, *Poaceae* and Rosales are represented in Antarctica today. *Brassicaceae* is found in the southern hemisphere in the tropical zone but only in elevated mountain areas suggesting that the representative we found was from a colder environment rather than a tropical one. *Salix* has a more widespread distribution in the northern hemisphere but is also well represented in the southern hemisphere in South America and South Africa in temperate to tropical environments (Davy, 1922). This genus has an unresolved phylogenetic history (Azuma, et al., 2000) but is thought by some to have diverged both northward and southward from East Asia during the Quaternary with the North American willows immigrating both from Asia and from South America in the late Tertiary (Zhen-Fu, 1987).

In our interpretation of the taxa diversity, we expect to find the most recent ecosystem and, therefore, no tropically adapted taxa as these would date back to the Eocene and beyond the expected survival of ancient DNA. Moreover, such taxa would have been outcompeted by a more recent DNA signal from colder adapted taxa if interglacial periods provided ice-free environments. Hence, we evaluated the taxa groups detected in terms of where they overlap in their coldest tolerating common climate zone. For example, for *Sapindaceae,* that is most commonly distributed in northern temperate to tropical climate zones and competes with boreal conifer forest in Canada (Bell, et al., 2011) and therefore exists under more cold conditions than northern temperate regions, we weigh the boreal zone as half of the other zones. *Fagaceae* is known to

occur in the northern temperate to boreal ecotone (Cao, et al., 1995) and is therefore also supportive of a signal of cold temperate to boreal climate. *Pinus*, *Nardus*, *Urtica, Polygonaceae* and *Festuca* are found across several zones, but all overlap in the boreal to temperate zone (see Figure 3).

Two taxa occurring in the temperate to tropical zones were excluded because of uncertainties in assignment, or possible contaminant nature. These were the following:

> i) *Solanaceae* that contains the tomato and potato is a possible food contaminant and was also found in the Greenland samples.However, it is native to South America and could be a true variant in the southern hemisphere, providing a temperate climate signal.
> ii) *Neosprucea* within the *Salicaceae* family is found in South America in the tropical zone and is assigned with a 100% posterior probability in SAP. However, by inspecting the sequence match to *Neospruceae*, we observed two mismatches to the sample sequences, whereas the match to *Salix* only had one mismatch caused by one deletion in the sample sequences (all in error-prone homopolymer sites). Therefore, following the substitution model implemented in SAP (Tavaré, 1986, Munch, et al., 2008a), the most likely phylogenetic assignment is to *Neospucea*. However, the two-mismatches/one deletion could also indicate that there is no good match in the database so we should not assign these sequences further than to family level just as the other two assignment programs suggest, and we therefore choose not to include the tropical signal from *Neospucea* in the climate interpretations.

Other representatives being assigned but not accepted after the conservative approach above include *Fagus* and *Acer*, only occurring in the northern hemisphere in cold/northern temperate to tropical zones. Even though none of these are distributed in the southern hemisphere today and *Nothofagaceae* occupies the ecological niche of *Fagaceae* in the southern hemisphere there are findings from the Mezocoic and Tertiary floras of the South American and Australasian record. *Acer*, *Quercus, Pinus, Fagus* and *Sequoia* have been found as macrofossils in the tertiary southern hemisphere but not confirmed from the microfossil record (Couper, 1960). Following

the substantial *Nothofagus* fossil findings in Antarctica these northern hemisphere taxa have been discarded as being representative of an ancient distributional signal, although it is noteworthy that we have an overlap with three of the five aforementioned taxa.

**Flora and Past Climate**

The trend from the comparison of taxa at their most informative level (genus and family) with a climate-delimited distribution shows the greatest overlap within the range of boreal to temperate zone. Due to the degradation of DNA with time, although slower in cold temperatures, we would not expect to find DNA from a warm/tropical environment earlier than Pliocene with the temperatures from the Taylor Valley that have been stable and cold, below the -17°C found today. We ran a temperature model (Dahl-Jensen, unpublished) based on ice core and ocean core $d^{18}O$ records from the last 3MY that showed average temperatures below –20°C (Bintanja, et al., 2005, de Boer, et al., 2010), (see Appendix E). Furthermore, we would only expect to find a genetic signal from the preserved DNA originating from the most recent ecosystem as any previous signal would be out competed. Therefore, if we interpret the present day climate distribution of these taxa towards their most cold environment we would end up with a northern temperate to boreal climate signal.

**Comparing aDNA with fossil findings**

The warmest previous Antarctic climate supported by fossil findings and climate models is from the late-Pliocene/Pleistocene transition (Salzmann, et al., 2011). Dating of the Sirius Group findings in the Transantarctic Mountains by paleosols and marine diatoms provides an age of ~3 MY with a range from 1.3 – 4.1 MY (Francis, et al., 2007a). The flora findings included mosses, liverworts, conifers and a range of angiosperm plants (Ashworth & Cantrill, 2004). The aDNA findings detect signals of conifers and angiosperms together with grasses, although these are from different families than those reported from the Sirius Group strata and the sediment core DSDP 274, except for *Polygonaceae* found in both the sediment core and with aDNA.

A combination of biodiversity and climate models have confirmed that the climate within this time period could have supported a tundra environment in the Sirius Group

Strata that occurs at an elevation ranging from 1750-3000 m (Haywood, et al., 2002, Ashworth & Cantrill, 2004, Francis, et al., 2007a, Salzmann, et al., 2011). Furthermore, the tundra-vegetated ecosystem in the Transantarctic Mountains is located at a latitude of 85° South, corresponding to roughly 900 km pole wards from the Taylor Valley at 77° South. The climate and temperature range could, therefore, have been warmer at lower latitudes and lower elevations in an environment west of the Transantarctic Mountains.

A boreal climate at these high latitudes would have had long winters without solar heating for 5.5 months although the oceans would have been warmer than today. Hence, the mean annual temperature would have been affected by such long periods without light and the short growth seasons.

The characteristics of a boreal or taiga ecosystem today is that it occurs on the cold range of where forests exist, with conifer trees being in majority in boreal forests that have an understory of vegetation influenced by the moist or dry conditions prevailing locally (Whittaker, 1975). The boreal biome is also characterized today by a cold season of 8 months and where the period with temperatures above 10°C is less than 4 months (Walter & Breckle, 2002). Annual temperatures in the boreal/taiga zone can have a great range depending on a continental or coastal location with maximum temperatures of 30°C and minimum down to -60°C. Precipitation is low but varies according to location and boreal ecosystems can be humid to semi-arid with a mean annual average precipitation level of 40-100 cm/year according to the Köppens Climate Diagram (Rubel & Kottek, 2010).

Boreal ecosystems today in Siberia and Alaska-Yukon have mean annual average temperatures of -10°C, although the more coastal boreal zones have mean annual average temperatures above 0°C (Ricklefs, 2008). Moreover, mixed boreal forests with conifers and deciduous shrubs do occur today e.g. in northern China and Siberia with a tolerance of permafrost as long as the summer temperatures allow for tree growth (Burger & Shidong, 1988, Qu, et al., 2010). Therefore, the boreal ecosystem we detect from the aDNA sequences could potentially have existed at a time when the pollen were deposited in the DSDP 274 site and when the Sirius Group Strata occurred with mean annual temperatures of *ca*. -12°C in the Transantarctic Mountains

and average summer temperatures that were *ca*. 16°C warmer than today as suggested by Francis *et al.* (2007a).

**Fauna DNA**

Diptera, Calliphoridae and Lucilia, representing the blowflies, were replicated and assigned by both programs and other taxa found also included the Coleoptera but were not assigned below order level. The orders are common worldwide and especially the Coleoptera could have been interesting with regards to providing more ecological constraints on the ecosystem; however, with the approaches chosen we could not achieve deeper taxonomical resolution. The blowflies are common pests in livestock and their lifecycle includes laying eggs in necrotic or dead tissues and hatching when conditions are moist and temperatures are above freezing (cool to warm) (Wall, et al., 2001). Overall, there is not much climatic information to be interpreted from these amplified insect sequences.

**Dating**

The cosmogenic $^{10}Be/^{36}Cl$ method and the $^{234}U/^{238}U$ recoil dating method were employed for dating of the Taylor and Suess Glacier basal ice.

The uranium recoil dating method measures the recoil of $^{234}U$ daughter products into the ice matrix from $^{238}U$ mother atoms in the dust or sediment particles in the ice. When constraining the initial input of $^{234}U/^{238}U$ in the ice before the recoil process, it is necessary to account for a pure marine input in clean glacial ice, or an input from basal sediments that would affect this proportion. A ratio of 1.14 for $^{234}U/^{238}U$ is applicable in clean glacial ice, where the dissolved $^{234}U/^{238}U$ component is from sea salt Na and Cl (marine source) and in basal ice, the $^{234}U/^{238}U$ component is 1.0 when the input from basal sediments is taken into account (Aciego, et al., 2011).

Ages were calculated for the glacier samples according to both models of the dissolved $^{234}U/^{238}U$ values of 1.0 or 1.14. Estimates using the model that takes a signal from sediments into account would be considered most appropriate. However, the method is limited by the detection accuracy of the fractional input of daughter products from the dust grains that are ejected into the ice matrix and the fraction that has been ejected within the dust grain itself and should not be counted in the ice. The ability to differentiate among these inputs relies on the measurement accuracy of the

surface volumes of different sediment particles and could cause a significant error in the age estimates. This was shown to account for a re-setting of the age of the basal ice in the Dome C core to 85 KY whereas the glacial ice above was dated to 800 KY (Aciego, et al., 2011).

Taylor Glacier was dated with the $^{238}U/^{238}U$ recoil dating method to a minimum age of 230 KY ± 25 when the signal from bedrock was not taken into account because the measurements could not be performed on the model including sediment input.
The age of the Suess Glacier was dated to a minimum age of 240 KY ± 27 with the model taking a signal from bedrock into account.
Both of these ages are minimum ages and are not constrained by a maximum age, but fall out as very similar in age although a comparison of values within the same model suggest a 100 KY difference between the two sites where Taylor is older than Suess (see Supplementary Material).

The $^{10}Be/^{36}Cl$ dating method was calibrated against surface glacial ice from Taylor Glacier to provide a minimum age estimate of the basal ice from this glacier. These proportions provided a minimum age of 100 KY. Suess could not be calibrated against younger glacial samples from the same source region due to limited sample material available but if we compare the $^{10}Be/^{36}Cl$ ratio from Suess to the Taylor sample we get a signal of an older age, perhaps a minimum of 200 KY (see Supplementary Material). However, this age estimate is uncertain and it needs to be calibrated with younger ice from Suess glacier.

The two dating methods are both affected by processes at the base of the glacier and therefore these estimates are influenced by additional factors that could not be investigated in detail in this study. The ages should be seen as minimum ages whereby re-setting due to unknown processes could potentially decrease the estimates to a large extent.

Therefore, we have to rely on interpreting the likely age from the biological results and do not restrict the time period by the young ages of the supplementary dating methods, but rather use these as a confirmation of an ancient signal of the sample material itself.

The incorporation process, chemical and physical properties, and, ice crystal structure within the amber-coloured silt is different from the coarse textured debris below this zone in the basal ice and suggests that different processes were causing this type of sediment entrainment (Sleewaegen, et al., 2003, Samyn, et al., 2005c, Mager, et al., 2009). The two glaciers have different source areas with Taylor Glacier having its main source from the EAIS and Suess being a local alpine glacier in the Taylor Valley (Fitzsimons, 1996) The source of the fine grained material does most likely stem from a permafrost origin that later got incorporated into the glacier (Fitzsimons, S. and Tison, J.L., personal communication). As Suess Glacier is a small glacier within the Taylor Valley it suggests that the past permafrost source could have been more local to the valley system and not from a distant source across the Transantarctic Mountains like the ice in Taylor Glacier.

Climate models support a warm period at the Plio-Pleistocene transition where sea level was 25 m higher than today with significant reductions of the Greenland Ice Sheet (50%) and Antarctic Ice Sheet (33%). The equator-to-pole temperature gradient was reduced and the EAIS was fully retreated or absent, supporting a dynamic behavior of the EAIS during the mid-late Pliocene period (Dowsett, 2010, Salzmann, et al., 2011).

# CONCLUSIONS AND PERSPECTIVE

Basal ice from Suess and Taylor Glaciers in the Taylor Valley of Antarctica were tested for aDNA content from plants, insects and animals. Only the Suess basal "amber" ice showed results of flora and a few insect taxa. Taylor did not show any results and we believe this is due to the coarse textured nature of the debris that shows no indications of belonging to a terrestrial ecosystem prior to incorporation into the glacier and hence does not contain aDNA from flora and fauna. The biodiversity found in the Suess Glacier points to a palaeo-environment with the variation in taxa suggesting a boreal ecosystem. Previous results from fossil findings point to a warm period in Pliocene around 3MY where tundra vegetation could occur only 500 km from the pole in the Transantarctic Mountains and where pollen had been deposited off the coast of Victoria Land. The biodiversity revealed by aDNA is positioned between these two locations and adds support to a climate much warmer than today. We interpret the biodiversity to stem from a boreal ecosystem, which could exist in the low elevation areas within valleys on the western slopes of the Transantarctic Mountains. The mean annual temperatures could go down to -10°C and summer temperatures could rise above 10°C to support tree growth in less than 4 months a year. With such a climate existing in the area of Victoria Land, the WAIS must have been absent and supports the theory of a dynamic behavior of the EAIS ice sheet during the recent geological time. The preservation of ancient DNA in up to 3 million year-old samples is considered likely due to the cold preservation conditions in Antarctica. Hence, genetic tools can be applied on a broader geographical scale on this continent and offer new means to investigate the past ecological and climatic history of Antarctica.

**Acknowledgements**
This study would not have been possible without the sharing of samples from the Laboratoire de Glaciologie ULB, Brussels, Belgium and I thank Jean-Louis Tison, Reginald Lorrain and Denis Samyn for valuable discussions on basal ice dynamics and processes and the local history of the sample, not to mention helpful sharing of previous findings and new ideas. Moreover, I would like to add a special thank you to Tina Brand (Centre for GeoGenetics) for sharing long hours in the freezer helping me processing the samples. Thanks to Per Hartvig (Natural History Museum of Denmark) for providing valuable discussions on the plant history and ecosystem niches. A huge effort was invested in dating the material so a special thank you to the team of people involved in dating (R. Muscheler (Lund University), A.M. Berggren and Ala Ladahan (Uppsala University) and S. Aciego (University of Michigan)). Last thank you to M. Raghaven for proof reading and to my supervisors for making this study possible.

**Table1**. Presence-absence data at the order level of assigned taxa as determined by each of the three assignment methods; SAP, EcoTag and OGA. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, blue by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample. Dark colours indicate where the taxa have been replicated within labs.

| Samples | Assignment method | Orders | *Brassicales | Caryophyllales | Coniferales | *Fagales | *Malpighiales | *Rosales | *Poales | *Sapindales | Solanales |
|---|---|---|---|---|---|---|---|---|---|---|---|
| #Su98/3 Murdoch | SAP | | | | | | | | | | |
| | EcoTag | | | | | | | | | | |
| | OGA | | | | | | | | | | |
| #Su98/3 Copenhagen | SAP | | | | | | | | | | |
| | EcoTag | | | | | | | | | | |
| | OGA | | | | | | | | | | |

* marks where taxa were replicated on the subset of samples sent to the Murdoch University clean lab, Perth, Australia.

**Table 2**. Presence-absence data at the family level of assigned taxa as determined by each of the three assignment methods; SAP, EcoTag and OGA. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, blue by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample. Dark colours indicate where the taxa have been replicated within labs.

| Samples | Assignment method | *Brassicaceae | *Fagaceae | Pinaceae | *Poaceae | Polygonaceae | *Salicaceae | Sapindaceae | Urticaceae |
|---|---|---|---|---|---|---|---|---|---|
| #Su98/3 Murdoch | SAP | | | | | | | | |
| | EcoTag | | | | | | | | |
| | OGA | | | | | | | | |
| #Su98/3 Copenhagen | SAP | | | | | | | | |
| | EcoTag | | | | | | | | |
| | OGA | | | | | | | | |
| | *Families* | | | | | | | | |

* marks where taxa were replicated on the subset of samples sent to the Murdoch University clean lab, Perth, Australia.

**Table 3**. Presence-absence data at the genus level of assigned taxa as determined by each of the three assignment methods; SAP, EcoTag and OGA. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, blue by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.

| Samples | Assignment method | *Festuca | Nardus | *Neospruceae | Pinus | *Salix | Urtica |
|---|---|---|---|---|---|---|---|
| #Su98/3 Murdoch | SAP | | | | | | |
| | EcoTag | | | | | | |
| | OGA | | | | | | |
| #Su98/3 Copenhagen | SAP | | | | | | |
| | EcoTag | | | | | | |
| | OGA | | | | | | |
| | *Genera* | | | | | | |

* marks where taxa were replicated on the subset of samples sent to the Murdoch University clean lab, Perth, Australia

**Figure 1**. Map of locations where fossils have been found, modified from Flemming et al. (1996) and Francis, et al. 2007a.

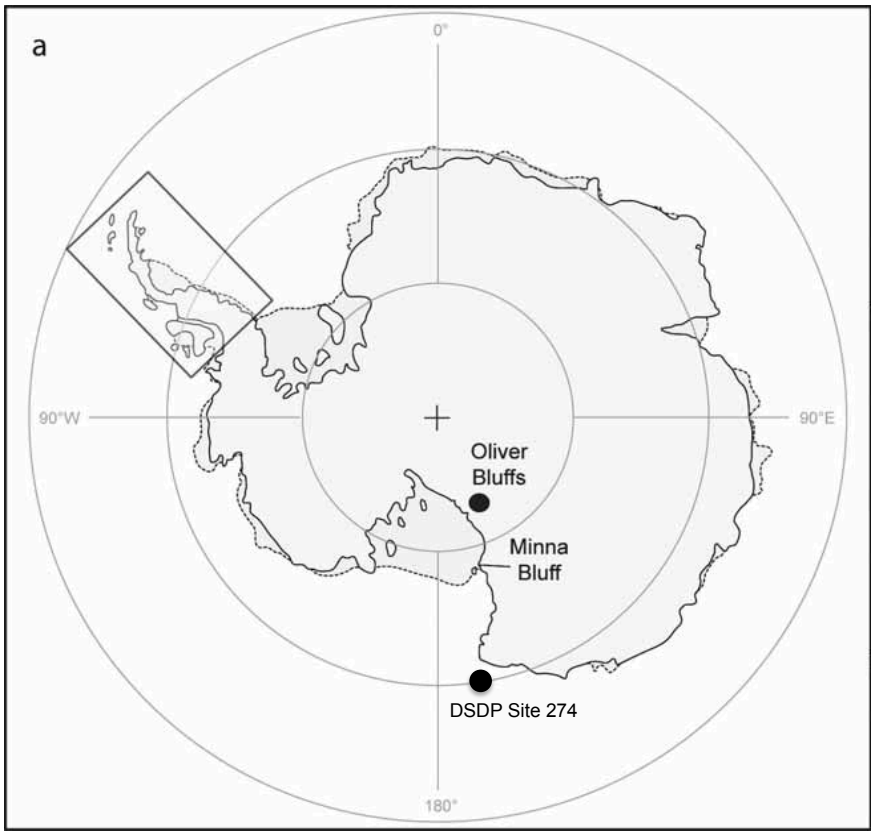


**Figure 2**. Map of Antarctica with sample location marked in yellow within the Dry Valleys of Antarctica. The inset illustrates the two sample sites, Suess and Taylor Glaciers, marked with stars (modified from Samyn (2005a).

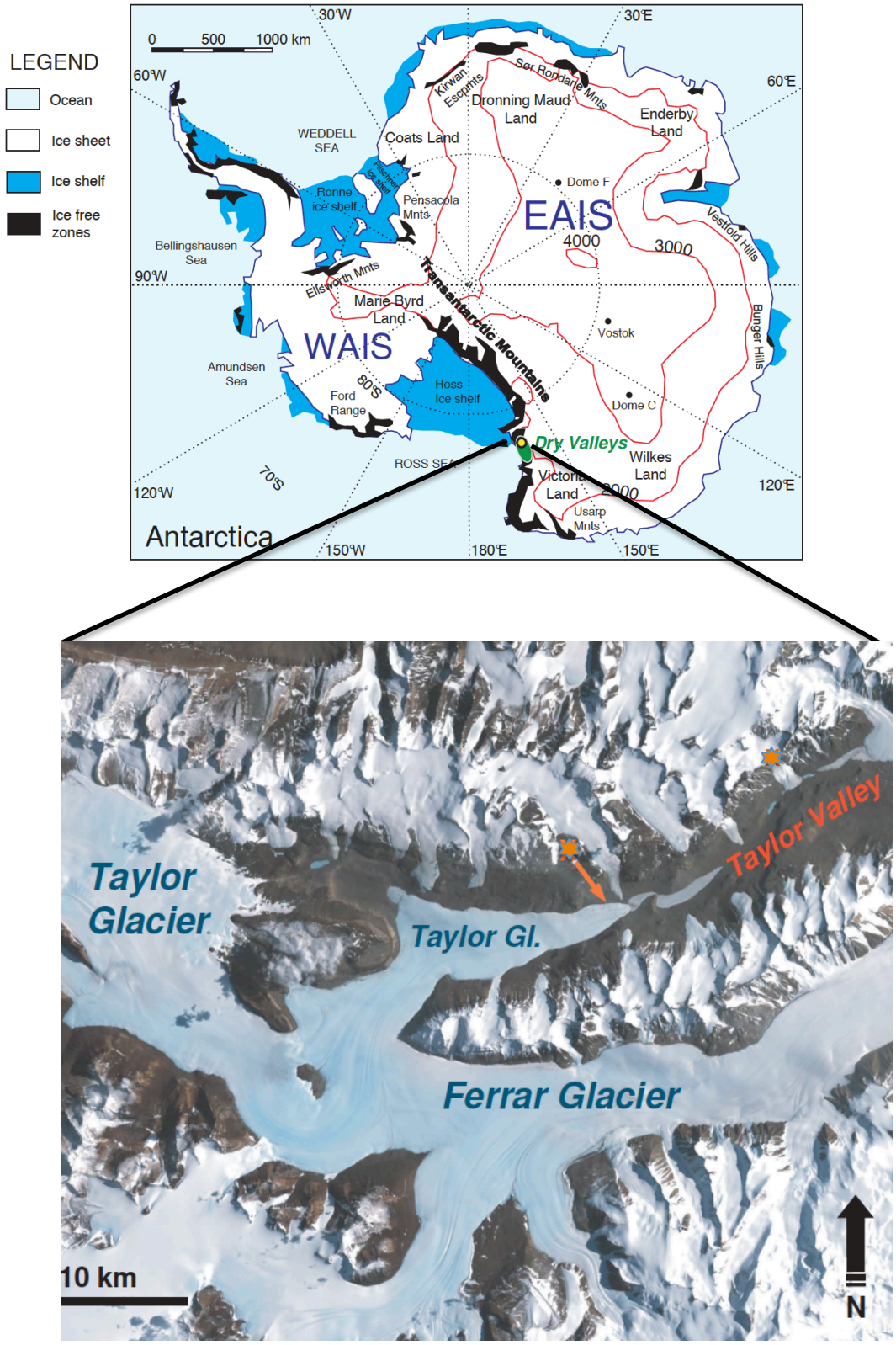
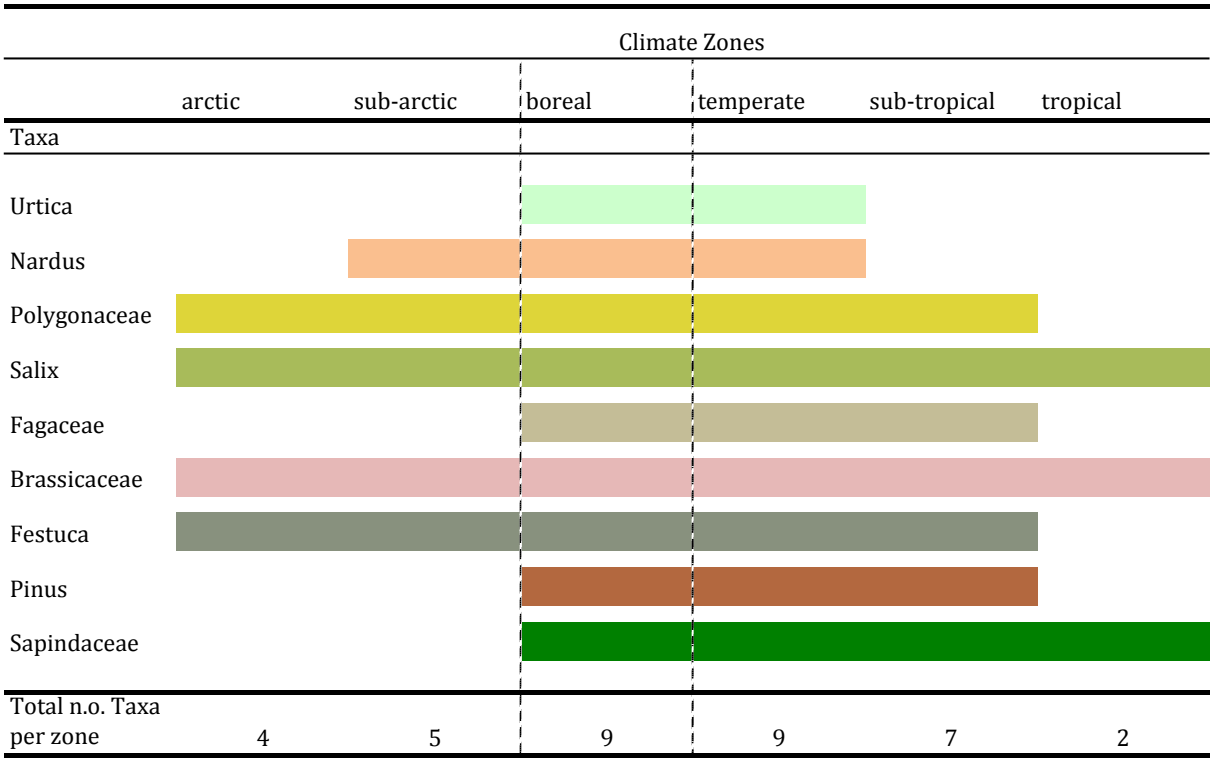


**Figure 3**. Illustration of the distribution of the different taxa groups across climate zones at the most specific taxonomic level assigned.




Vertical lines indicate the coldest common climate zone of the vegetation detected

References

Aciego S, Bourdon B, Schwander J, Baur H & Forieri A (2011) Toward a radiometric ice clock: uranium ages of the Dome C ice core. *Quaternary Science Reviews* **30**: 2389-2397.

Aciego SM, Bourdon B, Lupker M & Rickli J (2009) A new procedure for separating and measuring radiogenic isotopes (U, Th, Pa, Ra, Sr, Nd, Hf) in ice cores. *Chemical Geology* **266**: 194-204.

Anderson JB, Warny S, Askin RA, et al. (2011) Progressive Cenozoic cooling and the demise of Antarctica's last refugium. *Proceedings of the National Academy of Sciences* **108**: 11356.

Ashworth AC & Cantrill DJ (2004) Neogene vegetation of the Meyer Desert Formation (Sirius Group) Transantarctic Mountains, Antarctica. *Palaeogeography, Palaeoclimatology, Palaeoecology* **213**: 65-82.

Azuma T, Kajita T, Yokoyama J & Ohashi H (2000) Phylogenetic relationships of Salix (Salicaceae) based on rbcL sequence data. *American Journal of Botany* **87**: 67-75.

Barrett JE, Virginia RA, Hopkins DW, et al. (2006) Terrestrial ecosystem processes of Victoria Land, Antarctica. *Soil Biology and Biochemistry* **38**: 3019-3034.

Bell FW, Kershaw M, Aubin I, Thiffault N, Dacosta J & Wiensczyk A (2011) Ecology and Traits of Plant Species that Compete with Boreal and Temperate Forest Conifers: An Overview of Available Information and its Use in Forest Management in Canada. *The Forestry Chronicle* **87**: 161-174.

Bintanja R, van de Wal RSW & Oerlemans J (2005) Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437**: 125-128.

Burger D & Shidong Z (1988) An Introductory Comparison of Forest Ecological Conditions in Northeast China and Ontario, Canada. *The Forestry Chronicle* **64**: 105-115.

Cao K-f, Peters R & Roelof AAO (1995) Climatic Range and Distribution of Chinese Fagus Species. *Journal of Vegetation Science* **6**: 317-324.

Convey P, Stevens M, Hodgson D, et al. (2009) Exploring biological constraints on the glacial history of Antarctica. *Quaternary Science Reviews* **28**: 3035-3048.

Couper RA (1960) Southern Hemisphere Mesozoic and Tertiary Podocarpaceae and Fagaceae and their Palaeogeographic Significance. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **152**: 491-500.

Davy JB (1922) The Distribution and Origin of Salix in South Africa. *Journal of Ecology* **10**: 62-86.

de Boer B, van de Wal RSW, Bintanja R, Lourens LJ & Tuenter E (2010) Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic delta(18)O records. *Annals of glaciology* **51**: 23-33.

Denton GH, Prentice ML & Burckle LH (1991) *Cenozoic history of the Antarctic Ice Sheet*. Oxford University Press.

Dowsett H (2010) The PRISM3D paleoenvironmental reconstruction. *Stratigraphy* **7**: 123.

Fitzsimons SJ (1996) Formation of thrust-block moraines at the margins of dry-based glaciers, south Victoria Land, Antarctica. *Annals of glaciology* **22**: 68-74.

Fleming RF & Barron JA (1996) Evidence of Pliocene Nothofagus in Antarctica from Pliocene marine sedimentary deposits (DSDP Site 274). *Marine Micropaleontology* **27**: 227-236.

Francis J, Ashworth A, Cantrill D, et al. (2007b) 100 million years of Antarctic climate evolution: Evidence from fossil plants. *Antarctica, a keystone in a changing world. Proceedings of the 10th International Symposium on Antarctic Earth Sciences, Santa Barbara, California* 19–28.

Francis JE & Poole I (2002) Cretaceous and early Tertiary climates of Antarctica: evidence from fossil wood. *Palaeogeography, Palaeoclimatology, Palaeoecology* **182**: 47-64.

Francis JE, Haywood AM, Ashworth AC & Valdes PJ (2007a) Tundra environments in the Neogene Sirius Group, Antarctica: evidence from the geological record and coupled atmosphere-vegetation models. *Journal of the Geological Society* **164**: 317-322.

Haywood AM, Valdes PJ, Sellwood BW & Kaplan JO (2002) Antarctic climate during the middle Pliocene: model sensitivity to ice sheet variation. *Palaeogeography Palaeoclimatology Palaeoecology* **182**: 93-115.

Lewis A & Marchant D (2008) Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proceedings of the National Academy of Sciences* **105**: 4.

M. Haywood A, Valdes PJ, Sellwood BW & Kaplan JO (2002) Antarctic climate during the middle Pliocene: model sensitivity to ice sheet variation. *Palaeogeography, Palaeoclimatology, Palaeoecology* **182**: 93-115.

Mager S, Fitzsimons S, Frew R, Samyn D & Lorrain R (2009) Composition and origin of amber ice and its influence on the behaviour of cold glaciers in the McMurdo Dry Valleys, Antarctica. *Journal of Glaciology* **55**: 363-372.

Marchant DR, Swisher CC, Lux DR, West DP & Denton GH (1993) Pliocene Paleoclimate and East Antarctic Ice-Sheet History from Surficial Ash Deposits. *Science* **260**: 667-670.

Munch K, Boomsma W, Huelsenbeck J, Willerslev E & Nielsen R (2008a) Statistical Assignment of DNA Sequences Using Bayesian Phylogenetics. *Systematic Biology* **57**: 750-757.

Pearson PN, Foster GL & Wade BS (2009) Atmospheric carbon dioxide through the Eocene-Oligocene climate transition. *Nature* **461**: 1110-1113.

Pollard D & DeConto RM (2009) Modelling West Antarctic ice sheet growth and collapse through the past five million years. *Nature* **458**: 329-332.

Poole I & Cantrill DJ (2006) Cretaceous and Cenozoic vegetation of Antarctica integrating the fossil wood record. *Geological Society, London, Special Publications* **258**: 63-81.

Qu L, Makoto K, Choi DS, Quoreshi AM & Koike T (2010) *The Role of Ectomycorrhiza in Boreal Forest Ecosystem*
*Permafrost Ecosystems*. Springer Netherlands.

Ricklefs RE (2008) *The economy of nature*. W.H. Freeman, New York.

Rubel F & Kottek M (2010) Observed and projected climate shifts 1901-2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorologische Zeitschrift* **19**: 135-141.

Salzmann U, Williams M, Haywood AM, Johnson ALA, Kender S & Zalasiewicz J (2011) Climate and environment of a Pliocene warm world. *Palaeogeography, Palaeoclimatology, Palaeoecology* **309**: 1-8.

Samyn D (2005a) Structural and Geochemical Analysis of
Basal Ice from Taylor Glacier, Antarctica:
On the Role and Behaviour of the Interstitial Fluid Phase. Phd
Thesis, Universit´e Libre de Bruxelles, Belgium, Bruxelles.

Samyn D, Fitzsimons SJ & Lorrain RD (2010) Rotating micro-structures in Antarctic cold basal ice: implications for glacier flow and its interpretation. *International Journal of Earth Sciences* **99**: 1849-1857.

Samyn D, Svensson A, Fitzsimon SJ & Lorrain RD (2005c) Ice crystal properties of amber ice and strain enhancement at the base of cold Antarctic glaciers. *Annals of Glaciology, Vol 40, 2005* **40**: 185-190.

Sleewaegen S, Samyn D, Fitzsimons SJ & Lorrain RD (2003) Equifinality of basal ice facies from an Antarctic cold-based glacier. *Annals of Glaciology, Vol 37* **37**: 257-262.

Souchez R, Samyn D, Lorrain R, Pattyn F & Fitzsimons S (2004) An isotopic model for basal freeze-on associated with subglacial upward flow of pore water. *Geophysical research letters* **31**: 4.

Sugden DE (1996) The East Antarctic Ice Sheet: unstable ice or unstable ideas? *Transactions of the Institute of British Geographers* 443-454.

Tavaré S (1986) Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. *American Mathematical Society: Lectures on Mathematics in the Life Sciences* **17**: 57-86.

Tingey RJ (1991) *The regional geology of the Archean and Proterozoic rocks in Antarctica*. Clarendon Press, Oxford.

Wall R, Pitts KM & Smith KE (2001) Pre-adult mortality in the blowfly Lucilia sericata. *Medical and Veterinary Entomology* **15**: 328-334.

Walter H & Breckle S-W (2002) *Walter's Vegetation of the earth : the ecological systems of the geo-biosphere*. Springer, Berlin ; New York.

Webb PN & Harwood DM (1987) The Sirius Formation of the Bearmore Glacier region. *Antarctic journal of the United States* **22**: 8-12.

Whittaker RH (1975) *Communities and ecosystems*. Macmillan, New York.

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

Willerslev E, Cappellini E, Boomsma W, et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**: 111-114.

Zachos J, Pagani M, Sloan L, Thomas E & Billups K (2001) Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* **292**: 686-693.

Zhen-Fu F (1987) On the Distribution and Origin of Salix in the World. *Acta Phytotaxonomica Sinica* **25**: 6.

# Chapter VII

TITLE:

Ancient DNA analysis on the Kap København formation confirms the presence of an early Pleistocene forested ecosystem

ABSTRACT

Kap København is a geological formation located on the northern tip of Greenland where an arid climate typical of the high Arctic prevails today. A limited number of studies on this site have used micro- and macrofossils to describe a palaeo-environment from the composition of floral and faunal fossil findings. These point to the existence of a forest-tundra covered landscape dated to the early Pleistocene by geological evidence. Climate at this time was mild and has been suggested to be the last interglacial when Greenland was without a stable ice sheet. The exact age of this period, however, has been debated. In this study, we compare results from ancient DNA (aDNA) analysis of ice cores from the Kap København formation to both the fossil record and aDNA results from the Camp Century ice core site and propose that these sites might have co-existed in a more recent geological time. Preservation of aDNA in the permafrost from Kap København could also indicate a younger age of the site than the current age estimate of 1.8-2.5 million years. The composition of aDNA-detected taxa at the Kap København formation supports previous fossil findings of a palaeo-environment with an open boreal ecosystem with trees, shrubs as well as flowering plants. The warm period at the Kap København formation has been suggested to support forests all over Greenland. We propose that this period supported the aDNA findings of forests in the southern parts of Greenland (Dye 3), the northwestern part (Camp Century) and the north (Kap København), just before a stable continuous Greenland ice sheet was established. The age of this period could be younger than previously described and be linked to the glacial orbital cycles when changing from the milder 40,000-year cycles in the Pliocene to the colder 100,000-year cycles in the Pleistocene. This change is thought to coincide with the origin of the Greenland ice sheet around 0.9-1 million years ago. Similar studies in the future, that combine aDNA analysis and alternate dating methods, will ultimately aid in the understanding of the history of Greenland from both a climatic and ecological perspective.

## INTRODUCTION

In the Kap København formation in northernmost Greenland remnants of a rich and diverse flora has been preserved in an area, which today is characterised by Polar desert. The sedimentary succession was interpreted to reflect the early part of an interglacial, which culminated with tree growth at the forest-tundra ecotone with scattered trees and heath landscape (Funder, 1989). This is at odds with regions where such an environment is found today, primarily 2500 km further south in Labrador, Canada. The composition of flora suggest that the climate at the time represented the Low Arctic type - cool, moist boreal climate with summer temperatures of around 10°C and winter temperatures well above the present temperatures in the area of -30°C, and, with mean winter temperatures above -17°C (Funder, et al., 2001). Extrapolation from this indicates that the rest of Greenland was covered by boreal conifer forest in the lowlands, and possibly deciduous forest in the south and western parts, and with the more heath to forest-tundra dominated environment towards the northern parts and where glaciers had retreated from mountain slopes (see Figure 1) (Funder, 1996). Besides the fossil record, the warmer temperatures at this climatic optimum have been supported by ice flow models, the higher solar insolation at this time period (Funder, 1996, Solgaard, et al., 2011), and, temperature extrapolations of the current decreasing trend when moving from the northern latitude towards southern latitudes (Funder, et al., 2001, Solgaard, et al., 2011). The age of the Kap København Formation was determined by a combination of bio-, allo-, and cyclostratigraphic methods as well as palaeomagnetism seting the age to 2.4MY and its duration to ca. 20,000 years (Funder et al. 2001). However, the age of this site has also been placed at the boundary for the Plio-Pleistocene transition, which has been argued to be as young as 0.6MY and as old as 1.8MY based on palaeo-magnetism and foraminifer faunas (Bennike, 1990).

The extensive macrofossil and microfossil analyses carried out at this site, (Bennike, 1990, Funder, et al., 2001) showed some overlap on a subset of the flora taxa with the findings in Camp Century (see Table 4). Ecological details were described from the composition of the pollen and macrofossils from Kap København and pointed to a forest-tundra ecotone with a mix of habitats ranging from dry-land dwarf shrub heaths and scattered trees to more humid bog vegetation, ponds and lakes. Such a

combination of vegetation indicates that the climate was cool and humid in a sub-arctic to low-arctic environment where summer temperatures were at least 10°C to support tree growth (Funder, et al., 1985).

Previous studies have found that aDNA methods can detect taxa where the physical remains of past biota are no longer detectable in sediments (D'Andrea, et al., 2006, Haile, et al., 2007, Sønstebø, et al., 2010, Boere, et al., 2011, Jørgensen, et al., 2011). Hence, the aim of this study was to test if we could both complement as well as supplement the results from previous fossil findings at the Kap København formation with ancient DNA (aDNA) methods that could potentially detect taxa not identified by traditional methods. Moreover, we aimed at comparing the flora and fauna detected between Kap København and the Camp Century site to evaluate if these two locations could have co-existed at a time they were both ice-free. This would add to the understanding of the timing and dynamics of how the Greenland Ice Sheet formed. By comparing the aDNA from flora and fauna from several Greenland sites we can evaluate the past climate conditions at a more regional scale before the sites got preserved in their frozen state and contribute with details on biodiversity and ecological constraints on past climates.

METHODS:

**Sampling and DNA extractions**:

Samples were extracted from eight permafrost cores sampled at the Kap København formation at latitude 82°30'N, Pearyland, Greenland in the summer of 2006, see sample list in Chapter 3. These were sampled in parallel to provide for replication material of exactly the same depth profile. Two locations within the Kap København formation, localities 50 and 75, were included (see map in Figure 2).

Sub-sampling of cores, decontamination and DNA extractions were carried out in accordance to the protocols described in Chapter 3.

**DNA Amplification and Sequencing**:

Short DNA fragments from plants (trnL/rbcL, cpDNA), insects (Cox1, mtDNA ) and mammals (16S, mtDNA) were PCR-amplified and sequenced, as described in Chapter 3.

**Assignment Analyses:**

Three pipelines described in detail in Chapter 3 and 4 were used to filter and assign the plant sequences amplified for the trnL and rbcL markers.

The Cox1 (insect) and 16S mammalian sequences were filtered and assigned according to the same criteria as described for the plant markers in Chapter 4, although only with two programs, EcoTag, available at www.prabi.grenoble.fr/trac/OBITools, and SAP (Munch, et al., 2008a).

**Assigned Presence-Absence Data**:

We generated taxonomic diversity tables for each of the taxonomic levels (order, family and genus) for the two plant markers, rbcL and trnL. Tables were designed to

illustrate the diversity found per sample and method. The taxonomic results are presented as presence-absence data and illustrated in Tables 1-3.

Results from the EcoTag and SAP generated assignments for 16S and Cox1 markers are discussed below.

RESULTS AND DISCUSSION

**Plant diversity in Presence-Absence Data**

Eight samples were tested and two of these provided very few sequence results that passed the filtering and assignment criteria.

In the previous chapters, we found a discrepancy between the diversity detected with the two plant markers, trnL and rbcL. Due to a much more limited sample size this signal is not as obvious here but we follow the same approach as with the other two sites and only present taxa that is found by the trnL marker or where the rbcL marker assignments overlap with those of the trnL marker (see Tables 1-3). Furthermore, due to the fact that samples have not been replicated, we do not follow the same strict selection criteria of only presenting results that have been found at least twice between samples but keep the criteria of only accepting results that were assigned by more than two methods or found in more than one sample. The raw data is listed in Appendix F.

**Flora and Past Climate**

The diversity of taxa (listed in Tables 1-3) found at the genus level support of a boreal to sub-arctic environment with *Dryas*, *Salix*, *Juniperus*, *Saxifraga* and *Cannabis*. Within the grass family we find *Muhlenbergia*, which is normally found in more temperate regions but occurs in central to northern Canada and was only detected by the similarity based assignment method for two different samples, whereas the phylogenetic based assignments only take it to the family level of *Poaceae*. We also find *Juglandaceae* that is today more common in temperate climates but has been found as fossil remains in the high Arctic in Canada at the Prince Patrick and Duck Hawk Bluff formations, supposedly dating to the mid-Pliocene period *ca.* 3MY (Matthews Jr & Ovenden, 1990). A more recent study finds it as both pollen and aDNA in sediments from the Taymyr Peninsula in northwestern Siberia dating to 46-12.5KY (Jørgensen, et al., 2011). This last result shows that *Juglans* was found together with much more cold adapted taxa in the arctic, which in the study is suggested to come from an extinct, more cold adapted species.

On the family and order level, we moreover find *Apiaceae* and Fabales that are both globally distributed plant groups and found in the arctic and the boreal climate zone

today. *Fabaceae* is found both in the macrofossils from the site and in the raw presence-absence data but only with the rbcL marker and in one sample and was therefore not presented below order level where it was confirmed with the trnL marker.

The flora taxa groups overlap with the previous findings but also detect other taxa not found by the traditional methods (see Table 4). Moreover, the taxa groups found by aDNA at the Camp Century site also show close resemblance to the taxa from Kap København. On a taxonomic level, there are five groups (termed groups due to an overlap at the different taxonomic levels) detected with aDNA that coincide with the fossil findings from Kap København and eight aDNA taxa that are shared with the aDNA taxa from Camp Century (six and eleven, respectively, if we do not restrict our analyses to taxa that have been identified by at least two different assignment methods but instead consider the whole information, originating from the full set of sequences characterized, from the presence-absence tables).

Despite Kap København's small dataset, we found an overlap in the climate zone represented by both the fossil and aDNA results. The results are found to exist within a boreal climate zone that allows for tree growth with summer temperatures to a minimum average of 10°C and support families such as *Juglandaceae* and *Apiaceae*. Previous findings of *Thuja* within *Cupressaceae* was used to constrain winters temperatures to be above -17°C, to allow for tree growth (Funder, et al., 2001). The tree taxa found from fossils, within the families *Cupressaceae* and *Taxaceae,* were not detected in the current aDNA dataset. Although we also found a member of the *Cuppresaceae* family it belonged to a different genus, *Juniperus*, which occurs in Greenland today in the southern part and exists generally between the arctic and temperate climate zones. Ongoing replicate experiments will determine if the latter finding is due to a sampling bias rather than a true under-representation of gymnosperms in the aDNA molecules preserved in the ice-core. It is noteworthy though that gymnosperms were already identified at the DNA level at Camp Century, that showed the presence of both *Taxaceae* and *Picea*. Within the angiosperms we find Fabales, *Salicaceae*, *Rosaceae* and *Saxifragaceae* that overlap with the fossil data. The following groups were shared at the aDNA level between Camp Century

and Kap København; *Apiaceae*, *Poaceae*, *Cupressaceae, Fabaceae, Juglandaceae, Poaceae, Rosaceae (Dryas), Salicaceae (Salix), Saxifragaceae (Saxifraga)*.
Overall, we find taxa that all occur in the boreal zone today. It is not an uncommon finding in sedimentary DNA analyses if not all of the taxa groups detected by aDNA are found by macrofossil or pollen (Jørgensen, et al., 2011). Especially in ecotones where some trees may be existing at the boundary of their distribution zone, they may not produce seeds that would be detected in the fossil record (Zazula, et al., 2003). However, such taxa would still leave degraded tissues behind, and therefore possibly aDNA. Fossil findings only provide evidence of what could have been preserved in the sediments and not what had once existed but did not get preserved as an assignable fossil (Jørgensen, et al., 2011). Following the same argument, aDNA might not detect all the previous taxa groups that existed or were found as fossils if not all were equally well preserved. Furthermore, our sample material only represents a limited sub-set of the extensive sampling that was done on the Kap København formation.

The current composition of taxa found by aDNA support a boreal ecosystem with mixed vegetation and by the variety of taxa detected, we suggest an open habitat with mixed herbs and shrubs and random tree growth in sheltered areas from the coast.

Following the Köppens Climate Classification Map and Whittaker diagram (Whittaker, 1975), a climate zone of the boreal/taiga type is characterized by annual average temperatures below 5°C and annual average precipitation of 40-100 cm/year. The summers have midnight sun and would be compensating for the long cold and dark winters if temperatures during the warmest months were above 10°C. In the study by Bennike (1990), it was moreover inferred that the winters could not have been colder than –15°C to –17°C and that the precipitation was around 50cm/year with heavy snowfall in winters to shield the trees. This is comparable with the findings of boreal to temperate taxa in the current study.

**Faunal DNA**

Previous fossil findings from the site have only detected a few vertebrates like rodents and a fish from this time period. Reindeer are believed to have immigrated to this area around 8KY but older fossil evidence is limited. The preliminary aDNA results of the 16S fragment showed a signal of bovids from two different samples and assigned by two programs. However, we ascribe the bovid results as contamination as we cannot distinguish it from a contaminant bovid signal originating from the domesticated genus *bos* (cattle) also contained within this family and that has been shown, together with other common domestic animals, to represent a source of DNA contamination in PCR assay (Leonard, et al., 2007).

**Previous aDNA results**

The DNA findings from this site are on the boundary of how long we would expect to be able to find aDNA preserved in the Arctic sediments. Sampling was performed in the permafrost layers that are representative of the time period that the Kap København formation has been dated from. Previous unsuccessful attempts performed in the Willerslev et al. study (2007) might be explained by several reasons.
Firstly, the amplifications in 2007 were performed with a different polymerase enzyme, the *pfu* DNA polymerase (Stratagene) that does not amplify miscoding lesions due to damage, e.g. uracil produced by deamination of cytosine that leads to a common damage pattern of C/G to T/A substitutions. In this study we used the Platinum Taq High Fidelity polymerase (HiFi, Invitrogen) that is more successful in by-passing miscoding lesions, approximately on the order of one magnitude compared to the *pfu* polymerase (Heyn, et al., 2010).
Secondly, a different extraction protocol was utilized in the current study where larger amounts of soils (10 g) were extracted, whereby DNA can be concentrated from a larger initial volume in 50 mL tubes (MOBIO) compared to the smaller amounts of sediment (0.25 g) extracted in 2.5 mL tubes (modified extraction protocol from Bulat et al. (2000) (see Chapter 3 for details).
Lastly, amplification products were regularly cloned before sequencing instead of the deep sequencing approach on the Roche GS-FLX platform in this study. The combination of aforementioned factors with a more sensitive deep sequencing

approach explains why the current approach could detect rare amplification results in samples that had previously not provided any DNA signal.
In addition, we extended the sample material from the 2007 study with another three samples taken in 2006 that also contributed to the total diversity found in the samples.

**Age and DNA preservation**

The permafrost at the Kap København site has remained frozen since the climate changed from the warm interglacial period that sustained the boreal ecosystem (Funder, et al., 2001). This is supported by the well preserved fossils from this period and from the lack of a more intermediate and more cold tolerant flora in above lying layers following this warm period (Bennike, 1990, Funder, et al., 2001). The cold and anoxic conditions favorable for preserving pollen and macrofossils at the site are also favorable for DNA preservation (Willerslev, 2004a). The permafrost conditions in the Kap København formation are believed to have remained frozen and cold since the end of the warm climate period that supported the boreal vegetation (Funder, 1996). Stable, cold and dry conditions are known to be favorable for DNA preservation. However, with a mean annual temperature of around –14°C, we are getting results previously not believed to be possible if the age of Kap København site is indeed 1.8-2.5MY, and not younger as was proposed in the Camp Century chapter. An extrapolation from the Arrhenius equation with regards to DNA preservation would suggest preservation until approximately 1 million years and not over this time period (Willerslev, 2004a).
However, three areas of interest are important to take into account. First, the fact that the expectation of finding preserved DNA under average temperatures of –15°C were extrapolated from a simplified model testing the robustness of DNA in solutions from 0 – 70°C, and not taking factors into account other than a pH of around 7.4 (Lindahl & Nyberg, 1972, Willerslev, 2004a). Second, the preservation conditions are optimal in that the permafrost is dry with a very low water and oxygen content – two factors in addition to the low temperatures known to decrease the rate of DNA decay (Willerslev, 2004a). Third, the age estimate of the Kap København formation is challenged from modeling studies of combined ocean and ice core temperature profiles and the effect of orbital glacial cycles (Bennike, 1990, de Boer, et al., 2010). Models describing orbital glacial cycles detect the cooling trend that started in the late-Pliocene and offer an alternative point of view to when the Greenland ice sheet

(GIS) began building up to its current state from the glaciations that followed the onset of the 100KY cycles from around 0.9 – 1MY (Bintanja & van de Wal, 2008, Crowley & Hyde, 2008). These results could suggest that the origin of GIS is actually younger than previously thought and the DNA is thereby younger than 1.8 – 2.5MY. A younger age of around one million years would be an alternative to the 1.8 – 2.5MY and could explain the DNA survival in these samples.

Furthermore, in terms of DNA preservation, as mentioned in previous chapters, it seems that other and more complex factors need to be taken into account when extrapolating how long DNA can survive in frozen conditions, for example pH, anoxic or oxidative conditions (Willerslev, 2004a). Moreover, abiotic environmental factors like oxygen availability, grain size, surface to volume ratio, mineralogical composition of the sediment particles as well as the organic content represent additional components that could interact and influence adsorption and preservation of DNA in basal ice or sediments (Corinaldesi, et al., 2008).
There is a great need to investigate these factors to be able to validate and explain the results of these samples.

In the 2007 study by Willerslev et al. (2007) the lack of aDNA results were supported by amino acid racemization results indicating low levels of preservation of biomolecules. However, after additional amino acid racemization tests on both Camp Century and Antarctic basal ice, we know that the amino acid racemization signal can be biased when performed on sediments with few replicates, as a high intra-sample variation was detected on material from both the Camp Century and Antarctic samples that were influenced by ancient bacterial degradation signals skewing the ratio of chemically decayed amino acids (see Supplementary Material for a more thorough discussion). Therefore, permafrost (and in these cases basal ice) cannot be assumed to follow an amino acid racemization pattern for a closed system and hence does not directly reflect the DNA preservation level within a sample (Willerslev, et al., 2007).

## CONCLUSIONS AND PERSPECTIVES

aDNA results from the Kap København formation showed similarities to previous fossil findings and supported the proposition of a climate zone with vegetation that supports an ecotone with boreal and tundra taxa. These results shows that aDNA analysis can be supplementary and complementary to fossil findings when reconstructing palaeo-environments from small sample sets on material that contain low diversity. Moreover, the aDNA data shared taxa groups with aDNA findings from the Camp Century study. We propose that this might be due to a co-existence of these two ecosystems within the same warm interglacial period although with a warmer local climate at the Camp Century location.

If the two sites co-existed, it would fit with a time period where all of Greenland was forest covered prior to a GIS-covered Greenland, as suggested by Funder et al. (2001). The additional basal ice study of Dye 3 also showed a forested ecosystem by aDNA and we propose that all three sites date to the period of 0.9-1MY when the orbital glacial cycles changed to the 100KY periodicity. The forest-tundra ecosystem found at Kap København and the boreal open forest ecosystem at Dye 3 could have existed together with the boreal to northern temperate ecosystem at Camp Century in a period before the onset of a continuous GIS when temperatures were warmer than in any following interglacials. However, such a hypothesis is mainly relying on aDNA results and climate models and needs to be verified by independent age estimates and further replicate samples of the diversity to allow for a refining of such an extensive theory.

The proposal of a younger age for the Kap Købehavn site than the previous dating of 1.8-2.5MY, could explain why we were able to amplify preserved DNA from a past palaeo-environment. However, more investigations need to be performed to clarify these age differences. One way to pursue this further could be by cosmogenic nuclide dating of $^{26}Al/^{10}Be$ in the sediments, which could provide another age measure to compare with the geological evidence.

Dating experiments together with further analyses of the molecular diversity preserved in replicate samples, within an integrative framework would supplement previous work and could enable a refining of the age estimate of the Kap København formation and the aDNA results retrieved from here.

**Acknowledgements**

I would like to thank Svend Funder for valuable discussions on the palaeo-environmental history and previous findings at the Kap København formation and to Kurt Kjær for helpful discussions and aid in sample descriptions and abiotic measurements. Thank you to Morten Rasmussen, Aurélien Ginolhac and Ludovic Orlando for developing pipelines to sort, filter and assign the data. Additionally, I thank Kenneth Andersen and James Haile (Centre for GeoGenetics) for laboratory help. Thank you Rute Fonseca and Tobias Mourier for helpful scripts and discussions on data. Thank you Maanasa Raghaven for proofreading. Last thank you to my supervisors for making this study achievable and providing sample material.

**Figure 1**. Illustration of how Greenland could have looked at the time Kap København was a forest-tundra ecosystem (From Funder, 1996).

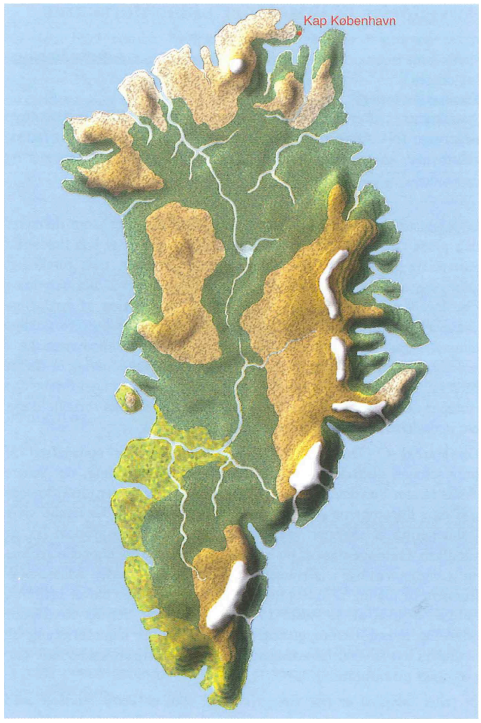



**Figure 2**. Map of Kap København with different sample positions. Localities 50 and 75 used in this study are marked with green circles (From Bennike 1990).

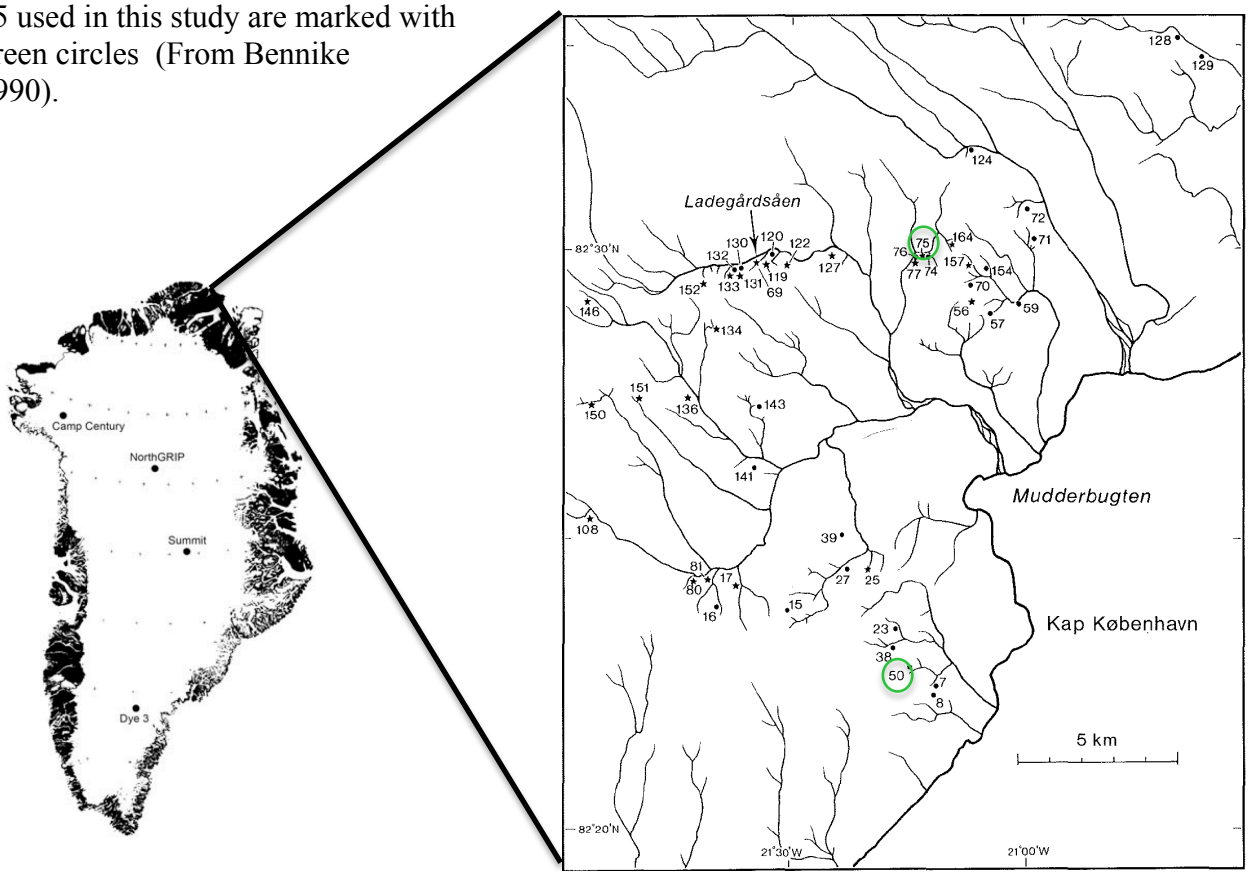
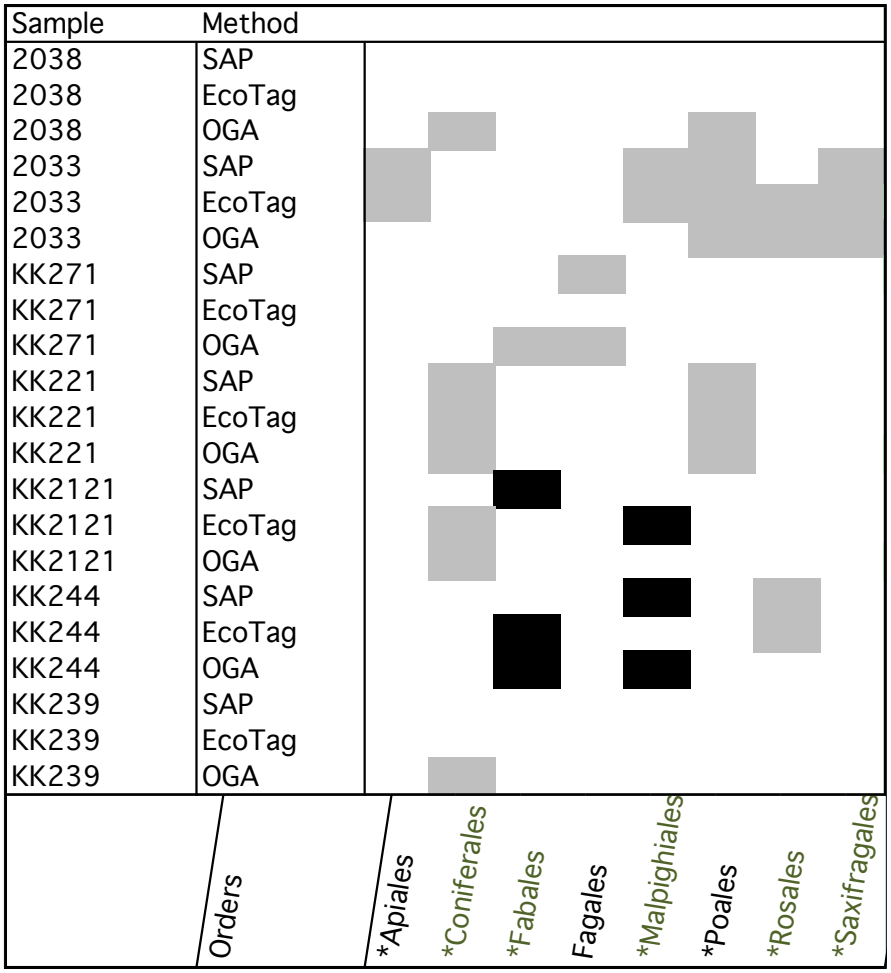
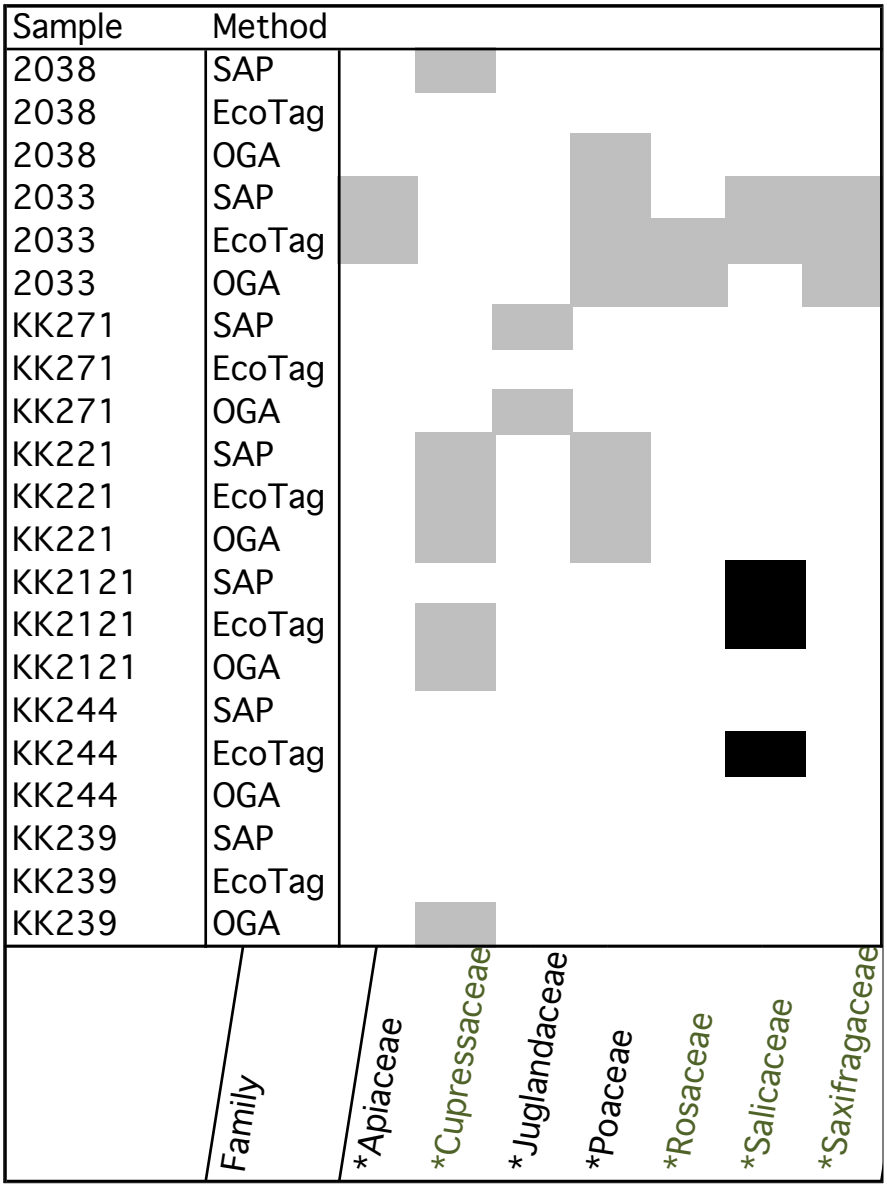


**Table 1**. Presence-absence data at the order level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL and black by rbcL.



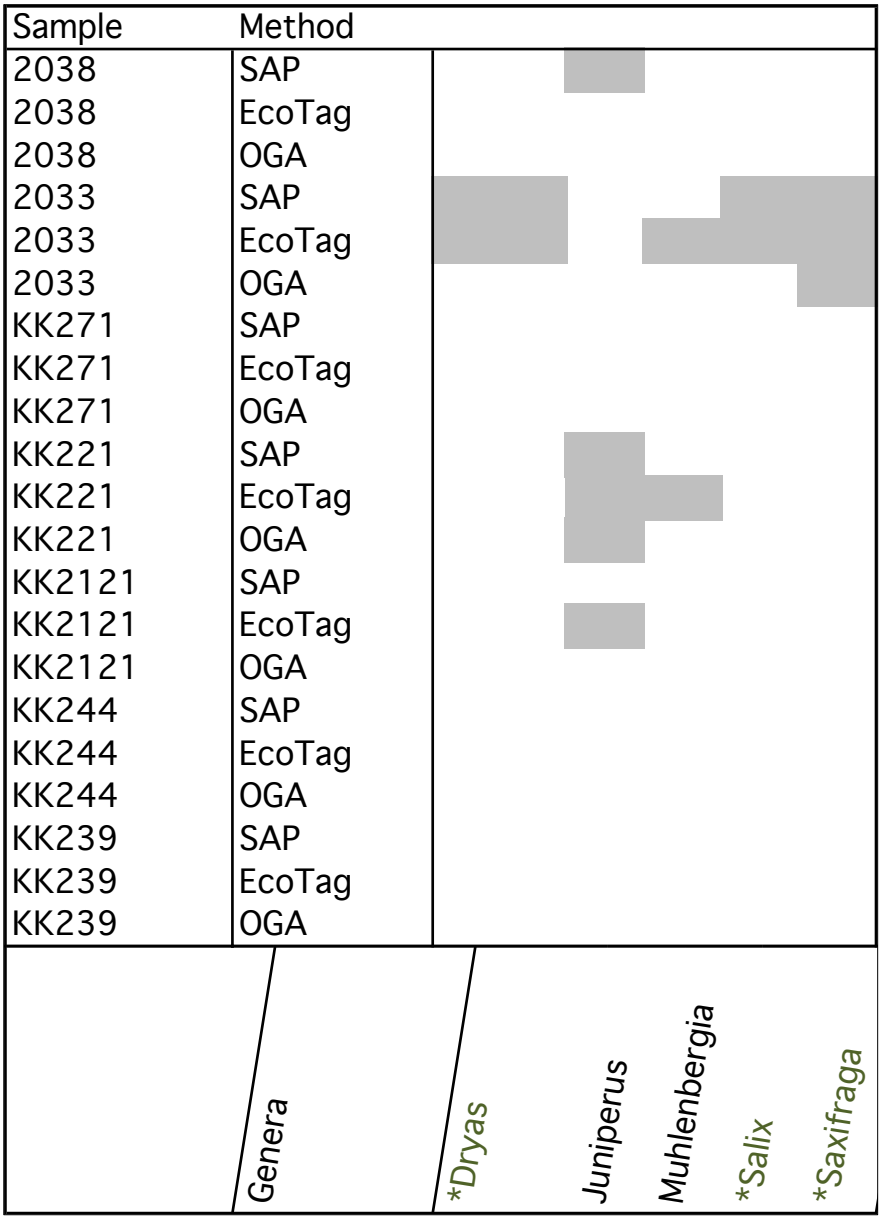

Green writing indicates the taxa that were found in the fossil record.
Black writing indicates the taxa that were uniquely found by aDNA.
* Illustrate that the taxon was also found in Camp Century.

**Table 2**. Presence-absence data at the family level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL and black by rbcL.




Green writing indicates the taxa that were found in the fossil record.
Black writing indicates the taxa that were unique found by aDNA.
* Illustrate that the taxon was also found in Camp Century.

**Table 3**. Presence-absence data at the genus level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL and black by rbcL.




Green writing indicates the taxa that were found in the fossil record.
Black writing indicates the taxa that were unique found by aDNA.
* Illustrate that the taxon was also found in Camp Century.

**Table 4**. List of taxa detected by aDNA and fossils at the family level from the Kap København site, and, aDNA findings from the Camp Century site.

| aDNA | fossil | aDNA |
|---|---|---|
| Kap København | Kap København | Camp Century |
| Apiaceae | | Apiaceae |
| | | Asteraceae |
| | | Asparagaceae |
| | | Anacardiaceae |
| | Betulaceae | Betulaceae |
| | | Brassicaceae |
| | | Bryaceae |
| | Caryophyllaceae | Caryophyllaceae |
| Cupressaceae | Cupressaceae | Cupressaceae |
| | Cyperaceae | Cyperaceae |
| Equisetaceae | Equisetaceae | Equisetaceae |
| | Ericaceae | Ericaceae |
| | | Euphorbiaceae |
| Fabaceae | Fabaceae | Fabaceae |
| | | Fagaceae |
| | Juncaceae | Juncaceae |
| Juglandaceae | | Juglandaceae |
| | | Oleaceae |
| | | Orobanchaceae |
| | Papaveraceae | Papaveraceae |
| | | Passifloraceae |
| | Pinaceae | Pinaceae |
| | | Plumbaginaceae |
| Poaceae | | Poaceae |
| | Ranunculaceae | Ranunculaceae |
| | Polygonaceae | Polygonaceae |
| | | Rapateaceae |
| Rosaceae | Rosaceae | Rosaceae |
| Salicaceae | Salicaceae | Salicaceae |
| | | Sapindaceae |
| Saxifragaceae | Saxifragaceae | Saxifragaceae |
| | | Solanaceae |
| | Taxaceae | Taxaceae |
| | | Urticaceae |
| Ulmaceae | | Ulmaceae |

References

Bennike O (1990) The Kap København Formation: stratigraphy and palaeobotany of a Plio-Pleistocene sequence in Peary Land, North Greenland. *Meddelelser om Grønland, Geoscience* **23**: 85.

Bintanja R & van de Wal RS (2008) North American ice-sheet dynamics and the onset of 100,000-year glacial cycles. *Nature* **454**: 869-872.

Boere AC, Rijpstra WI, De Lange GJ, Sinninghe Damste JS & Coolen MJ (2011) Preservation potential of ancient plankton DNA in Pleistocene marine sediments. *Geobiology* **9**: 377-393.

Corinaldesi C, Beolchini F & Dell'Anno A (2008) Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Molecular Ecology* **17**: 3939-3951.

Crowley TJ & Hyde WT (2008) Transient nature of late Pleistocene climate variability. *Nature* **456**: 226-230.

D'Andrea WJ, Lage M, Martiny JBH, Laatsch AD, Amaral-Zettler LA, Sogin ML & Huang YS (2006) Alkenone producers inferred from well-preserved 18S rDNA in Greenland lake sediments. *Journal of Geophysical Research-Biogeosciences* **111**.

de Boer B, van de Wal RSW, Bintanja R, Lourens LJ & Tuenter E (2010) Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic delta(18)O records. *Annals of glaciology* **51**: 23-33.

Funder S (1989) Quaternary Geology of the Ice-Free Areas and Adjacent Shelves of Greenland. *Quaternary Geology of Canada and Greenland* **K-1**: 741-792.

Funder S (1996) Et Anderledes Klima. *Varv* **4**: 18.

Funder S, Abrahamsen N, Bennike O & Feylinghanssen RW (1985) Forested Arctic - Evidence from North Greenland. *Geology* **13**: 542-546.

Funder S, Bennike O, Bocher J, Israelson C, Petersen KS & Simonarson LA (2001) Late Pliocene Greenland - The Kap Kobenhavn Formation in North Greenland. *Bulletin of the Geological Society of Denmark* **48**: 117-134.

Haile J, Holdaway R, Oliver K, et al. (2007) Ancient DNA Chronology within Sediment Deposits: Are Paleobiological Reconstructions Possible and Is DNA Leaching a Factor? *Molecular Biology and Evolution* **24**: 982-989.

Heyn P, Stenzel U, Briggs AW, Kircher M, Hofreiter M & Meyer M (2010) Road blocks on paleogenomes - polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Research* **38**: e161.

Jørgensen T, Haile J, Mølller P, et al. (2011) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* no-no.

Jørgensen T, Haile J, Møller P, et al. (2011) A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology* no-no.

Leonard J, Shanks O, Hofreiter M, et al. (2007) Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science* **In Press, Corrected Proof**.

Lindahl T & Nyberg B (1972) Rate of Depurination of Native Deoxyribonucleic Acid. *Biochemistry* **11**: 3610-3618.

Matthews Jr JV & Ovenden LE (1990) Late Tertiary Plant Macrofossils from Localities in Arctic/Subarctic North America: A Review of the Data. *Arctic* **43**: 364-392.

Munch K, Boomsma W, Huelsenbeck J, Willerslev E & Nielsen R (2008a) Statistical Assignment of DNA Sequences Using Bayesian Phylogenetics. *Systematic Biology* **57**: 750-757.

Solgaard AM, Reeh N, Japsen P & Nielsen T (2011) Snapshots of the Greenland ice sheet configuration in the Pliocene to early Pleistocene. *Journal of Glaciology* **57**: 871-880.

Sønstebø JH, Gielly L, Brysting AK, et al. (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*.

Whittaker RH (1975) *Communities and ecosystems*. Macmillan, New York.

Willerslev E (2004a) Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends in Ecology & Evolution* **19**: 141-147.

Willerslev E, Cappellini E, Boomsma W, et al. (2007) Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland. *Science* **317**: 111-114.

Zazula GD, Froese DG, Schweger CE, et al. (2003) Palaeobotany: Ice-age steppe vegetation in east Beringia. *Nature* **423**: 603.

# FINAL CONCLUSIONS AND PERSPECTIVES

## MAJOR CONTRIBUTIONS AND FUTURE PERSPECTIVES OF THE STUDY

Preserved ancient DNA (aDNA) from glaciated or permafrost locations offer a means to investigate the palaeoclimate and ecosystems of the past through ice cores or basal glacier samples. Novel DNA sequencing and analysis techniques have rapidly improved the data availability and throughput, which was more traditionally only available through macro- and micro- fossils from areas free of ice in Greenland and Antarctica. Such knowledge can elucidate answers to more of the classical questions within climate and biodiversity dynamics through time, as well as modes of adaptation to changing environments within different ecological biomes.

The link between biology and climate in ice material can be studied using different approaches. This thesis has investigated aDNA from flora and fauna in the basal ice samples and reconstructed the palaeo-environments and climatic characteristics from the biodiversity detected. Another angle was presented in the review chapter where biological remains within the ice is linked to past climate from a microbiological perspective. Research on bacteria within ice has increased and developed from being a model of how to study and detect extraterrestrial life to studies of microbial diversity, viable or ancient, linked to climatic fluctuations that could act as an independent climatic proxy.

The main contributions of this thesis have been to expand on the applicability of methods available to study aDNA in basal ice, from both the sequencing and data handling perspectives, to the evaluation of where and under what conditions aDNA is preserved in frozen environments. Abiotic measurements on chemical and physical properties of the soil were also performed to evaluate the likelihood of finding preserved terrestrial genetic remains from flora and fauna. Moreover, the study has assessed how assignments of amplified mini-barcodes from past flora to DNA sequence databases might affect the degree of detection of biodiversity at different taxonomical levels.

Comparisons of the amplified biodiversity from all of the three locations with present day ecological climate zones have provided new climate constraints on the palaeo-ecosystems from both Camp Century and Suess Glacier, and, complemented previous fossil findings in Kap København. We estimated the minimum age of the basal ice materials by three different dating methods: $^{10}Be/^{36}Cl$, $^{234}U/^{238}U$ and OSL. A

broad range of cross-disciplinary work was evoked for increasing the understanding of the basal processes within the ice and how these were affecting mineralogical, isotopic and biological elements that can elucidate more information on the age and dynamics of the ice. Based on the aDNA results, minimum ages of the sites, the ecologically-derived climate information, and, time periods when Greenland and West Antarctica were last free of ice coverage, were suggested. Such age determinations will add to the ongoing discussions on the stability of these ice sheets and contribute to the models describing past and future responses of the polar ice sheets to climatic fluctuations.

Future work on the biome in ice has the potential to increase the amounts of data considerably with the continuous development of sequencing and analytical technologies. Regardless of whether the biological interest in ice is, (i) aDNA and palaeo-environmental based, or, (ii) microbiological based research interest, both areas would benefit from continuously adapting to the highest – although continuously improving – standards within clean laboratory work for contamination-prone samples. Contamination can occur at many stages when working with such delicate material and will hamper the overall understanding of past and present dynamics of climate, ice sheets and ecosystems if not taken into account.

The next-generation sequencing techniques lift the field of ice core and sedimentary aDNA research into an era of deep sequencing opportunities of samples low in DNA concentration. With the increasing amounts of data there is a need for bioinformatics to design appropriate pipelines for verifying and sorting the data. In future work, the aim should be focused on improving the existing pipelines with regards to adding tests for damage patterns in the amplified sequences by statistical means to verify the ancient nature of these, and stress that biases in terms of sequence diversity need to be evaluated to account for false positives due to damage, amplification errors or assignment bias against reference sequence databases.

The field of aDNA studies in basal ice and permafrost samples draws upon diverse cross-disciplinary collaborations between genetics, glaciology, bioinformatics, geology, physics and geochemistry. It is a privilege and adventure to be a contributing part of such an interesting field where joint efforts drawn across a multitude of fields provide unique access to new data not previously possible

# SUPPLEMENTARY MATERIAL

# Dating methods appropriate for ancient basal ice


## Abstract

Basal ice has been dated by a series of different approaches, depending on the characteristics of the material in the ice and its handling since recovery. The methods involve radiometric dating of volcanic material in the ice, stable isotope measurements of gasses and their composition, and radiometric dating of radioactive isotopes from the silty ice and the entrained dust or debris particles in the ice. These methods on their own are often not applicable to a broad range of basal ice samples and a well-established procedure to estimate the age of basal ice samples has not yet been found.

Here we present a combination of biological and physical methods (recoil dating by $^{238}$Uranium/$^{234}$Uranium, optically stimulated luminescence dating, $^{10}$Be/$^{36}$Cl cosmogenic dating and Amino Acid Racemization dating) to date basal ice from Greenland and Antarctica, which together provide an age frame that can point to a minimum age of when these areas were last free of ice. The applicability of each method to these samples differs and here we present both the advantages and disadvantages of the techniques. The resulting ages represent a span that can define a minimum age of the ice for when material was incorporated into it.

# Introduction

Studying the characteristics of the basal ice and its debris content for stable isotopes, gas composition, ice crystallography and ion composition provides insights on the formation, ice sheet behaviour and dynamics of the lowermost ice, but determining the age of the basal ice in such a mixed environment is complex. Dating of basal ice in deep ice cores is challenging due to the deformation processes occurring at the bed that modify the basal ice differently than the above lying ice through thermal, strain and hydraulic conditions close to bedrock [*Knight*, 1997]. Formation of the above lying glacier ice happens mainly by compaction of seasonal snow precipitation that accumulates and changes first into firn and then into glacier ice due to the increased pressure of the ice mass [*Johnsen et al.*, 2001]. Besides the continuous formation process glacier ice has different flow characteristics than basal ice [*Knight*, 1997]. Basal ice further differs on one major area from the above clear glacier ice by containing a significant amount of debris, entrained as a consequence of ice flow dynamics of the sheets or glaciers. Both supra-glacial or sub-glacial mechanisms can cause debris entrainment during movement of the glacier, which will lead to mixing of the debris and the ice [*Samyn*, 2005a]. However, the amount of supra-glacial entrainment is minor compared to the debris volume entrained by sub-glacial processes from the bedrock (Souchez et al, 1993, Tison et al, 1989, Samyn, 2005). Basal ice can further be formed by an incorporation of relict glaciers that were formed prior to the formation of the ice sheet, which are then mixed with the advancing growing ice sheet [*Souchez*, 2000; *Souchez et al.*, 1994]. Due to these different formation mechanisms the basal ice is distinct from the overlying ice, by being conditioned primarily of the processes operating at the bed and by the chemical and physical characteristics of the debris [*Knight*, 1997].

The age of the basal ice is of interest for providing a *terminus post quem* for ice accumulation. This provides important data for climate modeling (identifying when a region was last ice free), and a better understanding of the local processes that lead to the build-up of ice sheets. The age of the basal ice will link climate history to palaeo-ecosystems and is critical for a more detailed understanding of the processes preceding great climate shifts.

So far, no dating method has been established as appropriate on a broad age scale for basal ice. There are however a myriad of different potentially available methods of which some have been applied for dating basal ice;

i) Radiometric dating of radioactive isotopes $^{40}Ar/^{39}Ar$, $^{40}K/^{40}Ar$ [*Sugden et al.*, 1995]
ii) U-series dating [*Aciego et al.*, 2009]
iii) Cosmogenic nuclide dating using radioactive isotopes like $^{10}Be$ and $^{36}Cl$ [*Willerslev et al.*, 2007b]
iv) Radiometric exposure dating e.g. optical stimulated luminescence dating [*Willerslev et al.*, 2007b]

In addition, in the presence of ancient biological material, methods such as amino acid racemization (AAR) and molecular clock analysis of specific genomic regions with fairly stable mutation patterns can also be used for dating [*Collins et al.*, 1999; *Hebsgaard et al.*, 2005; *Willerslev et al.*, 2007a].

The radioactive isotope approaches are limited in the way that they do not necessarily reflect the age of the ice but rather the age of volcanic ash or minerals or the time since they were incorporated into the ice.

Therefore different methods supply ages that represent two kinds of time scenarios, either by providing a minimum age for incorporation of the dated debris or dust material, or a maximum age of the basal ice by methods dating the time elapsed from snowfall to the formation of the basal ice. Combining different methods can then produce an age range that is bounded by a minimum and maximum from both ends of the scale, will supply a time frame of the age of the ice. It is therefore important to decide which fraction of the ice to date to get the best representation of the age of the basal ice.

In a previous study on basal ice samples from the Dye 3 ice core in Greenland [*Willerslev et al.*, 2007b] a combination of different methods were used to estimate an age range for the basal ice.

Here we present possibilities and limitations of combining biological and physical methods to date basal ice samples from both Greenland and Antarctica. The potential age frame of the studied samples is very broad, from 130,000 (130 ky) up to several

million years, and we consider a range of dating methods appropriate to cover this span.

## *Setup of study*

The samples included in this study are from the basal ice in the Camp Century ice core and from Taylor and Suess Glaciers in the Dry Valleys of Antarctica (figure 1). Additional and previously dated samples from the Dye 3 and GRIP core have been used as comparison for some of the methods. The time when these polar regions were last free of ice is debated and range in age from the Eemian (130 – 114 ky) for the Greenland sites on the margin of the present Greenland Ice Sheet (GIS) to the Miocene (8 my) for the glaciers in the Dry Valleys of Antarctica [*Otto-Bliesner et al.*, 2006; *Sugden et al.*, 1995].
To be able to cover such an age span we chose the methods for their ability to cover both a minimum age (of dust, debris or biomolecular incorporation) and a maximum age (time of precipitation build into glacier). Last these methods were chosen as they had all been tested on ice material previous to this study [*Aciego et al.*, 2009; *Willerslev et al.*, 2007b].

Each of the four methods will be presented in the following sections individually with introductions to the methods and how they have been applied on the samples and what the results showed.

# Materials and Methods

## *Physical methods*

### *$^{10}$Be/$^{36}$Cl*

[1]Raimund Muscheler, [2]Ann-Marie Berggren and [2]Ala Aldahan, [3]Göran Possnert, [4]Christof Vockenhuber

[1]Dept. of Earth and Ecosystem Sciences, Division of Geology, Lund University, Sweden

[2]Dept. of Earth Sciences, Uppsala University, Sweden

[3]Department of Engineering Sciences, Uppsala University, Box 534, SE-751 21 Uppsala, Sweden

[4]Laboratory of Ion Beam Physics, ETH Zurich, HPK G31, Schafmattstrasse 20, 8093 Zurich, Switzerland

#### Introduction

$^{10}$Be and $^{36}$Cl are radioactive nuclides that are produced by the interaction of high-energy galactic cosmic rays with particles of the atmosphere [*Lal and Peters*, 1967; 1968]. Independent of variations in the production rates both radionuclides are produced at a constant ratio [*Masarik and Beer*, 1999] and they get deposited relatively rapidly within 1-2 years mainly by wet deposition [*Heikkilä et al.*, 2009a; *Heikkilä et al.*, 2009b]. Assuming that both radionuclides are deposited at a constant ratio, i.e. that they are similarly influenced by atmospheric circulation and deposition processes, an initial $^{10}$Be/$^{36}$Cl ratio can be estimated. The different half lives of $^{36}$Cl (301 000 years) and $^{10}$Be (approx. 1 390 000 years [*Chmeleff et al.*, 2010; *Korschinek et al.*, 2010]) then lead to a time-dependent change in the $^{10}$Be/$^{36}$Cl ratio ("$^{10}$Be/$^{36}$Cl half life" = 384 000 years). This has the advantage that, for example, varying snow accumulation rates or production rate changes (due to solar and geomagnetic field changes) have not to be considered since these are expected to influence $^{10}$Be and $^{36}$Cl similarly.

Nevertheless, the $^{10}$Be&$^{36}$Cl dating method is not without problems. It has been observed that $^{36}$Cl and $^{10}$Be deposition rates do not change similarly with variations in climate. For example, an approximately 20% change in the $^{10}$Be/$^{36}$Cl ratio has been observed in Central Greenland over the transition from the last ice age to the Holocene [*Wagner et al.*, 2000]. In addition, it has been shown for low accumulation areas that $^{36}$Cl can diffuse in snow and firn after deposition [*Delmas et al.*, 2004]. Especially the acidity of the ice could influence the mobility of HCl in the firn. Migration of $^{10}$Be in the ice might be another process that complicates the $^{10}$Be/$^{36}$Cl interpretation. For example, it has been shown that $^{10}$Be accumulates on dust particles within the ice matrix over long tens of millennia back in time [*Baumgartner et al.*, 1997]. In such a case $^{10}$Be might be lost for the measurement process especially when filtering the melted ice before sample preparation. In the GRIP ice core older ice had up to 50% of the $^{10}$Be attached to the dust [*Baumgartner et al.*, 1997]. In addition, $^{10}$Be adsorbed on the silt (e.g. inherited from the period before a region got glaciated) could add significant amounts of $^{10}$Be that would distort the $^{10}$Be/$^{36}$Cl ratio. In addition, the input of $^{10}$Be attached to dust ("recycled $^{10}$Be" that is not freshly produced in the atmosphere) can change the $^{10}$Be/$^{36}$Cl ratio. Such processes limit the accuracy of the $^{10}$Be/$^{36}$Cl dating method but the influence of the different processes can be estimated and included in the age estimates [*Willerslev et al.*, 2007a].

Due to lack of reliable estimates of the importance of the different complicating factors we have to make the assumption that only the radioactive decay leads to a depletion of the radionuclides with time. However, potential additional losses can be estimated from $^{10}$Be measurements on the melt water and on the filters that are used before preparing the $^{10}$Be samples. In addition, crucial for any radioactive dating method is accurate knowledge of the initial conditions. Again, the aforementioned complications can makes the estimate of the initial conditions a difficult task and limit the accuracy of the method. Based on the results we can, however, distinguish different scenarios that lead to more or less reliable age estimates. These scenarios are outlined in the following:

The optimal scenario:

The most reliable ages can be obtained when there is a clear $^{36}$Cl (decreasing) trend with depth (or age). This is, for example, visible in the GRIP ice core where the

$^{10}Be/^{36}Cl$ ratio (combined half-life: $T_{1/2}$=384.000 yrs) is mainly driven by a decrease in the $^{36}Cl$ concentrations ($^{10}Be$ stays relatively stable) [*Willerslev et al.*, 2007a].

The less convincing scenario:
The $^{10}Be/^{36}Cl$ ratio increases but there is not clear trend in the $^{36}Cl$ values but the $^{10}Be$ values still lie in the normal range of variability. This could imply that lower accumulation rate or higher production rates have offset the $^{36}Cl$ decrease. The $^{10}Be/^{36}Cl$ ratio should then still be a good age indicator.

The problematic scenario:
No clear trend in $^{36}Cl$ and a strong increase in $^{10}Be$. This leads to an increase in the $^{10}Be/^{36}Cl$ ratio but it is very likely that there were additional $^{10}Be$ sources ($^{10}Be$ brought in by dust, $^{10}Be$ from the bedrock) that should not be included in an age estimate based on the $^{10}Be/^{36}Cl$ ratio. In this case the $^{10}Be/^{36}Cl$ ratios would overestimate the age.

In this context it is an important consideration what $^{10}Be/^{36}Cl$ ratios can be expected in the ice. This number probably depends not only on the ratio that is determined by the production processes (about a factor of 10 according to [*Masarik and Beer*, 1999]) but it also depends on atmospheric circulation, depositional processes and potentially complicating factors (see above). The following table shows a few measurements that are interesting for this consideration:

| | $^{10}$Be [$10^4$ atoms/g ice] | $^{36}$Cl [$10^4$ atoms/g ice] | $^{10}$Be/$^{36}$Cl average | comment |
|---|---|---|---|---|
| GRIP | 2 | 0.4 | about 5 | Ice age/warm period differences in the order of 20% Oldest ice the ratio goes up to 20-25 |
| Dye3 (last 500 years) | 1 | 0.15 (without bomb peak) | 6-7 | Similar in [*Willerslev et al.*, 2007a] $^{10}$Be/$^{36}$Cl ratio in the silty ice is about 22 (driven by $^{36}$Cl decrease and $^{10}$Be increase) |
| Camp Century | | | 3-6 | this study |

At least in Greenland one could expect an initial $^{10}$Be/$^{36}$Cl ratio in the order of 5 which is below the ratios expected from the production rate calculations. Results from Antarctic ice cores are closer to this expected production ratio. For the Dome Fuji ice core the ratios are close to 10 for ice from the Holocene [*Horiuchi et al.*, 2008; *Sasa et al.*, 2010].

## Camp Century considerations

To distinguish between meteoric and "dust-borne" $^{10}$Be input we measured $^{10}$Be after filtering the meltwater with 0.45 mikrometer mesh size filters. Both the $^{10}$Be in the meltwater and the $^{10}$Be on the filter are measured subsequently. Since $^{10}$Be can migrate to the dust in the ice matrix this separation does not allow us to exclude any meteoric origin of the $^{10}$Be on the filters. However, it provides additional information on the likely origin of the $^{10}$Be in the sample. In the following "$^{10}$Be on the dust" refers to the $^{10}$Be component measured on the filter.

For the Holocene samples we obtained $^{10}$Be/$^{36}$Cl ratios that are in the order of 6 to 7 ($^{10}$Be/$^{36}$Cl = 5.8 & 6.8 ($^{10}$Be ice/$^{36}$Cl ice) and 6.4 & 7.3 (($^{10}$Be in ice & on the dust)/$^{36}$Cl)). This agrees well with the estimates obtained from the Dye and GRIP. Therefore, a value of around 6-7 appears to be a good estimate for expected $^{10}$Be/$^{36}$Cl ratios during warm periods.

However, similar to the GRIP ice core we observe a significant dependency of the $^{10}$Be/$^{36}$Cl ratios on $\delta^{18}$O. Figure 1 indicates that this ratio could change by up to a factor of 2 depending on climate.

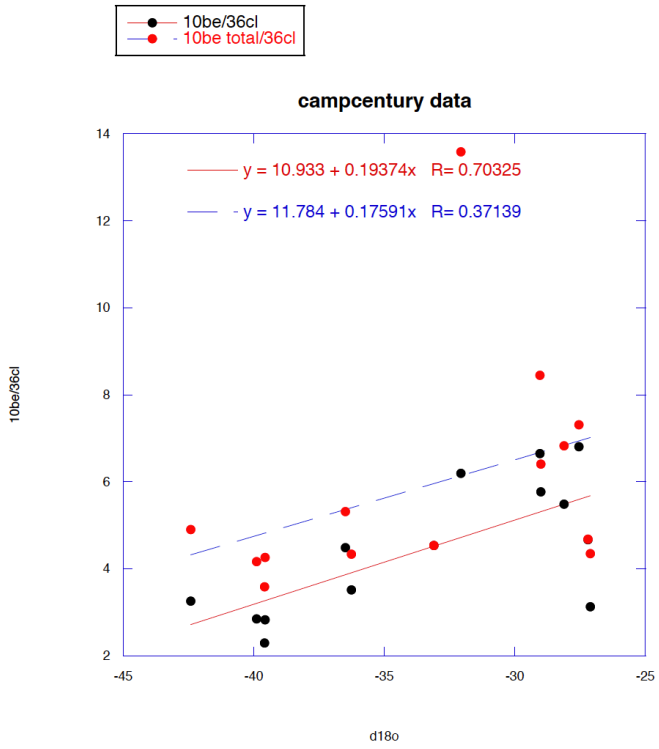



Figure 1: $^{10}$Be/$^{36}$Cl ratio versus $\delta^{18}$O . The plot shows all values obtained in this study (not corrected for decay). The results suggest that ice age ($\delta^{18}$O =-40‰) $^{10}$Be/$^{36}$Cl values are in the range between 3 (only $^{10}$Be in the meltwater) and 4 (all

$^{10}$Be included) which is significantly lower than in the ice from the Holocene. There is an age scale for Camp Century ice for the last 100,000 years. To investigate the origin of the climate dependency in more detail we focus on the samples where we have age estimates (last about 100,000 years) which enables us to correct for the known decay of $^{10}$Be and $^{36}$Cl. Figure 2 shows the $^{10}$Be on the filter afte correction for decay and in dependence on $\delta^{18}$O.

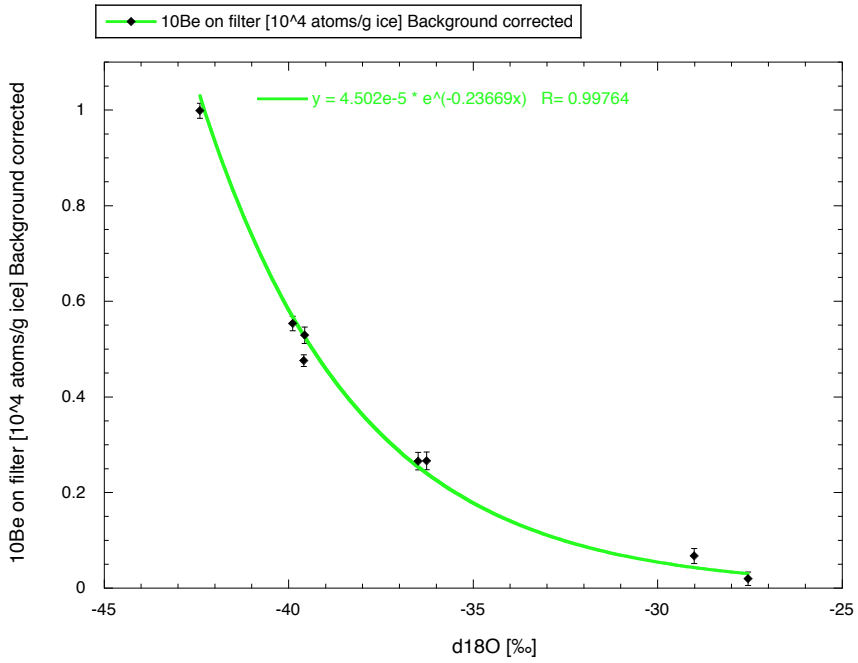



Figure 2: Decay-corrected $^{10}$Be measured on the filter.

Figure 2 shows a strong correlation between $\delta^{18}$O and the $^{10}$Be on the dust. There are several processes that could lead to this effect:

- More dust during glacial periods and $^{10}$Be (atmospheric) attaches to the dust (and gets lost for the $^{10}$Be measurements on the melt water). In this case the total $^{10}$Be ($^{10}$Be in the ice and $^{10}$Be on the filter) would be the best estimate for atmospheric $^{10}$Be input.
- More dust input (and recycled $^{10}$Be) during glacial periods. This is more likely also considering that Camp Century is closer to the ice margin. In this case only the $^{10}$Be in the ice should be used for the analysis.
- For GRIP $^{10}$Be it was observed that $^{10}$Be on the dust is time dependent (more $^{10}$Be attached to the dust the older the ice). This possible $^{10}$Be migration in the

ice would lead to lower $^{10}$Be concentrations in the melt water and to an underestimation of the $^{10}$Be/$^{36}$Cl ratios with age (and therefore to lower age estimates). This age/depth dependent effect cannot be seen in the CampCentury samples since these are dominated by climate effect (see figure 1 and 2).

Therefore, it is unclear if we have to include the $^{10}$Be on the dust in the age estimates and we will consider two cases that are based on different assumptions on the origin of the $^{10}$Be on the dust.

**Case I:**

**Assumptions:**

**I: $^{10}$Be on the dust should not be included in the $^{10}$Be/$^{36}$Cl ratios because of input of recycled $^{10}$Be**

**II: The $^{10}$Be/$^{36}$Cl ratio depend on climate which has to be corrected for before making any age estimates**

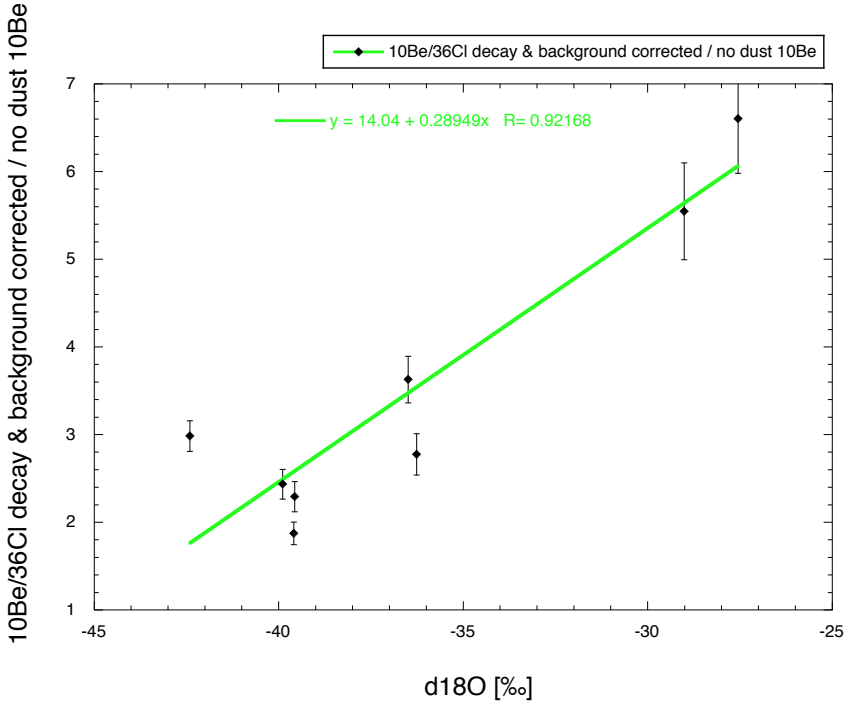



Figure 3: Decay-correct $^{10}$Be/$^{36}$Cl ratios (only $^{10}$Be in the melt water, last 100000 years).

Figure 3 shows the decay-correct $^{10}Be/^{36}Cl$ ratios. There is a strong dependency of the ratio on climate ($\delta^{18}O$) which can be fitted with a linear function. For the following this dependency will be removed from all samples. This will provide an estimate for a "climate corrected" melt water $^{10}Be/^{36}Cl$ ratios that then will be used for estimating the age of the lowermost ice. However, there is a large uncertainty in this correction since it is not clear if this suggested correction is indeed linear and if it applies to the older ice, too.

Figure 4 shows the climate corrected $^{10}Be/^{36}Cl$ ratios and $\delta^{18}O$ versus depth and tentative age. This age extension (black line, extending the known age scale for the last 100,000 years) suggests that the oldest sample could have a minimum age of around 150,000 years. The scatter in the data illustrates the uncertainty of this estimate. This estimate (150,000 years) would be rather a lower age estimate since any $^{10}Be$ on the dust is not included.

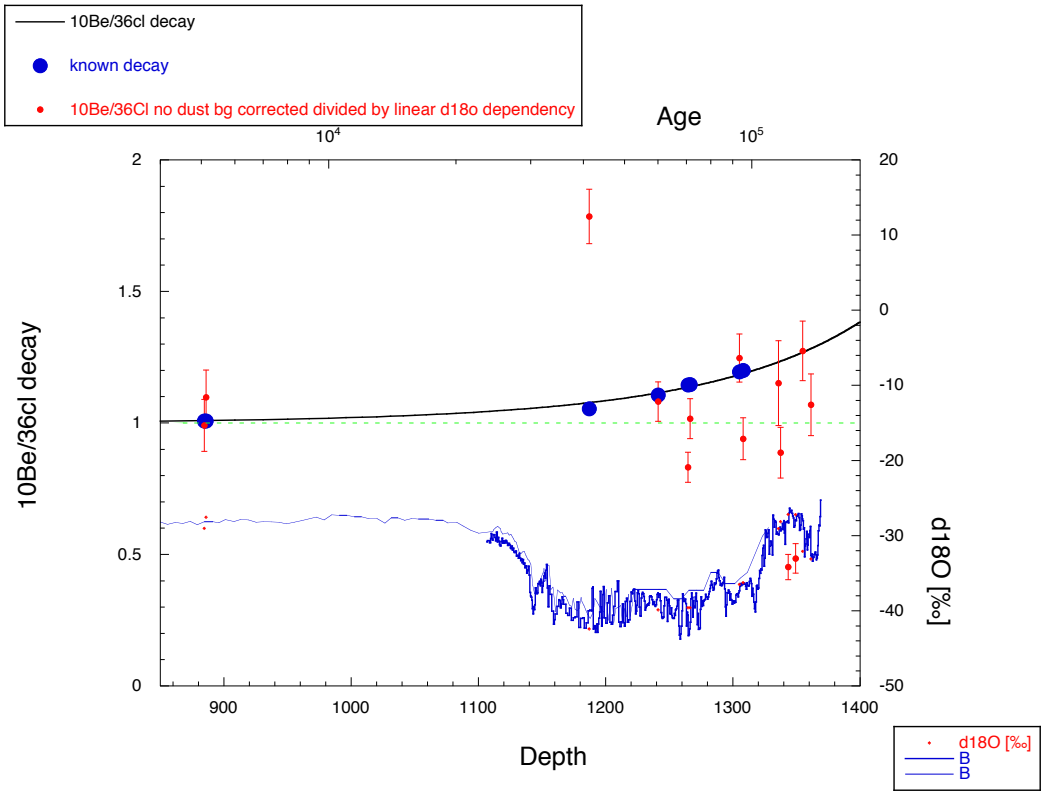


Figure 4: “Climate corrected” $^{10}Be/^{36}Cl$ ratio (only meltwater $^{10}Be$) and ($\delta^{18}O$). The blue dots shows the normalized $^{10}Be/^{36}Cl$ ratios that are expected for the samples with known age. The black curve shows an extension of this decay curve.

**Case II:**

**Assumptions:**

**I: $^{10}Be$ on the dust should be included in the $^{10}Be/^{36}Cl$ ratios because of $^{10}Be$ attaching to the dust after deposition**

**II: The $^{10}Be/^{36}Cl$ ratio depends on climate which has to be corrected for before making any age estimates**

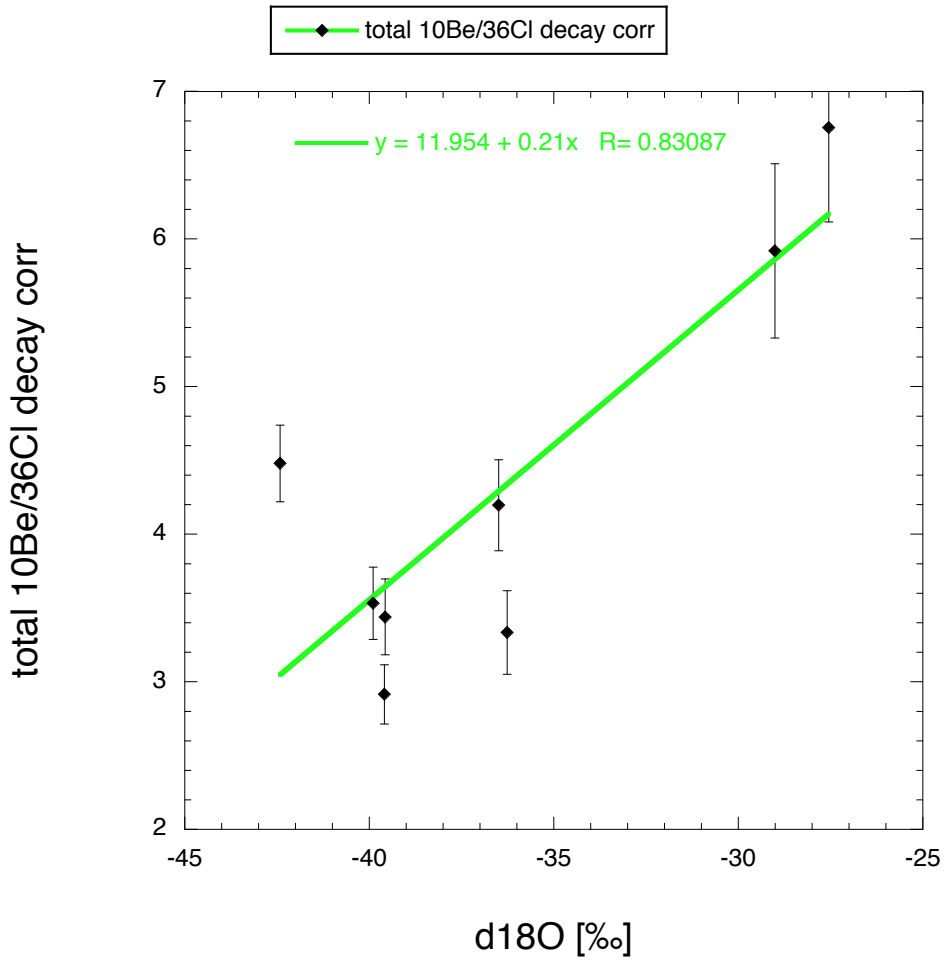



Figure 5: Total $^{10}Be/^{36}Cl$ after correction for known decay versus d18O.

Figure 5 shows the $^{10}Be/^{36}Cl$ ratios (for the samples with known age) after decay correction versus d18O. Again, there is a clear climate influence on the ratios. In the following we will correct for this with a linear function (green curve).
This correction will be applied to all samples (also unknown age) and the resulting climate corrected $^{10}Be$ (total) / $^{36}Cl$ ratios will be used for an age estimate.

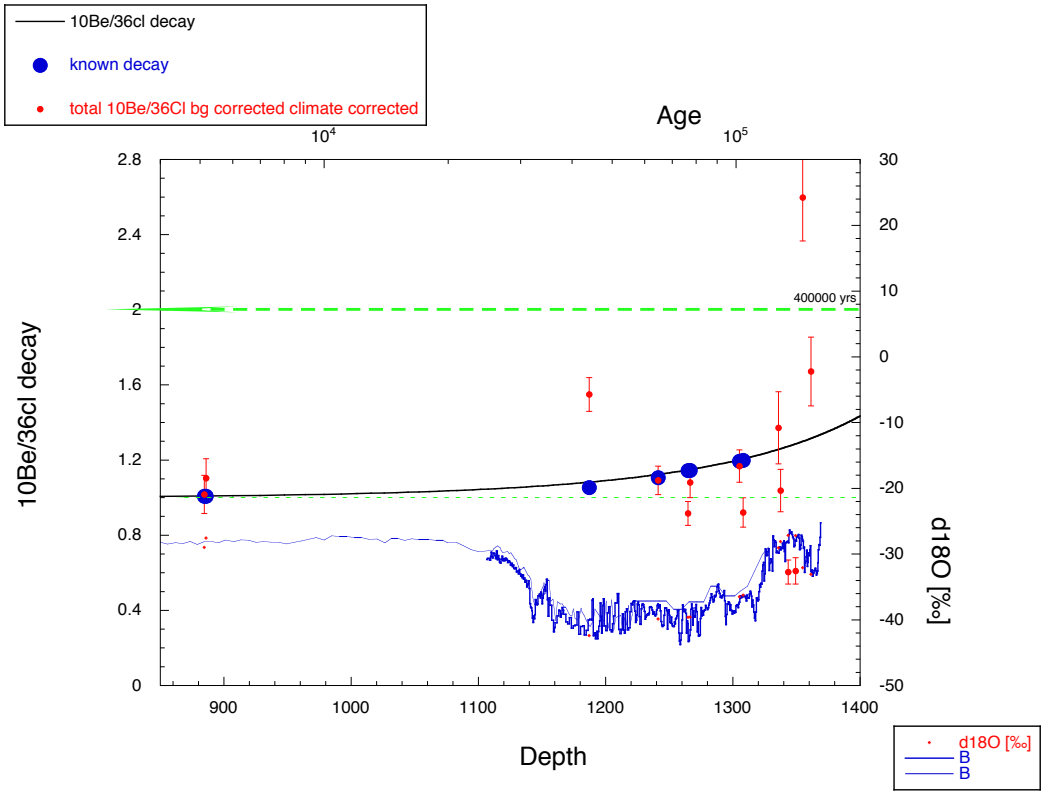



Figure 6: Total $^{10}Be/^{36}Cl$ versus depth and tentative age.

Figure 6 shows the $^{10}Be/^{36}Cl$ ratio after correction for the obvious climate influence as seen in figure 5. The black curve shows the expontentiell decay curve. It suggests that the oldest samples increase strongly in age. A doubling of the ratio corresponds to about 400000 years which lies in the same order of magnitude as the 2 lowermost samples. However there is one important complication: The high $^{10}Be/^{36}Cl$ values for the lowermost samples are due to the high $^{10}Be$ content on the dust/silt which might not be connected to a direct atmospheric $^{10}Be$ input.
To conclude: Age estimates for the lowermost ice at Camp Century are uncertain mainly due to the lack of a clear $^{36}Cl$ decrease that can be attributed to the decay. Additional studies about potential $^{10}Be$ and $^{36}Cl$ inputs from the bedrock or a better

understanding of the climate influence on the $^{10}$Be&$^{36}$Cl deposition are necessary to improve the age estimates based on the $^{10}$Be/$^{36}$Cl ratios.

## Dry valley, Antarctica considerations

Similar to the Camp Century considerations it is important for any age estimate to know the initial $^{10}$Be/$^{36}$Cl ratio as a reference value to infer an age. Initial values can lie in the range of 5 to 10 (see above) depending on location. By measuring $^{10}$Be and $^{36}$Cl from Taylor and Suess glacier from different locations we tried to estimate a realistic reference value to estimate the age of the oldest ice (that contained the DNA) from the Suess glacier.

## Data: (see attached Excel file) and table below

| New Code | Old Code (as on plastic bags) | Depth of sample (cm) | $\delta^{18}O$ | $\delta D$ | Weight (gr) | | sample name | weight (g) | carrier (mg) | $^{36}Cl/Cl$ (1 (%) | error | $^{36}Cl$ conc. ($^{36}Cl/10$ | | sample | sample(g) | carrier(g) | [Be10 atom | error | 10Be/36Cl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TGS I 1 | TGS I 1 | 123 | -44.18 | -353.8 | 426 | | | | | | | | | | | | | | |
| TGS I 2 | TGS I 2 | 148 | -43.01 | -344.6 | 511 | | | | | | | | | | | | | | |
| TGS I 3 | TGS I 3 | 174 | -44.68 | -358.6 | 610 | | | | | | | | | | | | | | |
| TGS II 1 | TGS II 1 | 143 | -42.72 | -339.6 | 386 | | | | | | | | | | | | | | |
| TGS II 2 | TGS II 2 | 190 | -43.54 | -347.9 | 343 | | | | | | | | | | | | | | |
| TGS II 3 | TGS II 3 | 253 | -43.06 | -341.3 | 411 | | | | | | | | | | | | | | |
| TGS III 1 | TGS III 1 | 141 | -42.25 | -338.9 | 475 | | | | | | | | | | | | | | |
| TGS III 2 | TGS III 2 | 196 | -42.55 | -339.8 | 352 | | | | | | | | | | | | | | |
| TGS III 3 | TGS III 3 | 246 | -41.28 | -329.1 | 395 | | | | | | | | | | | | | | |
| TGS IV 1 | TGS IV 1 | 141 | -41.96 | -333.6 | 377 | | | | | | | | | | | | | | |
| TGS IV 2 | TGS IV 2 | 206 | -41.48 | -328.9 | 356 | | | | | | | | | | | | | | |
| TGS IV 3 | TGS IV 3 | 228 | -41.41 | -329.2 | 332 | | | | | | | | | | | | | | |
| TGS V 1 | TGS H 1 | 141 | -40.25 | -321.4 | 352 | | TGS V | 982.0 | 3 | 0.056 | 6.4 | **0.289** | **±0.019** | TGS V | 982.0 | 6.00E-04 | 6.04 | 3.5 | 20.91518 |
| TGS V 2 | TGS H 2 | 182 | -41.81 | -333.3 | 292 | | TGS V | 982.0 | 3 | 0.056 | 6.4 | **0.289** | **±0.019** | TGS V | 982.0 | 6.00E-04 | 6.04 | 3.5 | 20.91518 |
| TGS V 3 | TGS H 3 | 249 | -39.81 | -315.1 | 370 | | TGS V | 982.0 | 3 | 0.056 | 6.4 | **0.289** | **±0.019** | TGS V | 982.0 | 6.00E-04 | 6.04 | 3.5 | 20.91518 |
| TGS VI 1 | TGS HH 1 | 140 | -45.08 | -361.8 | 417 | | TGS VI | 1157.9 | 3 | 0.126 | 5.2 | **0.557** | **±0.029** | TGS VI | 1157.9 | 6.00E-04 | 4.47 | 3.9 | 8.031772 |
| TGS VI 2 | TGS HH 2 | 201 | na | -362.9 | 364 | | TGS VI | 1157.9 | 3 | 0.126 | 5.2 | **0.557** | **±0.029** | TGS VI | 1157.9 | 6.00E-04 | 4.47 | 3.9 | 8.031772 |
| TGS VI 3 | TGS HH 3 | 245 | na | -363.6 | 418 | | TGS VI | 1157.9 | 3 | 0.126 | 5.2 | **0.557** | **±0.029** | TGS VI | 1157.9 | 6.00E-04 | 4.47 | 3.9 | 8.031772 |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | STblank1 | 800.9 | 3 | 0.002 | 40.8 | **0.011** | **±0.005** | STblank1 | 800.9 | 6.00E-04 | 0.13 | 23.2 | |
| | | | | | | | STblank2 | 802.1 | 3 | 0.001 | 57.8 | **0.004** | **±0.002** | STbl2 | 802.1 | 6.00E-04 | 0.32 | 11.9 | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | Suess98/2 | 932.7 | 3 | 0.054 | 6.2 | **0.296** | **±0.018** | Suess98/2 | 932.7 | 6.00E-04 | 3.47 | 5.7 | 11.73428 |
| | | | | | | | TGTB9 | 645.0 | 3 | 0.046 | 6.9 | **0.362** | **±0.025** | TGT B9 | 645 | 6.00E-04 | 3.15 | 6.6 | 8.690211 |

Table 1. Data from Taylor and Suess glacier that have been used for the following analysis.

## Summary of the results from Taylor and Suess glacier:

**$^{36}$Cl/$^{10}$Be = 21 for TGS V**

**$^{36}$Cl/$^{10}$Be = 8 for TGS VI**

**$^{36}$Cl/$^{10}$Be = 11.7 for Suess98/2**

**$^{36}$Cl/$^{10}$Be = 8.7 for TGT B9**

Suess98/2 (0.3), TGTB9 (0.36) and TGS V (0.29) have similar $^{36}$Cl concentrations. TGS VI is significantly higher (0.56).

### Interpretation:

From the ratios it is likely that TGS V represents the oldest ice. If the $^{36}$Cl/$^{10}$Be ratios are only influenced by age (minor influence of climate, $^{10}$Be dust input, $^{36}$Cl

diffusion) one can conclude that TGS V is about 500,000 yrs older than the other samples.

If we can take a $^{36}Cl/^{10}Be$ ratio of 6 as the modern reference value one could infer a minimum age of about 100000 years for TGS VI and TGT B). Suess 98/2 would be 200000 yrs old. However, as outline above this initial $^{36}Cl/^{10}Be$ ratio depends on the location and might be closer to 10 (the expected production ratio) at Antarctic sites. This would imply that TGS VI and TGT B are very close to initial $^{36}Cl/^{10}Be$ ratios and no significant old age can be inferred for these samples. Again, more research about additional $^{10}Be$ and $^{36}Cl$ sources (e.g. from exposed bedrock) are required to improve these estimates.

# *OSL*

Christine Thiel, Jan Pieter Buylaerta and Andrew Murray
The Luminescence Research Laboratory, Risø National Laboratory. Denmark

## Introduction

In optically stimulated luminescence dating, quartz and feldspathic grains incorporated in the ice are used for dating. The grains are exposed to ionizing radiation from the decay of naturally occurring radio-nuclides (mainly $^{40}K$, $^{238}U$ and $^{232}Th$; [*Aitken*, 1985]) from within the grain (internal dose). The ionizing radiation from other grains of any size (external dose) will contribute to the total dose rate to individual grains if the grains are very close to each other. The ionizing radiation causes free charge (i.e. electrons) to be trapped at defects in the crystal lattice of the minerals. When exposed to daylight, the electrons absorb energy, causing de-trapping from their meta-stable traps and recombination in the crystal; this results in release of energy, i.e. luminescence. This process occurs under natural conditions during transport and associated exposure to sunlight of the grains and is commonly known as bleaching or zeroing. At time of deposition, the luminescence clock is thus reset, i.e. any trapped charge population is removed or at least reduced to a negligible value. After burial, or in case of ice, after ice formation, the trapped charge population builds up again. In the laboratory, optical stimulation of the samples frees the electrons and causes them to recombine at luminescence centers; the emitting luminescence can be measured. The amount of luminescence is related to the total radiation dose absorbed by a sample since its last exposure to daylight. Hence, both the equivalent dose ($D_e$) and the dose rate, i.e. the rate of energy absorption from ionizing radiation during burial, have to be determined in order to derive an age.

Sub-sampling for luminescence dating

For this study, five sub-samples (1049A, 1053B, 1054A, 1055B, and 1057A) from the Camp Century ice core were investigated. All sub-samples originate from the inner core of the ice and have thus seen as little (room-) light as possible. Samples were melted and centrifuged at 14.000 g for 30 minutes under subdued orange light to

separate sediment grain particles from melted ice in ancient DNA clean lab facilities (Centre for GeoGenetics, University of Copenhagen). The grain particles were concentrated in the bottom of 50mL Nalgene® tubes and the melted ice separated. Then each tube containing the sediment samples was wrapped in foil to avoid any exposure to daylight during transport.

The following sample preparation took place under subdued orange light at the Nordic Laboratory for Luminescence Dating, Risø, Denmark.

For each sample, about 200-400 mg of sand-sized grains was available. To limit the loss of valuable grains, the sample preparation was limited to wet-sieving (90-125 µm, 125-180 µm and 180-250 µm) and HF-etching (10%, 15 minutes) of the fractions. The etching was followed by HCl treatment (10%, 15 minutes) and repeated rinsing with de-ionized water. Density or magnetic separation of the minerals was not considered practical, and thus the fractions are composed of a variety of minerals, which had to be distinguished based on luminescence characteristics.

Equivalent dose determination

Commonly, the $D_e$ is determined on aliquots composed of several tens to hundreds of sand-sized grains. However, when only a small amount of material is available, individual grains are used for dating (e.g. Duller, 2008). Special single grain sample holders (discs) are available for such measurements (Bøtter-Jensen et al., 2003). They are composed of arrays of 100 holes (300 µm deep, 300 µm in diameter) in a 10x10 grid. The position of the holes is recognized on the basis of three positioning holes, and thus an unequivocal attribution of grains and luminescence signal is given. The individual grains (n=4600 grains) were mounted on the single grain discs under subdued orange light.

The luminescence measurements were made on Risø TL/OSL-DA-15 (Bøtter-Jensen et al., 2003) and Risø TL/OSL-DA-20 (Thomsen et al., 2006) automated readers, equipped with dual laser systems (i.e. green laser and infra-red laser) for single-grain measurements. The grains were irradiated using $^{90}Sr/^{90}Y$ beta sources, calibrated with γ-dosed (4.81 Gy) calibration quartz provided by the Nordic Laboratory for Luminescence Dating.

Pretests on small aliquots (about 20 grains), for which mineral separates were available, showed that quartz extracted from the Camp Century ice core is very insensitive, i.e. it is unsuitable for dating. Willerslev et al. (2007) observed the same phenomenon for their quartz from the Dye3 ice core. Therefore, it was made use of infra-red (IR) stimulation, to which quartz is not sensitive; hence, only feldspathic grains were stimulated and correspondingly measured.

To avoid anomalous fading, i.e. an athermal signal loss with time commonly observed in feldspars (Wintle, 1973; Spooner, 1994), which leads to significant age underestimation, we made use of a single-aliquot regenerative (SAR) elevated temperature post-IR IRSL protocol (Buylaert et al., 2009) to the single grains. After a preheat of 250°C (60 s), the aliquots were stimulated for 100 s at 50°C with an array of IR light emitting diodes (LED). This was followed by IR stimulation (3 s) of the individual grains with a 150 mW 830 nm IR laser (20 μm spot diametre) at elevated temperatures (225°C). The luminescence was measured through a Schott BG39/Corning 7-59 filter combination in the blue-violet region (325-450 nm). The test dose response was measured in the same manner, and at the end of each SAR cycle an IR illumination (IR diodes) at 290°C for 40 s was conducted on order to avoid signal carry-over from one measurement cycle to the next. For each grain a full growth curve was generated, based on the initial 0.2 s of the post-IR IRSL signal less a background of the last 1.0 s. All grains with recycling ratio >15%, test dose error and/or equivalent dose error >20% and recuperation >10% were rejected. The rejection criteria were tested on single grains of a known age sample (FK952509; Nordic Laboratory for Luminescence Dating), for which the application of these criteria did not result in significantly different equivalent dose estimates compared to the slightly stricter criteria suggested by Wintle and Murray (2006).

The suitability of the measurement protocol to our samples was tested by means of a dose recovery test using grains from the outer core of sample 1054 A. The grains were bleached in a Hönle SOL2 simulator for 2 hours; this treatment ought to represent ‘natural’ bleaching conditions and should remove all trapped charge. The sample was then given a beta dose of ~40 Gy, and the given dose measured in the usual manner. The measured to given dose ratio resulted in 1.09 ± 0.08 (n=42), which is in 10% of unity and thus shows the applicability of our measurement protocol.

In total, 906 feldspathic grains fulfilled the above mentioned criteria and were analyzed for their chemical composition/mineralogy as part of the dose rate determination.

## Dose rate determination

The dose rate to the individual grains is more or less entirely made up of the internal dose, which is dependent on the $^{40}K$, $^{238}U$ and $^{232}Th$, because neither the external dose nor the cosmic dose rate is of any importance to the grains disperse in ice (Prescott et al., 1988; Willerslev et al., 2007). The internal dose rate to potassium-rich feldspar grains (K-feldspar) is dominated by $^{40}K$ (Zhao and Li, 2005), whereas for sodium- and calcium-rich feldspar grains great dependency on $^{238}U$ and $^{232}Th$ is given (Willerslev et al., 2007). Especially in the latter case, problems arise from the heterogeneity of the individual elements within the grains; this hampers the dose rate determination for sodium- and calcium-rich feldspars. Due to the smaller dependency on other elements and their heterogeneities, K-feldspar grains seem the best choice for deriving accurate age estimates. Furthermore, the luminescence characteristics and the fading behavior are best understood for K-feldspar, and thus only these grains were used for age determination.

To distinguish potassium-rich from other grains, scanning electron microscopy (SEM/EDX) was applied to all grains for which equivalent doses were determined. The grains were kept in their original grid to ensure an unequivocal attribution of equivalent dose and chemical composition. Both the images and the semi-quantitative concentration analyses were used for mineral identification. 209 grains were identified as K-feldspar, and an average K concentration of 12.5 ± 0.5% (Huntley and Baril, 1997) was used for dose rate calculations. The calculations further included average values for $^{238}U$ (0.08 ± 0.01 ppm) and $^{232}Th$ (0.56 ± 0.20 ppm) (Zhao and Li, 2005) as well as an a-value of 0.08 ± 0.02. Following this the average dose rate to the grains is 0.69 ± 0.14 Gy/ka.

## Results

The number of suitable grains varies from sample to sample (see below table). The smallest number of grains comes from sample 1049 (n=28), and samples 1053 and 1054 had the largest number (n=57).

The dose distributions for all samples are skewed, with a long tail towards higher doses > 1000 Gy (sample 1049). The origin of this skewness is difficult to assess; one cause could be incomplete bleaching of individual grains prior to deposition. However, it is evident that the majority of measured doses falls within the lower dose region, resulting in mean doses ranging from 63 ± 10 Gy (sample 1057) and 270 ± 40 Gy (sample 1053) (see below table); the median ranges from 42 Gy (sample 1057) to 200 Gy (sample 1055). For all samples except sample 1057, the mean ages are older than Eemian (~120 ka), and the median ages range from 60 ka (sample 1057) to 300 ka (sample 1055).

The unweighted averages of both the mean (260 ± 80 ka) and median (180 ± 50 ka) ages imply incorporation of the grains between OIS 9 and OIS 7.

It has to be noted that sample 1057 is an outlier with significantly younger mean ages than the other samples. Possible causes for the disagreement with the other samples are i) more light-exposure during sampling and subsequent handling or ii) underestimation the uncertainties for all samples, i.e. sample 1057 is actually - within errors - in agreement with the other samples. In case of i), sample 1057 should be dismissed from the calculations, and thus the unweighted average of the mean would increase to 300 ± 50 ka, and the unweighted average of the mean would result in 200 ± 40 ka.

| sample | n | De [Gy] Mean | De [Gy] Median | Age (ka) Mean | Age (ka) Median |
|---|---|---|---|---|---|
| 1049 | 28 | 200 ± 40 | 130 | 300 ± 80 | 190 |
| 1053 | 57 | 270 ± 40 | 160 | 400 ± 100 | 230 |
| 1054 | 57 | 110 ± 20 | 60 | 160 ± 40 | 90 |
| 1055 | 36 | 240 ± 30 | 200 | 350 ± 80 | 300 |
| 1057 | 31 | 63 ± 10 | 42 | 90 ± 20 | 60 |
| unweighted average | 5 | | | 260 ± 80 | 180 ± 50 |
| unweighted average without 1057 | 4 | | | 300 ± 50 | 200 ± 40 |

None of the presented ages include corrections for anomalous fading or for possible signal loss due to light exposure during sampling and subsequent handling. The latter was estimated by comparing the IRSL signal at 50°C with the post-IR IRSL signal measured at 225°C. The latter is known to bleach slower than the IRSL signal at 50°C (Buylaert et al., 2009), and thus a comparison of the two signals from a known age sample (FK952509; Nordic Laboratory for Luminescence Dating) after different light exposure times allowed for estimating the maximum signal loss to 20% (Thiel et al., in prep.).

A similar amount of signal loss can be expected due to anomalous fading (Thiel et al., in prep). Fading experiments were run on single grain discs used for $D_e$ determination. Prior to a preheat of 250°C (60 s) the aliquots were given doses of ~40 Gy, after which different storages times (varying from as brief as experimentally possible ('prompt') to delays of up to ~10 hours) were inserted. IR stimulation at 50°C (100 s) was followed by post-IR IR stimulation at 225°C (100 s) for both the given dose measurement ($L_x$) and the test dose measurement ($T_x$). At the end of each SAR cycle, an IR illumination was conducted at 290°C for 40 s. The obtained *g*-values (percentage decrease of signal intensity per decade of time; Aitken, 1985) for the post-IR IRSL signal range from 1.5 ± 0.3 %/decade (sample 1055; n=3) to 3.0 ± 0.6 %/decade (sample 1057; n=3).

Another uncertainty is associated with the estimated potassium content of 12.5 **±** 0.5 %. A lower concentration would result in a lower dose rate, and thus in an older age. Since a potassium concentration ≤ 9% is not possible in a K-feldspar grain, the

minimum dose rate to the grains can be constrained to 0.51 ± 0.09 Gy/ka, which is about 30% less than the dose rate used for age calculation.
The presented ages might therefore be underestimated by a maximum of 40%; a precise estimation is not possible because of too many unknowns. However, this maximum value allows for an upper age constrain of 700 ka (based on the oldest derived age and its errors; sample 1053) (Thiel et al., in prep.).

# $^{238}U/^{234}U$

Sarah Aciego,
Department of Earth and Environmental Sciences, University of Michigan, USA

## Introduction

Recoil dating has been successfully applied to Dome C ice core samples; recoil ages for non-basal ice samples fall directly on the published age scale. Samples of basal ice from Dome C with large inclusions of fine-grained solid material had significantly younger ages, indicating a re-setting of the ice age and that the ice is no longer ”primary”. Successful dating of basal ice at other locations will depend on the characteristics of that particular location (melting, recrystallization, mixing). We measured recoil ages of basal ice samples from Camp Century and Dye3, Greenland and the Suess and Taylor Glaciers in Antarctica with mixed results. Ice recoil ages are not the age of the ice, but the time the sediment, dust for non-basal samples, has been incorporated within the ice. For basal ice samples, the recoil ages are a combination of the ice age with the incorporation age of basal material and are thus minimum ages for ice.

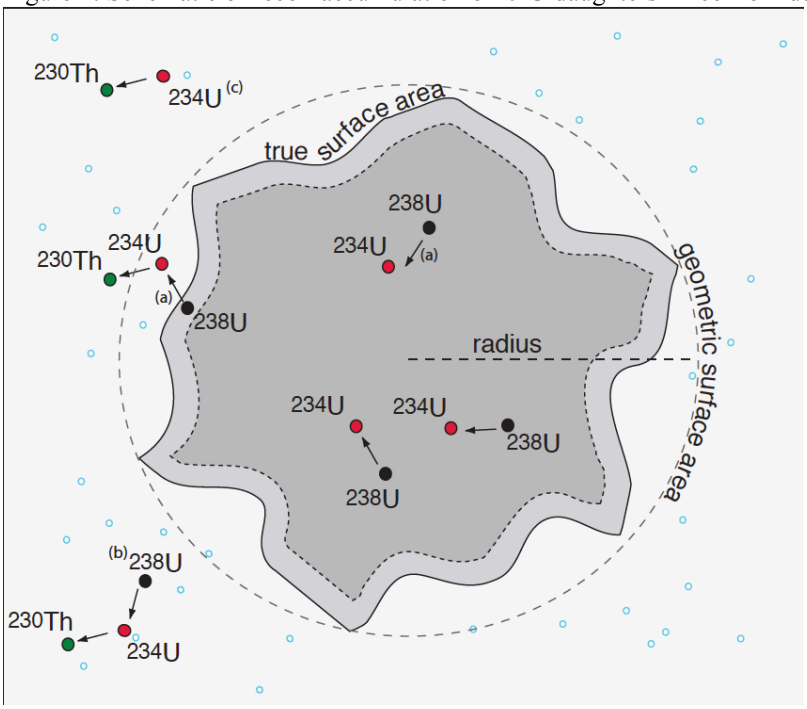
2. Basis of the Recoil Method
Recoil accumulation as a radiometric dating method for ice was first proposed by Fireman (1986) and more recently revisited by Goldstein et al. (2004), but a combination of analytical limitations, disturbed ice, and incomplete characterization

of the materials hindered their ability to produce accurate, precise ages. The principles of the dating scheme are illustrated in Figure 1, which shows a schematic of the particulate-ice, parent-daughter system for ice. The dust or sediment grains bound in the ice contain trace amounts of $^{238}U$ that alpha decays to $^{234}Th$ with a half-life of 4.4683 Ga, the energy of which results in the daughter nuclide moving a measurable distance, 20-25 nm (Hashimoto et al., 1985; Fleischer, 2003) in crustal minerals, with a subsequent rapid gamma decay to $^{234}U$ [Fig. 1, arrows a]. Ice also contains trace amounts of uranium, both $^{238}U$ and $^{234}U$ [Fig. 1, (b), (c)]. The concentration of $^{234}U$ in the ice depends on the initial concentration of
uranium in the ice, the concentration of $^{238}U$ in the dust or sediment, the fraction of daughters, f, that recoil out of the grain into the ice (e.g., in Fig. 1, f = 0.125), and the decay of $^{234}U$ to $^{230}Th$. This accumulation-decay defines the timescale over which the radiometric method is applicable, and ages can be calculated using the age equation:

$$t = -\left(\frac{1}{\lambda_{234}}\right) log\left(\frac{[U_{234}]_{ice} - f\,[U_{238}]_{dust} - [U_{238}]_{ice}}{[U_{234}]^{in}_{ice} - f\,[U_{238}]_{dust} - [U_{238}]_{ice}}\right) \quad (1)$$

where square brackets denote the activity of the respective nuclides.

Figure 1: Schematic of recoil accumulation of $^{234}U$ daughters in ice from dust or sediment.




In basal ice, the two sources of material are depositional, marine (dis

solved) and continental (particulate) aerosols, and basal sediment incorporation both as dissolved and particulate material. The large amounts of basal material incorporation results in measurably high concentrations of uranium and uranium isotopic compositions. Constraining the concentration of initial uranium and its isotopic composition from the combination of marine aerosols and basal dissolved species is much more difficult. If the dissolved uranium component has the same marine source as ssNa and Cl, the initialuranium isotopic composition should be constant; the $^{234}U/^{238}U$ activity of uranium in seawater for the past 800 ka has remained 1.14 ± 0.01 (Henderson, 2002). Direct measurements of young ice samples from the Dye3 location in Greenland confirm that the bulk of the dissolved uranium in the non-basal ice is from a marine component. The addition of a dissolved fraction from basal material will change this starting value considerably: basal sediments should have $^{234}U/^{238}U$ activities close to equilibrium, 1.00, thus dissolution, or fining of sediments to less than 0.05μm, would result in an initial $^{234}U/^{238}U$ activity between 1.00 and 1.14.

The fraction of daughter products accumulating in the ice from recoil depends on the surface to volume ratio of the sediment. Because the dust grains are of varying mineralogy, and therefore shape (Gaudichet et al., 1988; Delmonte et al., 2004), this fraction cannot be calculated precisely from a geometric model based on radii and crystallography. However, an approach based on measurements of the surface area can account for the varying surface roughness (Bourdon et al., 2009; Semkow, 1991). The recoil factor (f) is calculated from:

$$f = \left(\frac{1}{4}\right)\left(\frac{2^{D-1}}{4-D}\left[\frac{a}{R}\right]^{D-2}\right) R \cdot S_{BET} \cdot \rho_{dust} \qquad (2)$$

where D is the fractal dimension of the surface, a is the diameter of the adsorbate molecule ($N_2$, 0.314 nm), R is the recoil length (24 nm, calculated using the using the modal mineralogy (see supplementary information) and recoil lengths calculated for different minerals using the nuclear shell model (Hashimoto et al., 1985; Fleischer, 2003) and variations are less than 1% 2 S.D.), and $\rho_{dust}$ is the density of the sediment (2.64 g/cm$^3$; average continental crustal material).

Surface area calculations using the new nano-scale BET method make it possible to measure surface areas with 0.2-5 mg of material by measuring co-

variations of pressure with $N_2$ gas adsorption.

## Methods and Material

Camp Century (CC) ice core samples 1048A, 1052C and 1054A from approximately 1351-1360m depths were selected for this study (Sample table 1). All showed distinct banding with layers of high basal sediment incorporation and layers almost perfectly clear. Basal sediment was markedly brown-orange. CC samples ranged in size from 440 g to 640 g. The Dye3 basal ice, 1970m depth and 810 g, also showed visible inclusions, with less striking stripes or bands indicating variability in sediment incorporation. The Taylor Glacier (TG) and Suess Glacier (SG) basal ice samples were both approximately 1000 g. The TG sample had visible, large inclusions with several larger than 1 mm in diameter but little material in the < 30 mm size fraction. The SG sample had abundant sediment (more than 100 mg), but it was dominated by the < 30 mm size fraction.

The methods described here are identical to those described in Aciego et al., 2011. Tools for decontaminating and processing the core were machined out of PFA teflon and acid washed prior to use; chisels were constructed out of PFA teflon impregnated with quartz in order to achieve the hardness necessary to scrape the ice. All initial processing and subsequent chemistry was done in a class 100 clean room, under class 10 000 laminar flow hoods. Samples were decontaminated in a three-step process. First, while the core sections were still cold (-20$^{o}$C), the outer 5 mm was scraped away with PFA chisels, then rinsed with distilled ethanol. After allowing the cores to warm for 10 minutes, they were scraped with new, clean, chisels to remove another 1-2 mm then rinsed with 18.4 M$\Omega$ H$_2$O. This second scraping was repeated, then the ice placed in 2 L teflon beakers. The samples were then weighed and allowed to melt at room temperature. Seastar ammonia was added dropwise every 0.5 hours to maintain a pH of 7.5 while the samples melted.

Samples were filtered immediately after melting in 100 mL aliquots with a 2-stage PTFE filtration unit equipped with PFA filters to separate out the sediment in the > 30 mm and < 30 mm > 0.05 mm size fractions, and to collect the ice filtrate. EDTA, 1 mL 0.05 M, was added to the final 100 mL aliquot and 100 mL $H_2O$ rinse to remove

any adsorbed Th from the dust grains. All of the CC and SG filtrates were colored indicating a high dissolved (< 0.05 mm) load, while the Dye3 and TG filtrates were clear. A spike of $^{236}U$ was added to the filtrate prior to drying under a clean $N_2$ atmosphere. The residue was treated with concentrated $HNO^3$ and HF in 15 mL Teflon beakers to promote dissolution of the EDTA salts and equilibration with the spike, usually 2-3 days, then dried down. The ice residues were brought back into solution with 9 M HCl for chemical separations of the elements.

BET surface areas of the sediment were measured prior to dissolution; the filters with the > 30 mm and < 30 mm > 0.05 mm size fraction were placed in capped 60 mL Teflon beakers with 2-3 mL of double distilled ethanol and the beakers placed in an ultrasonic bath. The ethanol solutions were decanted into 15 mL beakers and the ethanol allowed to evaporate down to until the samples were a thick slurry. Aliquots of this slurry were loaded into small gold cups, after drying the aliquots ranged from 1-5 mg. The remaining slurry was allowed to fully dry and weighed. After BET measurement, the gold cups were placed back into their original 15 mL beakers with ethanol, the beakers ultra-sonicated then the ethanol decanted into the 60 mL beakers containing the original filters. In order to remove any dust grains adhering to the cups, 0.5 mL of 3 M $HNO_3$ was added to the small beakers and warmed at 100$^{o}$C overnight, the beakers ultra-sonicated then the $HNO_3$ decanted into the 60 mL beakers containing the original filters. This last step was repeated once more and the gold cups set aside for reuse. The spike of $^{236}U$ and $HNO_3$ and HF were added to the 60 mL beakers. After digesting for 2 days, the sediment solutions were decanted, saved, and the filter rinsed with $HNO_3$ and HCl before the combined rinses were dried down and brought back into solution with 9 M HCl for chemical separations. These procedures ensured quantitative recovery and dissolution of the dust grains both from the gold cups and off the filters.

Because the amount of sediment in these samples was high (>10 mg), aliquots of the sediment sample solutions were used for chemical separation and measurement. Ice and dust solutions were chemically separated using the methods described in Aciego et al., 2009.

Clean U fractions were measured on the ETH Nu Instruments MC-ICPMS equipped with multiple ion counters and a retardation filter. Uranium was measured statically with $^{234}U$ collected in ion counter IC1. Each sample was bracketed with a solution of SRM960, diluted to a similar concentration as the samples (2–5 ppb, equivalent to 1 ng of sample uranium). The $^{238}U/^{234}U$ ratio (18,918; Cheng et al., 2000) of the bracketing SRM960 was used to calculate the ion counter efficiency. Ratios were corrected for mass bias by normalizing to $^{238}U/^{235}U = 137.88$ using the exponential law; $^{238}U/^{236}U$ ratios of the samples were corrected for mass bias using the $^{238}U/^{235}U$ ratio of the bracketing standard. Measurements were acquired in 30 cycles of 8 second integrations after 30 s of baseline measurements at the half masses.

Sr and Nd fractions were measured on the ETH ThermoScientific Triton TIMS. Both elements were loaded onto rhenium filaments that had been previously outgassed for 2 h at 4 A and 10 min at 5 A. Solution standards of SRM987 for Sr and JNdi-1 for Nd were measured during the same TIMS runs as the samples. An internal mass fractionation correction was applied to $^{87}Sr/^{86}Sr$ using the exponential law and the isotope ratios were obtained by normalizing to $^{88}Sr/^{86}Sr=8.375209$. Rubidium interferences were corrected by monitoring $^{85}Rb$ in the center cup. Rubidium beam sizes were less than 10−13 A, such that corrections did not exceed 100 ppm. Nd measurements were made on beam sizes of greater than 100 mV of $^{144}Nd$. Data was acquired in 4 second integrations for at least 200 cycles. $^{143}Nd/^{144}Nd$ was normalized to $^{146}Nd/^{144}Nd$ to 0.7219
using the exponential law and mass 150 was monitored for Sm interferences; no Sm was ever detected.

Blanks were measured using 200 mL of 18.4 MOhm H2O; filters and filtrate were processed using the same amount of acids and procedures as the ice core samples with the addition of $^{84}Sr$ and $^{150}Nd$ spikes to $^{236}U$. Total procedural blanks were 0.5-1 pg for U and Nd and 125 pg for Sr.

## *Results*

Table 1: Basal Ice Data*

| Sample depth (m) | U ppt | $2\sigma$ | $(^{234}U/^{238}U)$ | $2\sigma$ | $^{87}Sr/^{86}Sr$ | $2\sigma$ | $^{143}Nd/^{144}Nd$ | $2\sigma$ | $S_{BET}$ $m^2/g$ | $2\sigma$ | D | $2\sigma$ | f | $2\sigma$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CC1 (D) | 9.81 | 2 | 1.03 | 1 | | | | | 24.92 | 16 | 2.444 | 10 | 0.0979 | 24 |
| CC1 (D)$> 30\mu m$ | 124.33 | 24 | | | | | | | 15.66 | 15 | 2.486 | 17 | 0.0541 | 14 |
| CC1 (I) | 31.23 | 7 | 1.173 | 8 | 0.715810 | 12 | 0.511399 | 34 | | | | | | |
| CC2 (D) | 27.56 | 3 | 1.03 | 1 | 0.746074 | 17 | | | 43.02 | 25 | 2.484 | 2 | 0.1496 | 36 |
| CC2 (D)$> 30\mu m$ | 424.76 | | | | | | | | 14.01 | 13 | 2.445 | 12 | 0.0549 | 13 |
| CC2 (I) | 49.77 | 0.11 | 1.315 | 8 | 0.723355 | 9 | 0.511896 | 30 | | | | | | |
| CC3 (D) | | | | | | | | | 36.59 | 38 | 2.429 | 17 | 0.1504 | 36 |
| CC3 (I) | 57.61 | 0.13 | 1.167 | 9 | 0.728792 | 9 | 0.511545 | 34 | | | | | | |
| Dye3 (D) | 12.55 | 3 | 1.00 | 1 | 0.732030 | 9 | 0.510653 | 24 | 22.61 | 10 | 2.502 | 7 | 0.0745 | 18 |
| Dye3 (D) $> 30\mu m$ | 86.43 | 18 | | | | | | | 1.00 | 3 | 2.437 | 30 | 0.0040 | 1 |
| Dye3 (I) | 10.27 | 0.02 | 1.054 | 10 | 0.773057 | 9 | 0.510506 | 30 | | | | | | |
| TG (D) | | | | | 0.714814 | 12 | 0.512001 | 24 | 10.71 | 21 | 2.525 | 30 | 0.0329 | 1 |
| TG (I) | 1.26 | 0.01 | 1.261 | 12 | 0.709322 | 12 | 0.512208 | 40 | | | | | | |
| SG (D) | | | | | 0.712537 | 10 | 0.512138 | 24 | 67.45 | 36 | 2.479 | 10 | 0.2381 | 57 |
| SG (D) $> 30\mu m$ | | | | | | | | | 2.18 | 3 | 2.341 | 22 | 0.0117 | 1 |
| SG (I) | 15.43 | 0.03 | 1.243 | 8 | 0.712024 | 10 | 0.512085 | 24 | | | | | | |

Surface adsorption measurements on the sediment prior to dissolution yielded both surface area and fractal dimension, D, values.

Recoil ages

Table 2: Basal Ice Ages

| Sample | Age[a] ka | $2\sigma$ | Age[b] ka | $2\ \sigma$ |
|---|---|---|---|---|
| CC1 | 131 | 19 | 384 | 39 |
| CC2 | 196 | 24 | 281 | 31 |
| Dye3 | N/A | | 182 | 25 |
| TG | 230 | -25, $+\infty$ | N/A | |
| SG | 122 | 13 | 240 | 27 |

a indicates initial $^{234}U/^{238}U$ of seawater, 1.14

b indicates initial $^{234}U/^{238}U$ has been reset, ratio of sediment used (1.03 and 1.00)

# Biological method

## *Amino Acid Racemization*

[1]Beatrice Demarchi, [1]Matthew Collins, [2]Enrico Cappellini and [1]Kirsty Penkman

[1]BioArch, Depart. of Archeology, University of York, UK

[2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen

### Introduction

The extent of degradation of the proteins within sediments can be quantified on the basis of different diagenesis indicators: protein hydrolysis, amino acid racemization (AAR) and decomposition. AAR has been widely used as a geochronological tool for dating a range of substrates, with a particular focus on biominerals, i.e. mineralised organisms where a fraction of the original biomolecules is trapped within an inorganic skeleton (e.g. Hare & Abelson, 1968; Bada & Schroeder, 1975). The technique is based on the *post-mortem* spontaneous interconversion between two chiral forms of a single amino acid (D- and L-enantiomers). L-enantiomers are exclusively represented in living organisms, with the important exception of the cell wall of micro-organisms such as bacteria (Corrigan, 1969) and, as recently discovered, in some specific proteins in higher organisms (Fujii, 2002). AAR geochronology is based on the principle that, as diagenesis proceeds with time, racemization approaches equilibrium and the ratio between [D-] and [L-] enantiomers (here referred to as DL ratio or D/L value) increases from 0 to 1. The DL ratios therefore yield an estimate of the time since death (or cessation of tissue turnover) of a biomineralised organism.
However, the rate of racemization of amino acids in the burial environment is affected by a range of factors, including temperature, presence of water in the environment, pH, pressure and the chelation of amino acids by some metal ions [*Bada and Schroeder*, 1975; *Poinar et al.*, 1996]; all these factors also affect the spontaneous degradation of DNA [*Bromham and Penny*, 2003; *Peterson and Butterfield*, 2005]. In the light of these multiple factors, past studies have focussed on establishing correlations between the extent of racemization of specific amino acids, namely

aspartic acid (Asx), and the probability of recovering DNA from ancient and degraded samples. A first threshold was established for Asx D/L < 0.08 for bone [*Poinar et al.*, 1996], but the reliability of protein diagenesis in bone as an indicator for DNA preservation has been further questioned (Collins et al., 1999; Ottoni et al 2008; Collins et al 2009). However, AAR is potentially useful as an indicator of the degree of preservation of organic material [*Willerslev et al.*, 2007b].

Extensive investigation of the patterns of protein diagenesis in a range of biominerals has shown that AAR can be used as a reliable geochronometer only when it can be assumed that protein diagenesis has occurred within a closed system (Brooks et al 1990; Penkman et al. 2007). It is essential that the original proteins contained in an organism have degraded *in situ*, both the degraded proteins and their by-products can be recovered from the system and diagenesis has been affected by temperature and time only (and not by other environmental factors) (Towe 1980; Brooks et al 1990; Sykes et al 1995, Collins & Riley 2000; Penkman et al 2008). Moreover, direct comparison of the D/L values can be safely performed only when the same protein fraction is isolated from a range of samples, due to the effect of the protein amino acid sequence on the overall racemization rate (e.g. Liardon & Ledermann 1986)

These conditions are satisfied for numerous taxa of mollusc shells and avian eggshells, due to the presence of an intracrystalline fraction of proteins within the mineral structure, behaving as a closed system with regard to protein diagenesis and which can be isolated by bleaching (e.g. Brooks et al 1990; Penkman et al 2008; Demarchi et al 2011). On the contrary, protein sources within basal ice are complex and diagenesis can not be assumed to have occurred within a closed system environment; therefore the racemization patterns are heavily influenced by differences in amino acid composition as well as by a suite of factors operating in the burial environment, including the contribution of bacterial activity. Due to the heterogeneity of the samples analysed, it is possible that these differences could operate on a very small spatial scale.

In a previous study by Willerslev et al. (2007) silty ice samples from the Greenland cores GRIP and Dye 3 underwent amino acid analysis and assessed for closed system behaviour. This initial test highlighted that the extent of racemization of multiple amino acids recovered from the soil can potentially be used as a geochronological indicator, if:

- The environmental factors affecting the racemization rate can be assumed to be unimportant
- The D/L values increase with age due to chemical degradation and not as a result of the incorporation of bacterial cell walls (peptidoglycans) which are enriched in D- amino acids
- The protein composition of the soil is known and therefore the appropriate kinetic parameters to use for modelling the reaction can be established

This study expanded on this previous research and investigated the feasibility of this approach on a range of silty ice specimens from both Greenland and Antarctica, with the aim of: a) assessing the level of natural variability within each soil specimen; b) evaluating the contribution of D-amino acids of bacterial origin on the overall extent of racemization; c) testing the reliability of a “closed system” model for ice cores and its potential to be used as a geochronometer.

## Materials and methods

Chiral amino acid analysis was performed by reverse-phase high performance liquid chromatography (RP-HPLC) on the amino acids extracted from a range of silty ice samples. The extraction protocol is described below.

The frozen basal ice samples were defrosted and, whenever possible, multiple subsamples were taken from each specimen (tube) (see below table).

| AAR samples | NEaar n. | Subsamples | Asx D/L ± 1 s.d. | Glx D/L ± 1 s.d. | Ala D/L ± 1 s.d. |
|---|---|---|---|---|---|
| 1049-B | 6521, 6619, 6620, 6641 | 4 | 0.135±0.020 | 0.071±0.018 | 0.070±0.024 |
| 1050A-2 | 6520 | 1 | 0.180±0.002 | 0.106±0.000 | 0.109±0.004 |
| 1052-C | 6623, 6624 | 2 | 0.136±0.003 | 0.061±0.001 | 0.054±0.002 |
| 1054-C | 6621, 6622, 6640 | 3 | 0.156±0.014 | 0.074±0.010 | 0.077±0.011 |
| 1055-B | 6523 | 1 | 0.111±0.003 | 0.056±0.001 | 0.054±0.003 |
| 1057-A | 6522 | 1 | 0.097±0.001 | 0.052±0.000 | 0.048±0.000 |
| 1057-B | 6617, 6618, 6643 | 3 | 0.157±0.004 | 0.081±0.001 | 0.091±0.011 |
| Su 98/3 Amber | 6525 | 1 | 0.109±0.003 | 0.091±0.002 | 0.124±0.005 |
| Su 98/3 | 6625, 6626, 6645 | 3 | 0.076±0.013 | 0.122±0.037 | 0.103±0.016 |
| TG 6-B | 6524 | 1 | 0.115±0.001 | 0. 089±0.001 | 0.100±0.000 |
| TGT-CB | 6627, 6628, 6644 | 3 | 0.129±0.023 | 0.135±0.019 | 0.114±0.010 |

Each subsample was accurately weighed (between 30-200 mg were generally taken, including both silty particles and water) and placed in a sterile glass vials; 500 µL of 7M HCl were added to each subsample. After flushing with $N_2$ in order to minimise oxidation, the vials were tightly capped and placed in an oven at 110ºC for 24 hours, in order to release the peptide bound amino acids (Total Hydrolysable Amino Acids, or THAA).

Table X: AAR sample analysed – sample name, NEaar unique number, number of subsamples analysed. Also reported are Asx, Glx and Ala D/L values for each basal ice analysed (average of all analytical and laboratory replicates and one standard deviation around the mean).

After hydrolysis, the samples were centrifuged in order to separate the hydrolysate from the residual soil and the acid solution transferred in new sterile glass vials. These were then placed in a centrifugal evaporator overnight in order to remove the acid by evaporation. The THAA fraction thus extracted, however, contains a range of interfering organic substances and metal ions (Amelung & Zhang 2001) which could hamper the chromatographic resolution. Successful purification had been previously achieved by solid phase extraction, using a strong cation exchange resin (Willerslev et al., 2007). Here we present an improved version of this method for the purification of amino acids from basal ice, which allows for higher throughput, reducing manipulation of hazardous substances. The possibility of processing faster multiple subsamples from each ice core sample allows improving the estimate of the natural variability. The new methodology was based on the use of Stage-tips (Rappsilber et al., 2007) fitted with an Empore™ Cation Exchange-SR membrane (3M, Neuss, Germany), combined with Dowex® 50 WX8-200 IE resin (Sigma, Copenhagen, Denmark). Both were placed in 200 µL pipette tips and each THAA sample was resuspended in 200 µL of 0.05M HCl and purified in an individual tip. Using an adapter, the tip was mounted on a 2mL tube and fit in a bench-top centrifuge. The purification protocol includes the following steps, which were performed at the same time for all the samples:

- Pre-rinse with 2M NaOH
- Condition with 2M HCl
- Wash with 0.05 M HCl
- Load sample
- Rinse with 0.1 M oxalic acid (pH 1.6-1.8)
- Wash with 0.01 M HCl
- Elute with 2.5 M $NH_4OH$

Between each step, centrifuging (5000 RPM for 3 minutes) is carried out in order to aid the penetration of each solution loaded onto the tips through the resin. The

volumes are between 150-200 µL for each solution. The solution of amino acids eluted in $NH_4OH$ is subsequently evaporated to dryness in a centrifugal evaporator and the purified amino acids are rehydrated in the rehydration fluid routinely used in the NEaar laboratory. This is a solution containing 0.01 mM HCl, 0.77 mM sodium azide at a pH of 2 and an internal standard (0.01 mM L-homo-arginine), enabling quantitative analysis of the amino acids. Two analytical replicates were obtained for each individual subsample.

Analysis was carried out using a modified method of Kaufman and Manley (1998) for RP-HPLC, which enables the routine quantification of the enantiomers of multiple amino acids. In this study we considered Asp/Asn, Glu/Gln, Ser, L-Thr, L-His, Gly, L-Arg, Ala, LhArg (internal standard), Val, Phe, Leu, Ile. Each samples was analysed twice on the HPLC. During preparative hydrolysis both Asn and Gln undergo rapid irreversible deamination to Asp and Glu respectively (Hill 1965). It is therefore not possible to distinguish between the acidic amino acids and their derivatives and they are reported together as Asx and Glx. Here we focus our discussion mainly on Asx, Glx and Ala, as these three amino acids yielded the best analytical resolution.

### *Results*

The first aim of this study was to test whether the fraction of amino acids extracted from a range of silty soils from basal ice showed similar relative compositions and low within-sample variability, as this would enable more reliable direct comparison between the racemization data. For this purpose, it is important to consider multiple laboratory replicates (subsamples) for each soil sample. Fig 4.2.1.A shows that the [THAA] recovered from each subsample were extremely variable; the relative composition also differed among subsamples taken from each soil specimen (Fig. 4.2.1.B). This strongly suggests that the original soil samples are very heterogeneous, that the sources of proteins are diverse and that their distribution is variable across few cubic centimetres.

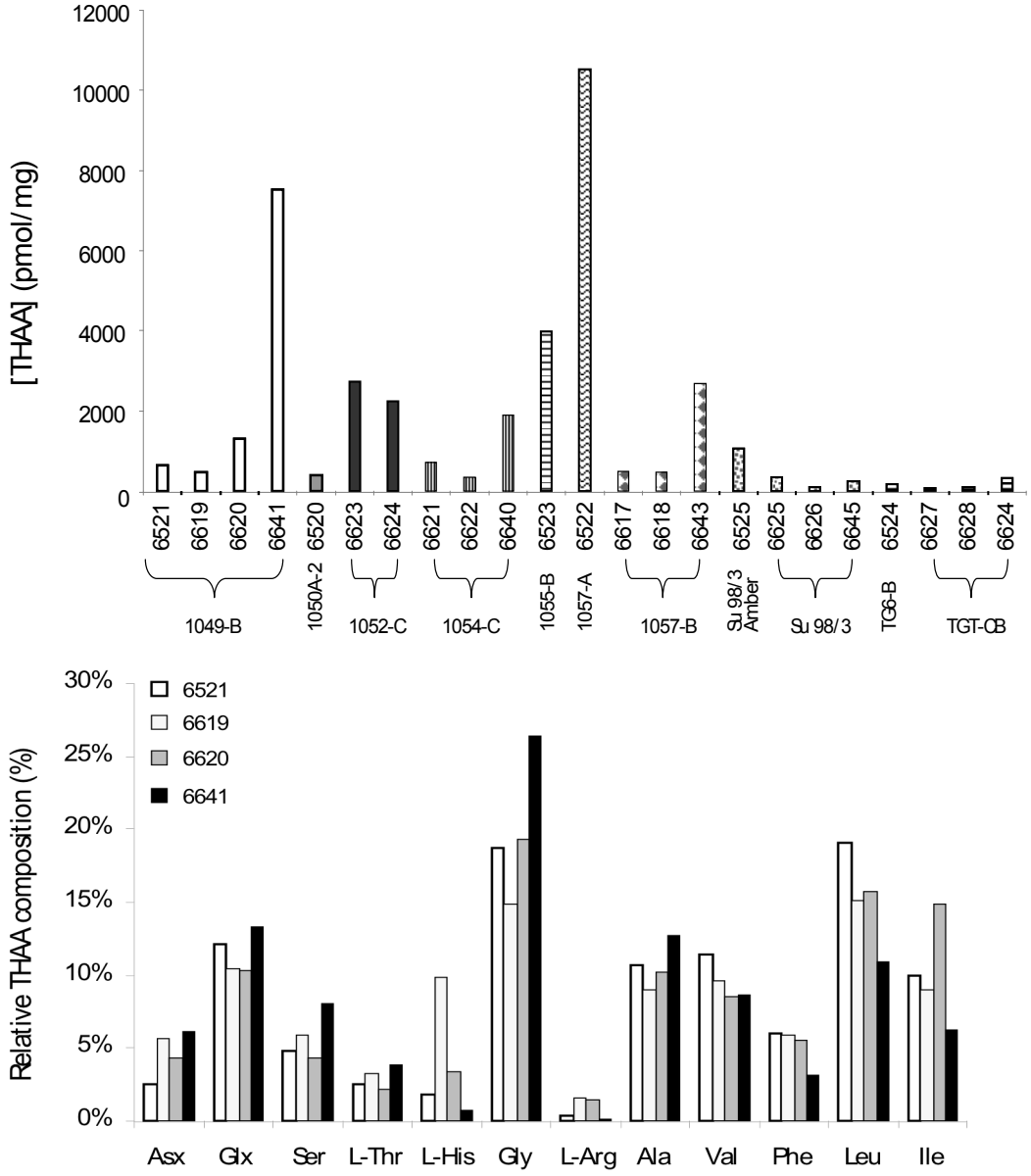



**Fig 4.2.1.A (top) and 4.2.1.B (bottom): Absolute THAA concentration recovered from each soil subsample and relative amino acid composition amongst four subsamples from soil sample 1049-B**

This compositional heterogeneity is reflected in the range of D/L values of multiple amino acids, measured in different subsamples (Fig. 4.2.1.C and Table X). A plot of the DL values of two amino acids is used in aminochronological studies as a powerful tool to assess the extent of racemization in different samples: within a closed system samples should fall on a trajectory of increase of degradation, from the origin of the axis to the top right corner of the plot (e.g. Penkman et al 2007; Demarchi et al 2011). Samples falling outside this trendline display non-closed system behaviour and should be discarded (Kosnik & Kaufman 2008).

Two main clusters of data can be observed in Fig. 4.2.1.C, which represent the Antarctica and the Greenland samples, respectively. It is therefore clear that two different patterns of diagenesis characterise the two subsets of samples. In particular, Glx and Ala D/L values are consistently *higher* in the Antarctic samples compared with the Greenland samples. This strongly suggests a contribution of proteinaceous material of bacterial origin, as it is known that peptidoglycans in the bacterial cell walls are enriched in D-Glx and D-Ala (Schleifer & Kandler, 1978), and has been used as a method to track recalcitrant nitrogen (McCarthy et al., 1998). This contribution is therefore of significant importance for the Antarctica soils. However, it is also evident that different subsamples taken from the same basal ice sample also display a range of Glx and Ala D/L values, which describe trends of increase of the extent of degradation (see for example samples 1049-B from Greenland and TGT from Antarctica, fig. 4.2.1.C). This could potentially be explained by modern bacterial contamination and differential protein breakdown during sample preparation. However, all the soil specimens were manipulated in a sterile or semi-sterile environment, stored at -20ºC and, upon defrosting, subsamples were taken at the same time. This would rule out bacterial contamination during preparation. It is therefore more likely that the heterogeneity of the bulk samples reflects the presence of ancient bacterial peptidoglycans due to elevated levels of active microbial recycling of biomass at some point in time, potentially when temperatures were more favourable for this in a pre-glacial environment.

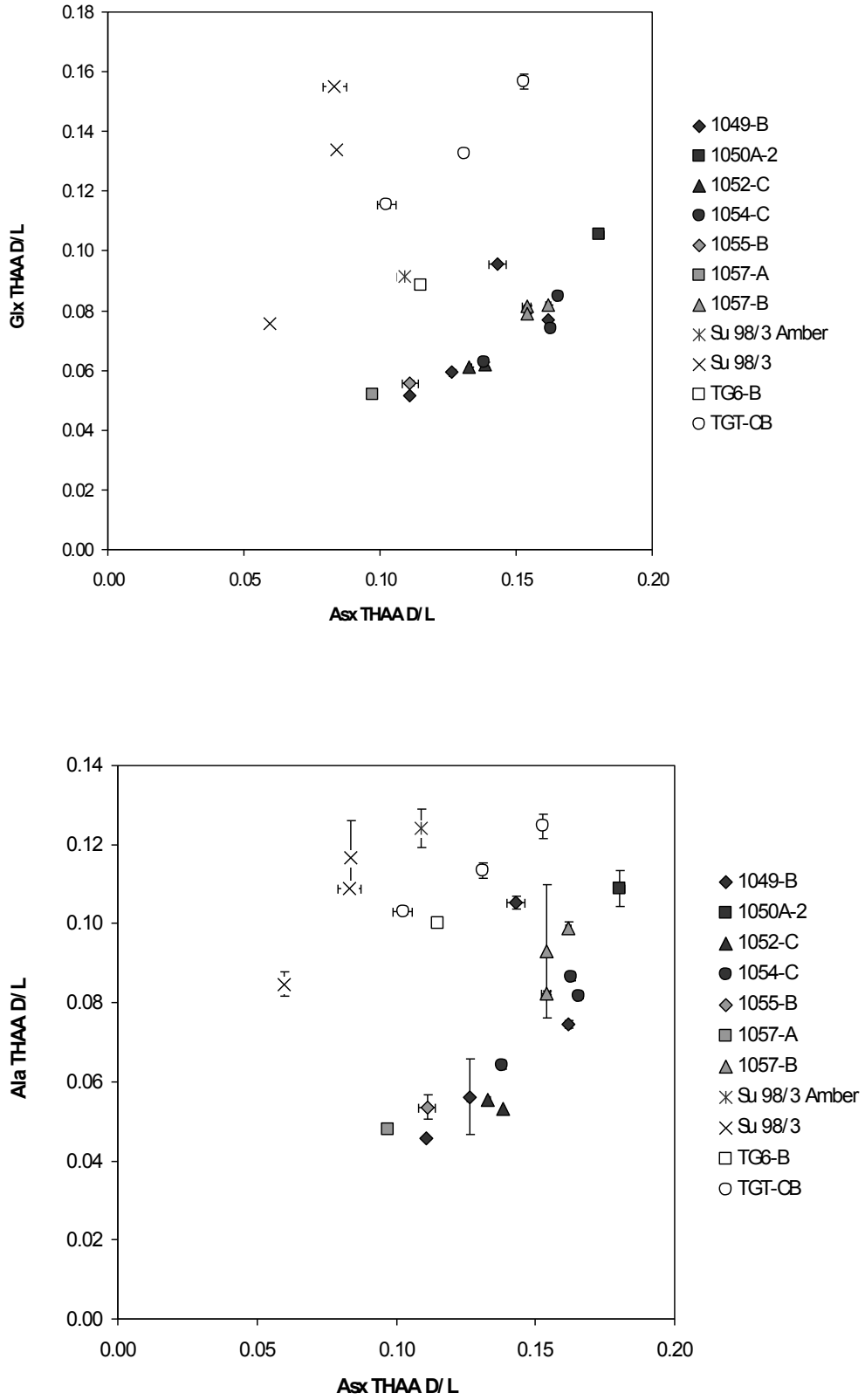



**Fig 4.2.1.C: THAA D/L values plots of Glx vs Asx (top) and Ala vs Asx (bottom) displaying both the difference between samples from Antarctica and Greenland and**

**between different subsamples taken from the same basal ice specimen. Error bars represent one standard deviation around the mean of two analytical replicates,**

In order to clarify this further, the samples analysed from Greenland and Antarctica in this study were compared with the previous basal ice samples reported in Willerslev et al., (2007). This study compared the 3 ka John Evans Glacier (JEG) with the extent of diagenesis recovered from old basal ice from the Greenland cores GRIP and Dye 3. Soil and ice were analysed separately for JEG, allowing for a distinction between the behaviour of diageneis in an open (soil) and a potentially closed (ice) system. This young glacier was then used as a control for the initial levels of racemization after incorporation of the soil into the ice.

The most significant observation which can be drawn from Fig. 4.2.1.D is that, while the "JEG soil" plots within the cluster of the Antarctica samples, thus suggestive of the soil-like nature of these samples; on the contrary, GRIP and "JEG ice" are at the two extremes of the "Greenland trajectory", supporting the hypothesis of an environment less affected by bacterial activity.

The study by Willerslev et al (2007) had applied a strategy for assessing closed and open system behaviour of basal ice, attempting to distinguish between chemical degradation and bacterial degradation. After normalisation against Asx D/L, the extent of racemization of Asx, Glx and Ala exhibited two major diagenetic trajectories, postulated to describe theoretical patterns for solely chemical vs solely bacterial degradation.

If the samples analysed in this study are plotted on a similar graph, it is clear that most of the Antarctica samples display values which fall closer to the "open system" trajectory (i.e. typical "soil" degradation), while most of the Greenland samples fall in an intermediate position, suggesting breakdown that is predominantly chemically driven (Fig 4.2.1.E). Nonetheless, we stress that the considerable intra-sample variability limits the age interpretation, as the differing composition of the original starting materials will strongly influence the observed extent of racemisation. Due to the wide range of D/L values measured within the same sample, we propose that more complex factors than a simple dichotomy between chemical/bacterial degradation must be considered in order to use AAR as a geochronological tool for basal ice (contra Willerslev et al., 2007).

We suggest that a better understanding of the patterns of diagenesis in this substrate could be achieved if the amino acid compositions of the bacterial peptidoglycans in each sample were known, enabling this to be accounted for in the initial "anomalous" high Glx and Ala D/L values. Moreover, the initial soil heterogeneity has proven to be a serious limitation to this type of investigation. Sample pre-treatments aimed at reducing such variability (e.g. separation of liquid and solid phases, homogenising the samples by vortexing) would most probably produce more reproducible AAR results. However, our study has highlighted the high degree of natural variability in the amino acid composition of basal ices, likely to be an inherent characteristic of such substrates; artificially reducing this variability would hamper our understanding of soil depositional dynamics as well as disregard the potential for material of widely different ages to be present within small portions of basal ice.

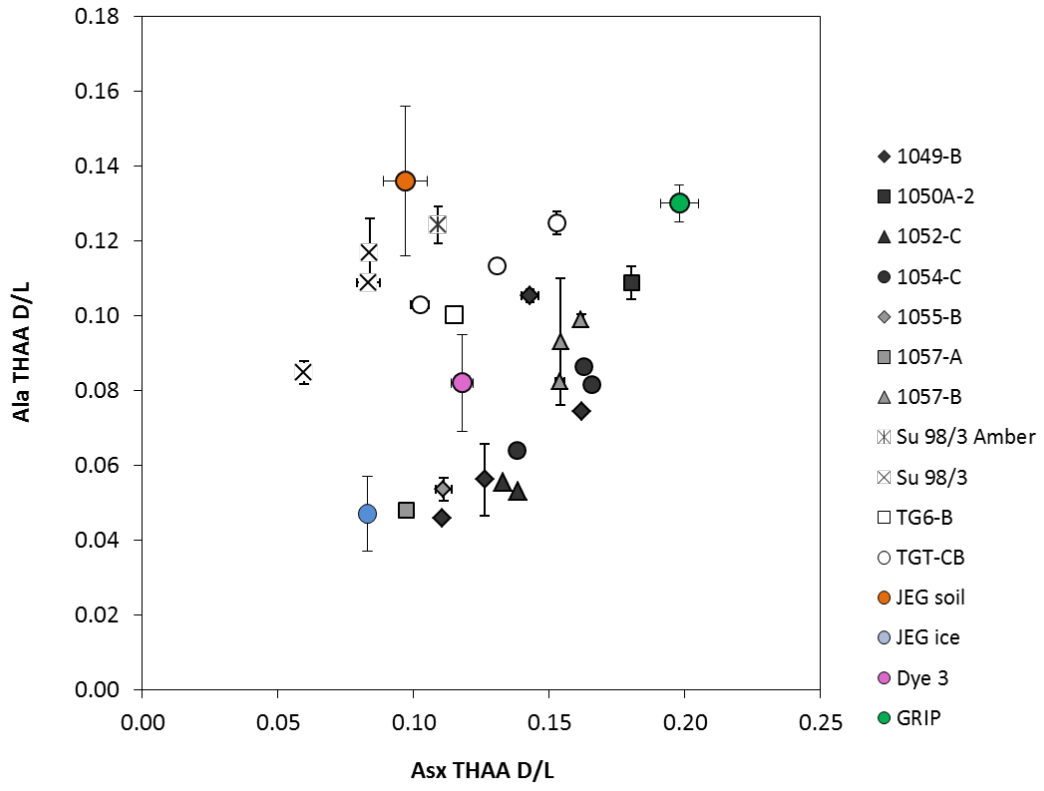



**Fig 4.2.1.D : Comparison between samples analysed in this study and samples from the Willerslev et al (2007) study on a THAA Ala D/L vs THAA Asx D/L plot.**

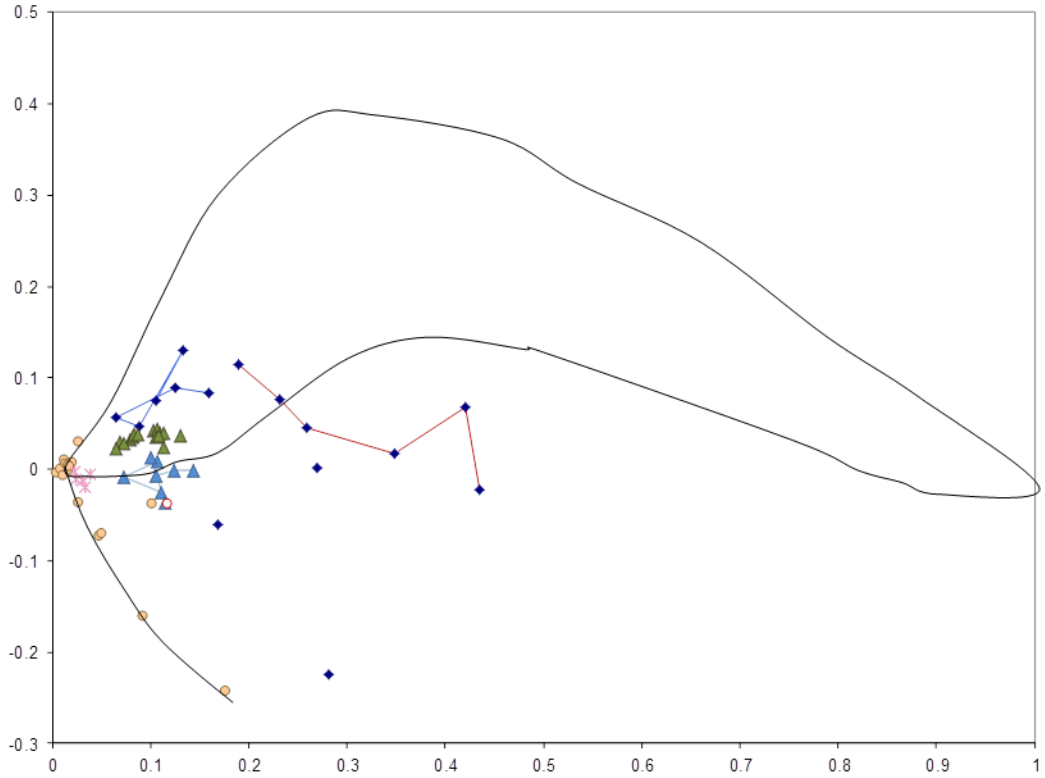



Fig 4.2.1.E : Trends of closed-system and open-system degradation. Closed system data from biominerals (Penkman et al, unpublished).

Table 1. List of the samples that have been dated by the different dating methods from Greenland and Antarctica.

| Geography | Location | Be/Cl dating | ~Depth(m) | Weight ice/soil | Method |
|---|---|---|---|---|---|
| Antarctica | Suess glacier | Su #98/2 | | 1067 g | U/U |
| Antarctica | Suess glacier | Amber #98/2 | | ~1000g | Be/Cl |
| Antarctica | Suess glacier | Amber #98/3 | | /~0.01 g | Racemization |
| Antarctica | Taylor glacier | TG-6-B 1 | | /0.66 g | Racemization |
| Antarctica | Taylor glacier | TG-6-B 2 | | ~1000g | Be_Cl |
| Antarctica | Taylor glacier | TGTB9 | | 1058 g | U/U |
| Greenland | Camp Century | 818 | ~1023 | ~1000g | Be_Cl |
| Greenland | Camp Century | 847 | ~1064 | ~1000g | Be_Cl |
| Greenland | Camp Century | 935 A | ~1187 | ~ 950 g | Be_Cl |
| Greenland | Camp Century | 972 B | ~1241 | ~1050 g | Be_Cl |
| Greenland | Camp Century | 988 B | ~1264 | 1300 g | Be_Cl |
| Greenland | Camp Century | 989 C | ~ 126 | 953 g | Be_Cl |
| Greenland | Camp Century | 1016 D | ~ 1305 | 933 g | Be_Cl |
| Greenland | Camp Century | 1018 B | ~ 1308 | 961 g | Be_Cl |
| Greenland | Camp Century | 1018 C | ~1308.5 | small piece ~300g? | Be_Cl |
| Greenland | Camp Century | 1033 A | ~1329 | small piece ~300g? | Be_Cl |
| Greenland | Camp Century | 1037 B | ~1336 | 1080 g | Be_Cl |
| Greenland | Camp Century | 1038 B | ~ 1337 | 928 g | Be_Cl |
| Greenland | Camp Century | 1047 D | ~1350.5 | 918 g | Be_Cl |
| Greenland | Camp Century | 1048A | ~1351 | ~1000g | U/U |
| Greenland | Camp Century | 1049 A | ~1353 | /0.26 g (incl. drop of water) | OSL |
| Greenland | Camp Century | 1049 B | ~1353.5 | /0.23 g | Racemization |
| Greenland | Camp Century | 1049 B | ~1353.5 | ~1000g | Be_Cl |
| Greenland | Camp Century | 1050 A | ~1354 | /0.28 g (incl. drop of water) | Racemization |
| Greenland | Camp Century | 1050 A | ~1354 | 1135 g | Be_Cl |
| Greenland | Camp Century | 1050 A | ~1354 | ~1000g | Be_Cl |

| Geography | Location | Be/Cl dating | ~Depth(m) | Weight ice/soil | Method |
|---|---|---|---|---|---|
| Greenland | Camp Century | 1050 C | ~1355 | ~1000g | U/U |
| Greenland | Camp Century | 1052 A | ~1357 | ~1000g | Be_Cl |
| Greenland | Camp Century | 1052 C | ~1358 | ~1000g | U/U |
| Greenland | Camp Century | 1053 B | ~1359 | ~1000g | Be_Cl |
| Greenland | Camp Century | 1053 B | ~1359 | /0.40 g | OSL |
| Greenland | Camp Century | 1054 A | ~1360 | ~1000g | Be_Cl |
| Greenland | Camp Century | 1054 A | ~1360 | /0.24 g (incl. drop of water) | OSL |
| Greenland | Camp Century | 1054 A | ~1360 | ~1000g | U/U |
| Greenland | Camp Century | 1054 B | ~1360.5 | ~1000g | Be_Cl |
| Greenland | Camp Century | 1054 C | ~1361 | ~1000g | Be_Cl |
| Greenland | Camp Century | 1055 B | ~1362 | /0.10 g | Racemization |
| Greenland | Camp Century | 1055 B | ~1362 | /0.26 g (incl. drop of water) | OSL |
| Greenland | Camp Century | 1056 A | ~1363 | ~1000g | U/U |
| Greenland | Camp Century | 1057 A | ~1364.5 | /0.18 g (incl. drop of water) | Racemization |
| Greenland | Camp Century | 1057 A | ~1364.5 | /0.15 g | OSL |
| Greenland | Camp Century | 1057 A | ~1364.5 | ~1000g | Be_Cl |
| Greenland | Dye 3 | 1943 A | ~1068.5 | ~1000g | U/U |
| Greenland | Dye 3 | 1953 B | ~1074.5 | ~1000g | U/U |
| Greenland | Dye 3 | 1970 B | ~1083.5 | ~1000g | U/U |

References


Aciego, S. M., B. Bourdon, M. Lupker, and J. Rickli (2009), A new procedure for separating and measuring radiogenic isotopes (U, Th, Pa, Ra, Sr, Nd, Hf) in ice cores, *Chem Geol*, *266*(3-4), 194-204.

Aitken, M. J. (1985), *Thermoluminescence dating*.

Bada, J. L., and R. A. Schroeder (1975), Amino-Acid Racemization Reactions and Their Geochemical Implications, *Naturwissenschaften*, *62*(2), 71-79.

Baumgartner, S., J. Beer, G. Wagner, P. W. Kubik, M. Suter, G. M. Raisbeck, and F. Yiou (1997), $^{10}$Be and dust, *Nuclear Instruments and Methods*, *B123*, 296-301.

Bromham, L., and D. Penny (2003), The modern molecular clock, *Nat Rev Genet*, *4*(3), 216-224.

Chmeleff, J., F. von Blanckenburg, K. Kossert, and D. Jakob (2010), Determination of the 10Be half-life by multicollector ICP-MS and liquid scintillation counting, *Nucl. Instr. Meth. Phys. Res. B.*, *268*, 192–199.

Collins, M. J., E. R. Waite, and A. C. T. van Duin (1999), Predicting protein decomposition: the case of aspartic-acid racemization kinetics, *Philos T Roy Soc B*, *354*(1379), 51-64.

Delmas, R. J., J. Beer, H.-A. Synal, R. Muscheler, J.-R. Petit, and M. Pourchet (2004), Bomb-test $^{36}$Cl measurements in Vostok snow (Antarctica) and the use of $^{36}$Cl as a dating tool for deep ice cores., *Tellus*, *56B*, 492–498.

Hebsgaard, M. B., M. J. Phillips, and E. Willerslev (2005), Geologically ancient DNA: fact or artefact?, *Trends in Microbiology*, *13*(5), 212-220.

Heikkilä, U., J. Beer, and J. Feichter (2009a), Meridional transport and deposition of atmospheric 10Be, *Atmos. Chem. Phys.*, *9*, 515–527.

Heikkilä, U., J. Beer, J. Feichter, V. Alfimov, H.-A. Synal, U. Schotterer, A. Eichler, M. Schwikowski, and L. Thompson (2009b), 36Cl bomb peak: comparison of modeled and measured data, *Atmos. Chem. Phys.*, *9*, 4145-4156.

Horiuchi, K., T. Uchida, Y. Sakamoto, A. Ohta, H. Matsuzaki, Y. Shibata, and H. Motoyama (2008), Ice core record of (10)Be over the past millennium from Dome Fuji, Antarctica: A new proxy record of past solar activity and a powerful tool for stratigraphic dating, *Quaternary Geochronology*, *3*(3), 253-261.

Johnsen, S. J., D. Dahl-Jensen, N. Gundestrup, J. r. P. Steffensen, H. B. Clausen, H. Miller, V. Masson-Delmotte, A. E. Sveinbj  rnsdottir, and J. White (2001), Oxygen isotope and palaeotemperature records from six Greenland ice-core stations:

Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP, *Journal of Quaternary Science*, *16*(4), 299-307.

Knight, P. G. (1997), The basal ice layer of glaciers and ice sheets, *Quaternary Science Reviews*, *16*(9), 975-993.

Korschinek, G., et al. (2010), A new value for the half-life of 10Be by Heavy-Ion Elastic Recoil Detection and liquid scintillation counting, *Nucl. Instr. Meth. Phys. Res. B.*, *268*, 187-191.

Lal, D., and B. Peters (1967), Cosmic Ray Produced Radioactivity on the Earth, in *Handbuch für Physik*, edited by S. Flügge, pp. 551-612, Springer, Berlin.

Lal, D., and B. Peters (1968), COSMIC RAY PRODUCED RADIOACTIVITY ON THE EARTH, *Journal Name: pp 551-612 of Handbuch der Physik. Band XLVI/2. Fluegge, S. Sitte, K. (eds.). Berlin, Heidelberg, New York, Springer-Verlag, 1967.; Other Information: Orig. Receipt Date: 31-DEC-68*, Medium: X.

Masarik, J., and J. Beer (1999), Simulation of particle fluxes and cosmogenic nuclide production in the Earth's atmosphere, *Journal of Geophysical Research*, *104*(D10), 12,099-012,111.

Otto-Bliesner, B. L., S. J. Marsha, J. T. Overpeck, G. H. Miller, A. X. Hu, and C. L. I. P. Mem (2006), Simulating arctic climate warmth and icefield retreat in the last interglaciation, *Science*, *311*(5768), 1751-1753.

Peterson, K. J., and N. J. Butterfield (2005), Origin of the Eumetazoa: Testing ecological predictions of molecular clocks against the Proterozoic fossil record, *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9547-9552.

Poinar, H. N., M. Hoss, J. L. Bada, and S. Paabo (1996), Amino acid racemization and the preservation of ancient DNA, *Science*, *272*(5263), 864-866.

Samyn, D. (2005a), Structural and Geochemical Analysis of
Basal Ice from Taylor Glacier, Antarctica:
On the Role and Behaviour of the Interstitial Fluid Phase, Phd
thesis, Universit´e Libre de Bruxelles, Belgium, Bruxelles.

Sasa, K., et al. (2010), Measurement of cosmogenic 36Cl in the Dome Fuji ice core, Antarctica: Preliminary results for the Last Glacial Maximum and early Holocene, *Nucl. Instr. Meth. Phys. Res. B.*, *268*, 1193–1196.

Souchez, R. (2000), Basal ice formation and deformation in central Greenland: a review of existing and new ice core data, *Special publication - Geological Society of London*, *176*(1), 13.

Souchez, R., J. Tison, L, R. Lorrain, M. Lemmens, L. Janssens, M. Stievenard, J. Jouzel, A. Sveinbjˆrnsdottir, and S. J. Johnsen (1994), Stable isotopes in the basal silty ice preserved in the Greenland Ice Sheet at summit; environmental implications, *Geophys. Res. Lett.*, *21*(8), 693-696.

Sugden, D. E., D. R. Marchant, N. Potter, R. A. Souchez, G. H. Denton, C. C. Swisher, and J. L. Tison (1995), Preservation of Miocene Glacier Ice in East Antarctica, *Nature*, *376*(6539), 412-414.

Wagner, G., J. Masarik, J. Beer, S. Baumgartner, D. Imboden, P. W. Kubik, H.-A. Synal, and M. Suter (2000), Reconstruction of the geomagnetic field between 20 and 60 kyr BP from cosmogenic radionuclides in the GRIP ice core, *Nuclear Instruments and Methods in Physics Research*, *B172*, 597-604.

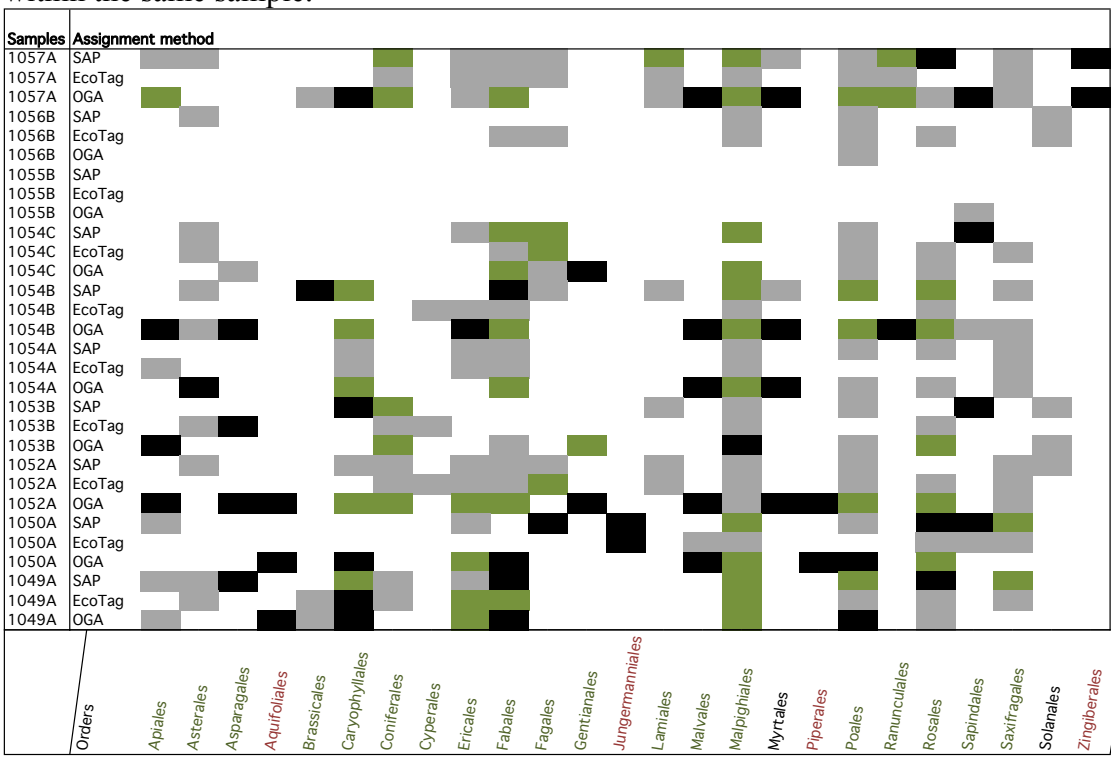
Willerslev, E., et al. (2007a), Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland, *Science*, *317*, 111-114.

Willerslev, E., et al. (2007b), Ancient Biomolecules from Deep Ice Cores Reveal a Forested Southern Greenland, *Science*, *317*(5834), 111-114.

# APPENDICES

# Appendix A

**Table 1**. Presence-absence raw data at the order level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL, black by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.



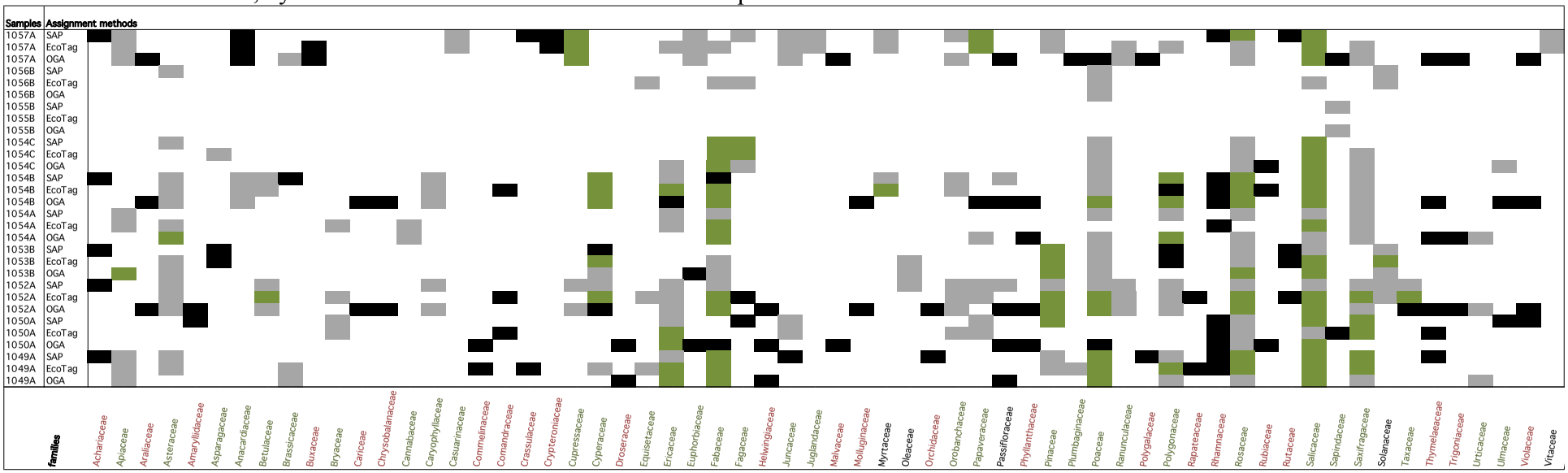

* mark where taxa were replicated on the subset of samples send to the Murdoch University clean lab, Perth, Australia.

Green writing indicates the taxa that were found by more than one method in one or several samples.

Black writing indicates the taxa that are possible contaminants from exotic food plants or misassignments.

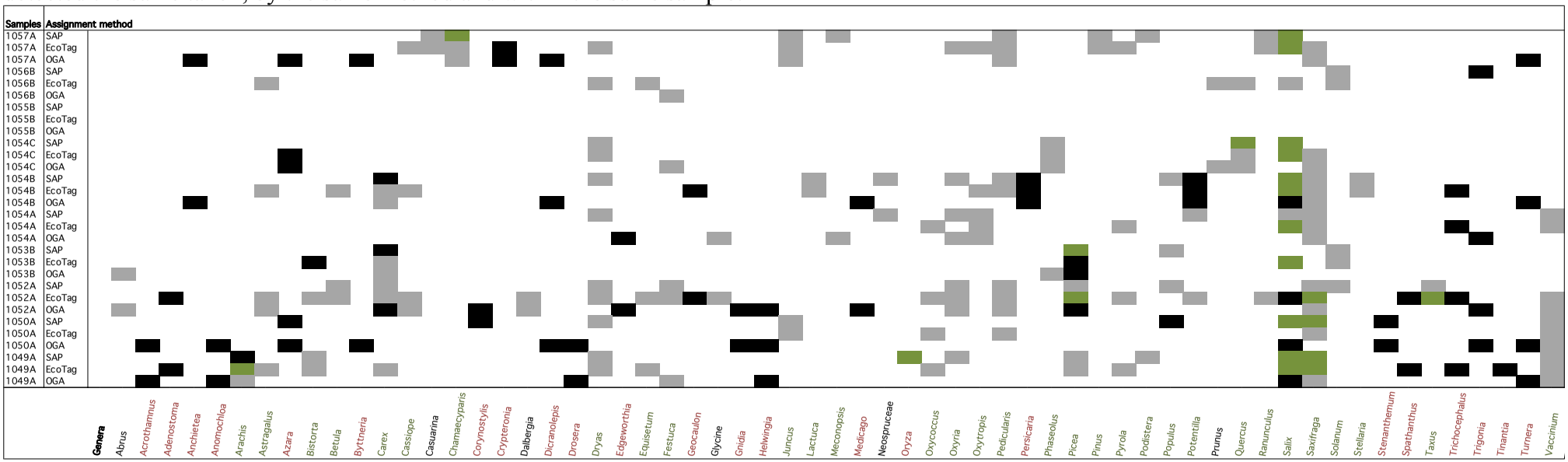
Red writing indicates which taxa have only been found by rbcL and represents tropical unlikely assignments

**Table 2**. Presence-absence raw data at the family level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL, black by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.




* mark where taxa were replicated on the subset of samples send to the Murdoch University clean lab, Perth, Australia.
Green writing indicates the taxa that were found by more than one method in one or several samples.
Black writing indicates the taxa that are possible contaminants from exotic food plants or misassignments.
Red writing indicates which taxa have only been found by rbcL and represents tropical unlikely assignments

**Table 3**. Presence-absence raw data at the genus level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL, black by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample.




* mark where taxa were replicated on the subset of samples send to the Murdoch University clean lab, Perth, Australia.
Green writing indicates the taxa that were found by more than one method in one or several samples.
Black writing indicates the taxa that are possible contaminants from exotic food plants or misassignments.
Red writing indicates which taxa have only been found by rbcL and represents tropical unlikely assignments

# Appendix B

## Temperature extrapolations for the Camp Century site based on $d^{18}O$ data

By Dorthe Dahl-Jensen, Centre for Ice and Climate, NBI, University of Copenhagen

Three scenarios were extrapolated; [(i)] the basal temperature, [(ii)] ice thickness and [(iii)] surface temperature and can briefly be described as follows.

[(i)]Basal temperature:
Red line describes the temperature fluctuations when the model was implementing a scenario where the ice retreated from Camp Century in MIS11 (420-360KY). It can be seen from figures 1-2 that the ice will be close to its melt point during the last 400KY, Moreover, it can be seen that when we move further back in time than the 400KY, the basal temperature remain cold.
The model can only simulate a primitive estimate of the ice-thickness back in time and it does not take into account if there is a threshold beyond which the ice does not build up after retreat. It is therefore not possible to deduce how long the glacial periods lasted from these temperature simulations.
Black line describes the temperature history based on currently accepted temperature models.

[(ii)] Ice thickness:
The difference between red and black line (Figure 3) is obvious and shows that the ice retreats from the Camp Century site in the warm inter-glacials whereas the black line shows a decrease in ice thickness from the current approximately 1400m during the warm periods.

[(iii)] Surface temperature:
There is not much difference between the red and black curves (Figure 4-5), except that it get warmer in the glacial periods for the red curve because the elevation of the surface goes down with the lapse rate (of 7.5°C per 100m).

Overall the plots show that if it was warm in the MIS11 and the ice melts away during this period, the ice would also have retreated in the following warm interglacials that were almost as warm. This is because the ice develops a state of equilibrium after 10,000 years and therefore it doesn't have a large effect that the MIS11 was longer than the other inter glacial stages.

**Figure 1**. Plot of the basal temperature during the last 2.5MY at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is based on the hypothesis that ice melted away during the MIS11 (420-360KY) and black line is following the current temperature models temperature profile.

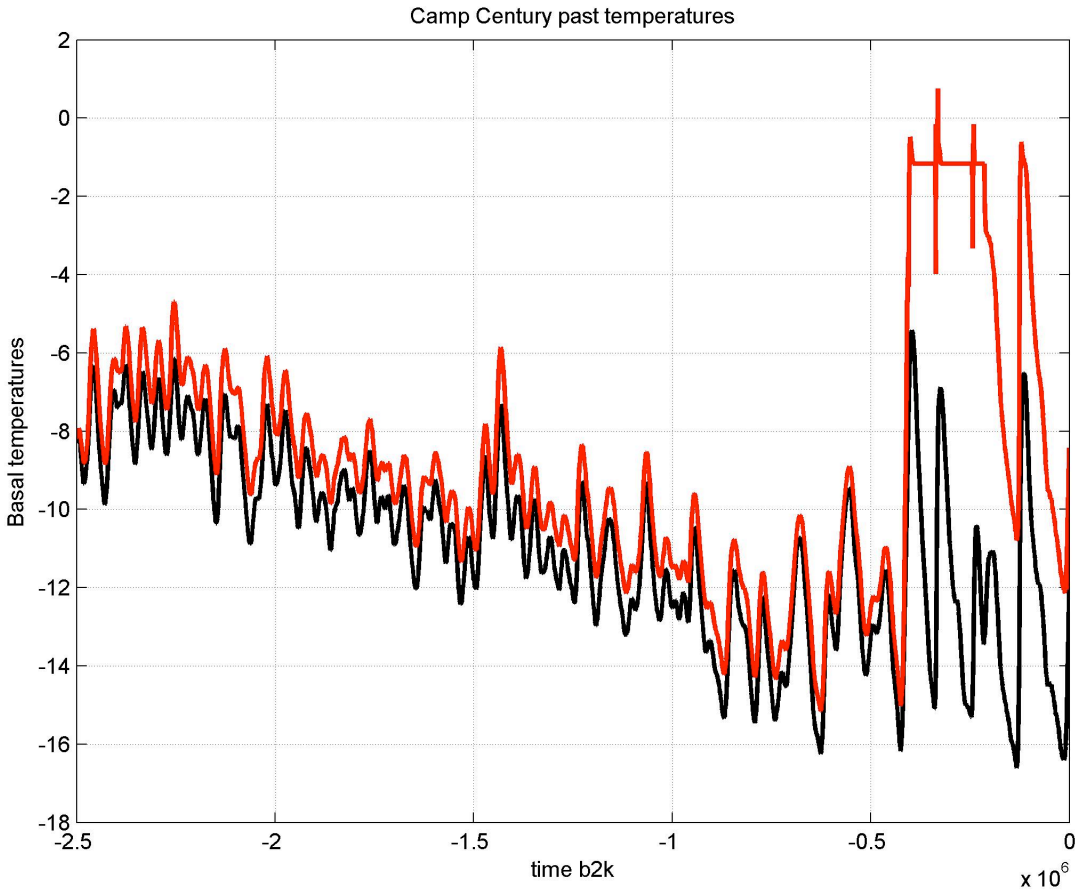


**Figure 2**. Plot of the basal temperature during the last 2.5MY when zoomed in on the period from 500KY until the present at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is fitted to a temperature scenario where the ice melted away during the MIS11 (420-360KY) at the Camp Century site and black line is following the current temperature models temperature profile.

.

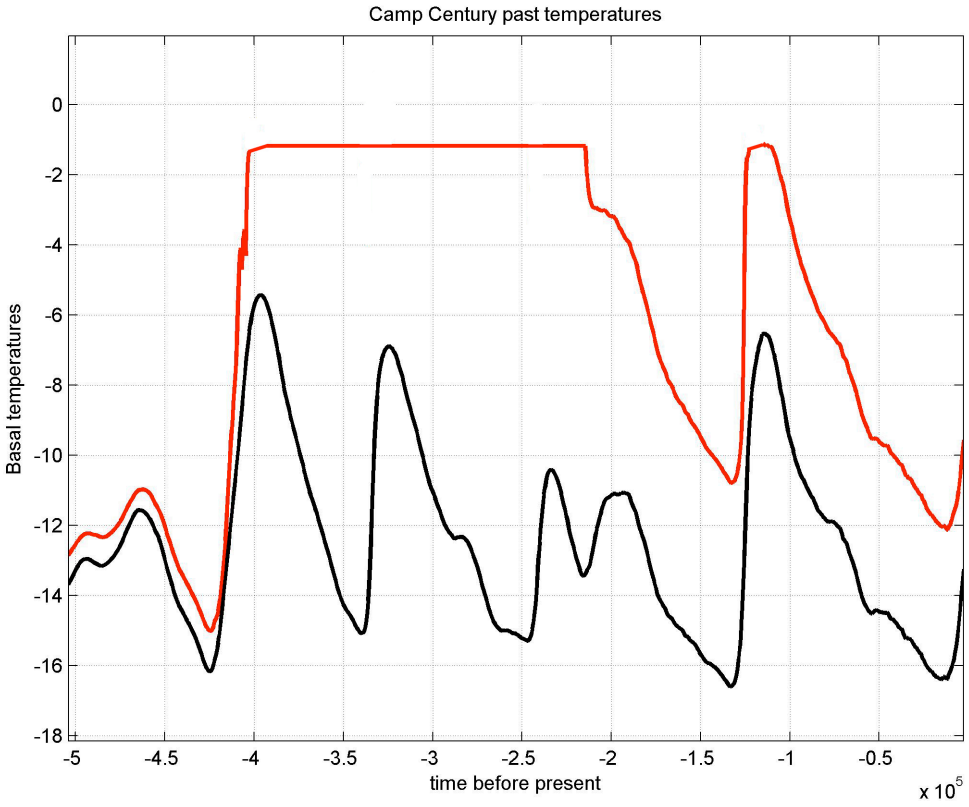


**Figure 3**. Plot of the ice thickness of the ice sheet during the last 500KY until the present at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is fitted to a temperature scenario where the ice melted away during the MIS11 (420-360KY) at the Camp Century site and black line is following the current temperature models temperature profile. It can be seen that if the ice sheet retreated during the MIS11, it would have occurred several times since then.

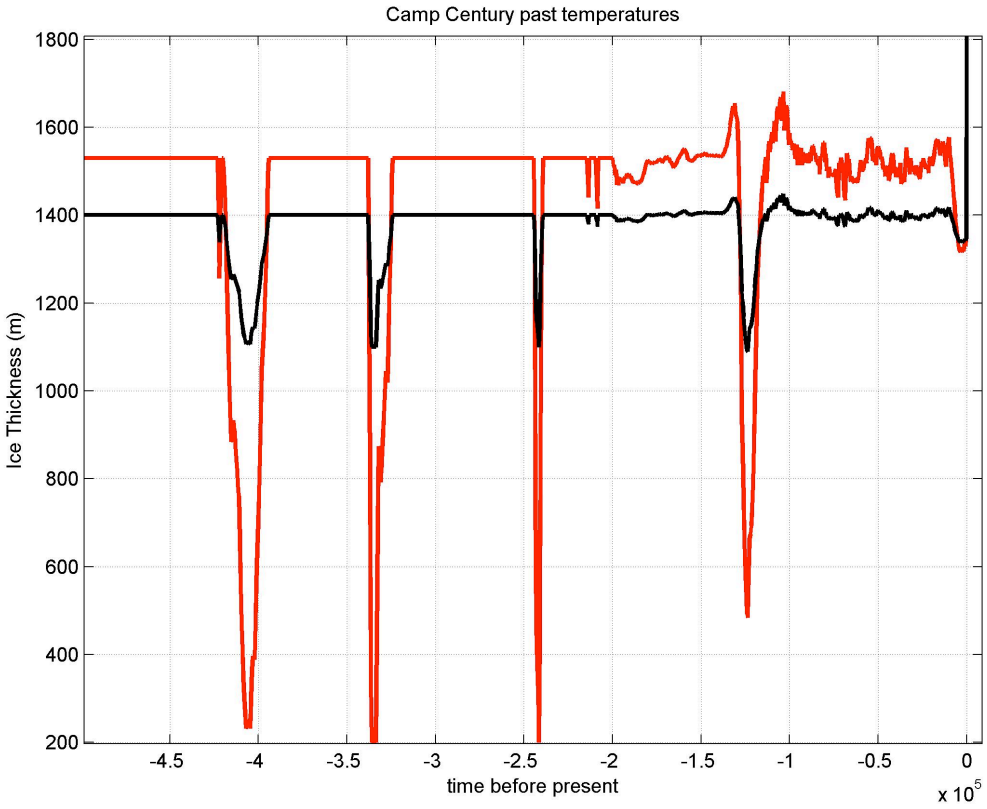


**Figure 4**. Plot of the surface temperature of the ice sheet during the last 2.5MY until the present at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is fitted to a temperature scenario where the ice melted away during the MIS11 (420-360KY) at the Camp Century site and black line is following the current temperature models temperature profile.

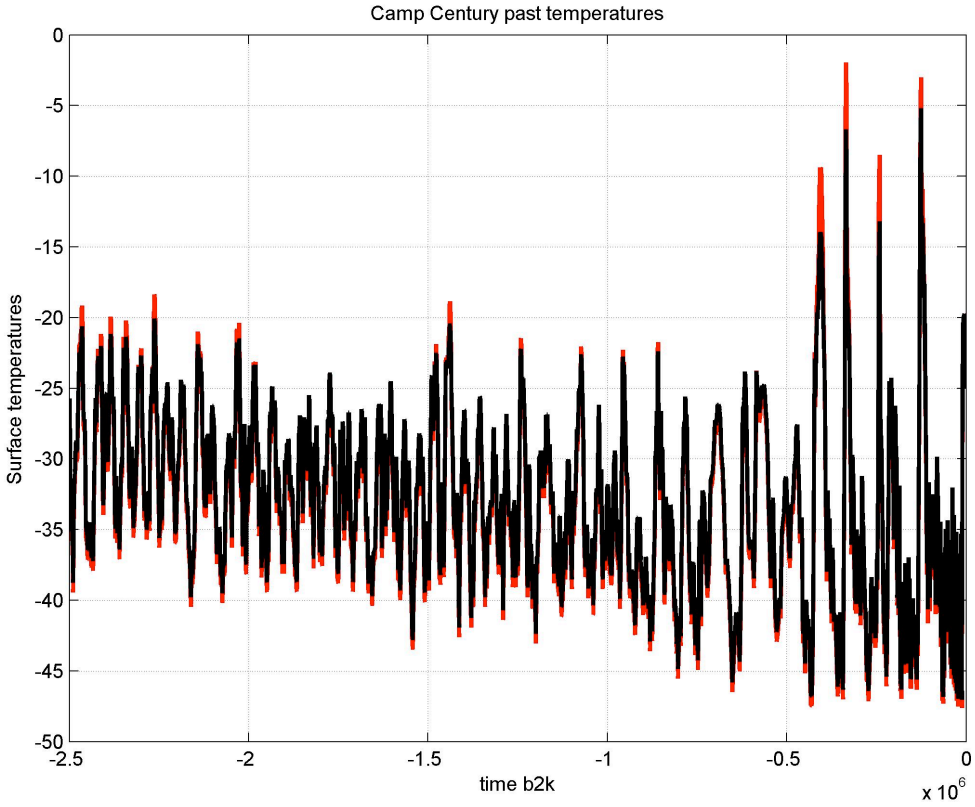
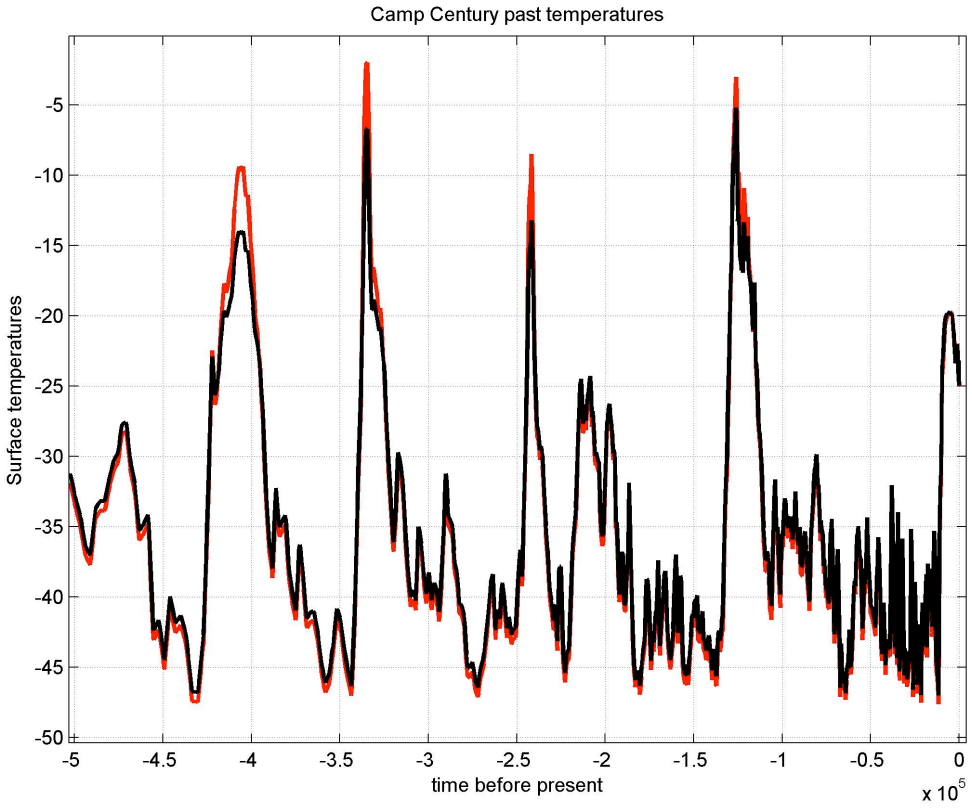


**Figure 5**. Plot of the surface temperature of the ice sheet during the last 5.0KY until the present at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is fitted to a temperature scenario where the ice melted away during the MIS11 (420-360KY) at the Camp Century site and black line is following the current temperature models temperature profile.



# Appendix C

**Pollen Analysis of Camp Century Basal Ice (Sample CC#1052A)**

Jocelyne C. Bourgeois, Geological Survey of Canada

Laboratory procedure and results

The 25 cm long sample was cut lengthwise twice: once at NEEM and then in our Ottawa cold room. The first (~ 90 mL) of the three pieces was melted, filtered, and analyzed at NEEM without any chemical treatment. It was difficult to see pollen on the microscope slides because of the large number of mineral grains. Yet, I found a few pollen grains (*Pinus* and possibly *Salix* and *Tsuga*). I also found long black particles which, after analysis with an SEM, proved to be iron oxide (probably contamination from the band saw rather than charred vegetal matter).

A second piece (200 mL) was melted inside a Whirlpack bag in Ottawa. The melted sample was filtered through a Millipore cellulose filter (8 μm pore-size) which was later dissolved in acids, using a laboratory method I've developed for snow and ice core samples. The method relies on the use of a small filtration apparatus instead of a centrifuge and gives me a better pollen recovery rate. At the end of the procedure, the pollen residue remains on a Nuclepore polycarbonate filter. The 47 mm filter is then cut in half and each half mounted on a microscope slide for analysis.

I found 75 pollen grains in the 200 mL sample (Table 1), corresponding to a concentration of 375 grains / L. *These are the results I showed you in November.* The pollen assemblage is dominated by *Betula, Pinus* and Poaceae but 25% of all the pollen grains could not be identified because of their poor condition. This is a much higher percentage than normally found in snow and ice core samples but is not surprising considering the age and depth of the sample. However, a few pollen grains looked remarkably fresh, including the 3 *Ambrosia* (ragweed) pollen. No pollen was found in the blank sample suggesting they didn't come from the laboratory.

For the third piece of ice, I went through a more rigorous laboratory procedure. I first cleaned the outside of the core by melting and scraping to remove the frost that had accumulated over the years. This is not something I've had to worry about (unlike most ice core biologists) but this ice core has probably been exposed to the ambient environment several times in the last forty years. I recuperated the 80 mL of that water for pollen analysis. It contained fresh looking *Ambrosia, Pinus* and *Betula* pollen as well as other older looking pollen types. The pollen concentration for the 80 mL is very similar to the one obtained for the 200 mL sample (400 grains/L and 375 grains/L respectively).

In the end, I was left with a 520 mL clean sample which I melted in a covered beaker. As for the other samples, I added *Lycopodium* spores to the sample to allow me to calculate the pollen concentration. I then proceeded with the usual pollen procedure. It was difficult to filter and, subsequently, difficult to count the pollen grains on the microscope

slides due to the large amount of mineral grains that remained ... even after chemical treatment. Four microscope slides were prepared from the 520 mL sample.

I counted 71 pollen grains and 47 unidentified spores in that sample (Table 1). The assemblage is dominated by Poaceae and *Betula* pollen but again, 24% of the grains could not be identified due to degradation. *Pinus* was also found but in lesser amount than previously reported. I didn't find *Ambrosia* pollen in the sample. Pollen concentration in the clean sample is lower than for the two previous samples thus suggesting that outer layer of the core might have been a source of contamination. In the comments that follow, I will only consider the results from this last sample (520 mL) which I believe to be more accurate.

Pollen concentration for the decontaminated sample is estimated to be 204 pollen grains/L. This is based on the number of *Lycopodium* spores introduced as standard, that were counted during pollen analysis. Although the concentration is lower than previously reported, it is still higher than other values found in snow and ice cores from Greenland and the Canadian Arctic (Table 2). The closest comparison comes from the Penny Ice Cap, Baffin Island, where a pollen concentration of 125 pollen grains/L was obtained in the ice core just above bedrock (at 333 m depth). This core also contained a large amount of silty ice.

In the Penny Ice Cap sample, the pollen assemblage is dominated, at 70%, by tree/shrub pollen (30% *Alnus*, 28% *Betula*, 4 % *Pinus*, 4% *Picea*). Only 5% of the grains could not be identified compared to 24% for the Camp Century sample. The basal ice of the Penny Ice Cap probably dates back to the end of the Eemian, when the ice cap started to build up again. Interestingly, in a pollen study on nearby lakes in the Cumberland Peninsula, Fréchette et al. (2006) also found very high concentrations and very high percentages of *Alnus* and *Betula* during the last interglacial. This suggests a low Arctic vegetation comparable to that which is presently found in southwest Greenland.

The pollen assemblage of the Camp Century basal ice is not easy to interpret. It contains a large percentage of unidentifiable pollen grains, suggesting the grains were exposed to air for some time before being incorporated in the ice or have been affected by stresses in the ice. Tree/shrub pollen and herb pollen (including *Salix*) are almost equally represented in the assemblage, at 34% and 31% respectively. The presence of *Betula*, *Salix*, Ericaceae, along with Poaceae and other herbaceous types could suggest a low to mid Arctic environment but *Betula* could also derive from long distance transport. However the near absence of *Alnus* (a prolific pollen producer) is surprising. The sample has a relatively high percentage of tree pollen (e.g. *Quercus*, *Ulmus*, *Fraxinus*) that are normally associated with temperate, broadleaf or mixed forests but these pollen types are often found in surface snow and ice core samples. Based on the data, it is difficult to determine if the pollen grains were blown to the ice surface or came from the sediment underneath the ice sheet.

**TABLE 1**.

| CAMP CENTURY BASAL ICE, SAMPLE CC1052A: Pollen analysis | | |
|---|---|---|
| | Split 1 | Split 2 |
| Taxa | 200 mL | 520 mL |
| Quercus | 3 | 1 |
| Ulmus | 0 | 2 |
| Fraxinus | 1 | 2 |
| Juglans? | 0 | 1 |
| Carpinus | 1 | 0 |
| Tsuga? | 0 | 1 |
| Larix? | 0 | 2 |
| Pinus | 8 | 3 |
| Pinus strobus | 1 | 0 |
| Picea | 3 | 0 |
| Alnus | 0 | 1 |
| Alnus crispa | 0 | 1 |
| Betula | 10 | 8 |
| Salix | 0 | 4 |
| Ambrosia | 3 | 0 |
| Iva | 1 | 0 |
| Caryophyllaceae | 2 | 0 |
| Cyperaceae | 2 | 0 |
| Dryas | 2 | 1 |
| Ericaceae | 3 | 2 |
| Oxyria | 1 | 1 |
| Papaver | 0 | 1 |
| Poaceae | 7 | 10 |
| Ranunculaceae | 1 | 2 |
| Asteraceae-Tub | 2 | 1 |
| Asteraceae-Lig | 1 | 0 |
| Brassicaceae | 1 | 0 |
| Equisetum | 0 | 2 |
| Trilete spore | 1 | 3 |
| Lycopodium lucidulum | 0 | 1 |
| Sphagnum | 1 | 0 |
| Unknown | 1 | 3 |
| Indeterminable | 19 | 17 |
| | | |
| Total | 75 | 71 |
| Concentration | 375 | 204 |
| | | |
| Undiff. spores | 39 | 47 |
| found by DNA | | |
| Found in decontaminated inner part | | |

**TABLE 2**.

POLLEN CONCENTRATION AND PERCENTAGES IN SNOW AND ICE CORE SAMPLES FROM GREENLAND AND THE CANADIAN ARCTIC

| Site | Sample | Pollen ∑ | Conc. grains / L | Percentages* Trees/ Shrubs | Herbs + *Salix* | References |
|---|---|---|---|---|---|---|
| **Greenland** | | | | | | |
| Dye 3 | pit samples | 232 | 15 | 60 | 33 | Bourgeois, 1990 |
| | ice core, 1936-1939 A.D. | 19 | 25 | 58 | 42 | Fredskild & Wagner, 1974 |
| GISP-2 | pit samples | 73 | 5 | 55 | 37 | Bourgeois, unpublished data |
| Camp Century | ice core, 1936-1938 A.D. | 60 | 96 | 43 | 53 | Fredskild & Wagner, 1974 |
| | ice core, 8800 B.P. | 4 | 21 | 25 | 75 | Fredskild & Wagner, 1974 |
| | ice core, 14 930 B.P. | 9 | 49 | 67 | 33 | Fredskild & Wagner, 1974 |
| | **ice core, basal ice (1052A)** | **71** | **204** | **31** | **34** | **This study** |
| **Canadian Arctic** | | | | | | |
| Penny-95 | ice core, basal ice | 115 | 125 | 70 | 19 | Bourgeois, unpublished data |
| Agassiz-79 | melt water**, basal ice | 777 | 14 | 49 | 24 | Koerner *et al.*, 1988 |
| Agassiz-87 | ice core, 1900-1950 A.D. | 104 | 15 | 64 | 25 | Bourgeois et al., 2000 |
| | bulk sample, basal ice-2 | 418 | 38 | 58 | 30 | Bourgeois, unpublished data |
| | bulk sample, basal ice-1 | 290 | 17 | 48 | 34 | Bourgeois, unpublished data |

* Percentages are based on the sum of tree and herb pollen, fern spores, Sphagnum, unknown and indeterminables grains.
** This sample, from the lowermost 3.5 m, was obtained by filtering the meltwater stored in the tank of a thermal drill.

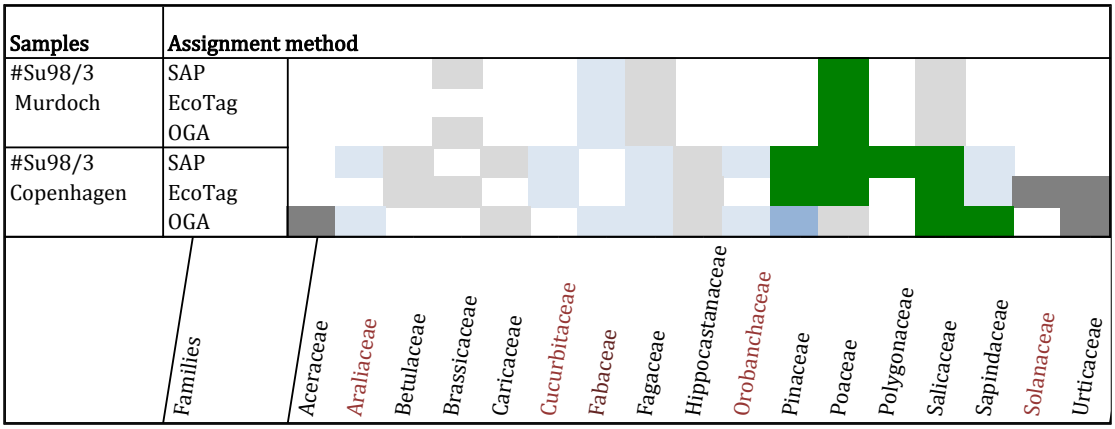
# Appendix D

**Table1**. Presence-absence raw data at the order level of assigned taxa as determined by each of the three assignment methods; SAP, EcoTag and OGA. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, blue by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample. Dark colours indicate where the taxa have been replicated within labs.

| Samples | Assignment method | Apiales | Brassicales | Caryophyllales | Coniferales | Cucurbitales | Fabales | Fagales | Malpighiales | Rosales | Poales | Sapindales | Solanales |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| #Su98/3 Murdoch | SAP | | | | | | | | | | | | |
| | EcoTag | | | | | | | | | | | | |
| | OGA | | | | | | | | | | | | |
| #Su98/3 Copenhagen | SAP | | | | | | | | | | | | |
| | EcoTag | | | | | | | | | | | | |
| | OGA | | | | | | | | | | | | |

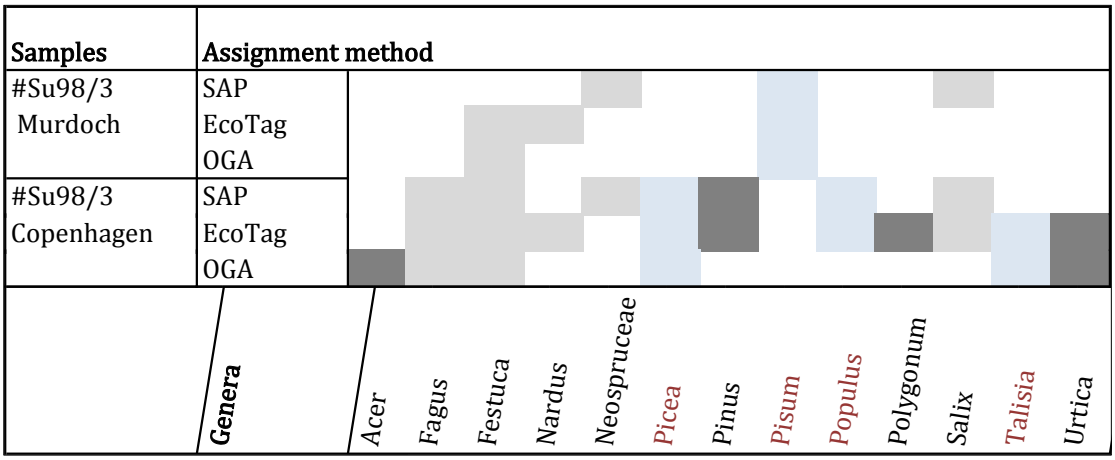
Red writing indicates which taxa have only been found rbcL and represents tropical unlikely assignments or contaminants from exotic food plants.

**Table 2**. Presence-absence raw data at the family level of assigned taxa as determined by each of the three assignment methods; SAP, EcoTag and OGA. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, blue by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample. Dark colours indicate where the taxa have been replicated within labs.




Red writing indicates which taxa have only been found rbcL and represents tropical unlikely assignments or contaminants from exotic food plants.

**Table 3**. Presence-absence raw data at the genus level of assigned taxa as determined by each of the three assignment methods; SAP, EcoTag and OGA. Colours indicate the markers that detected the taxon listed for the different samples and methods. Grey is detected by trnL, blue by rbcL and green is where both of the markers detected the same taxon, by the same method and within the same sample. Dark colours indicate where the taxa have been replicated within labs.




Red writing indicates which taxa have only been found rbcL and represents tropical unlikely assignments or contaminants from exotic food plants.

# Appendix E

Temperature history for the Dry Valleys, Antarctica based on data from Bintanja *et al*. (2005) and de Boer *et al*. (2010).

By Dorthe Dahl-Hensen, Centre for Ice and Climate, NBI, University of Copenhagen (2011)

**Figure 1**. Plot of the basal temperature during the last 2.5MY at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is based on the hypothesis that ice melted away during the MIS11 (420-360KY) and black line is following the current temperature models temperature profile.

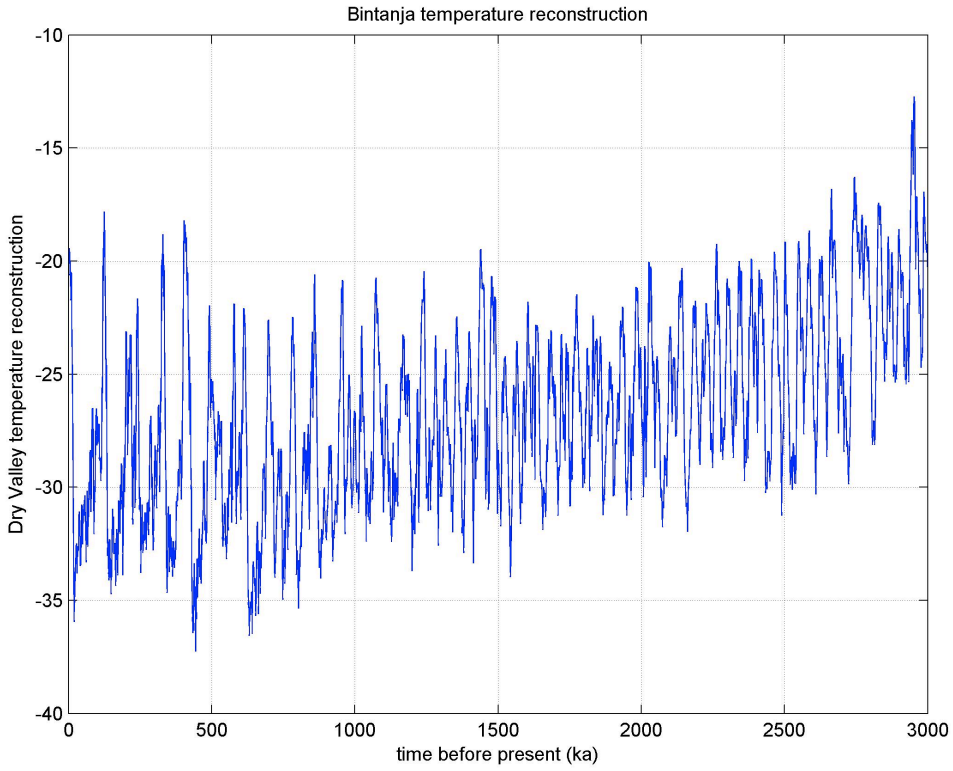
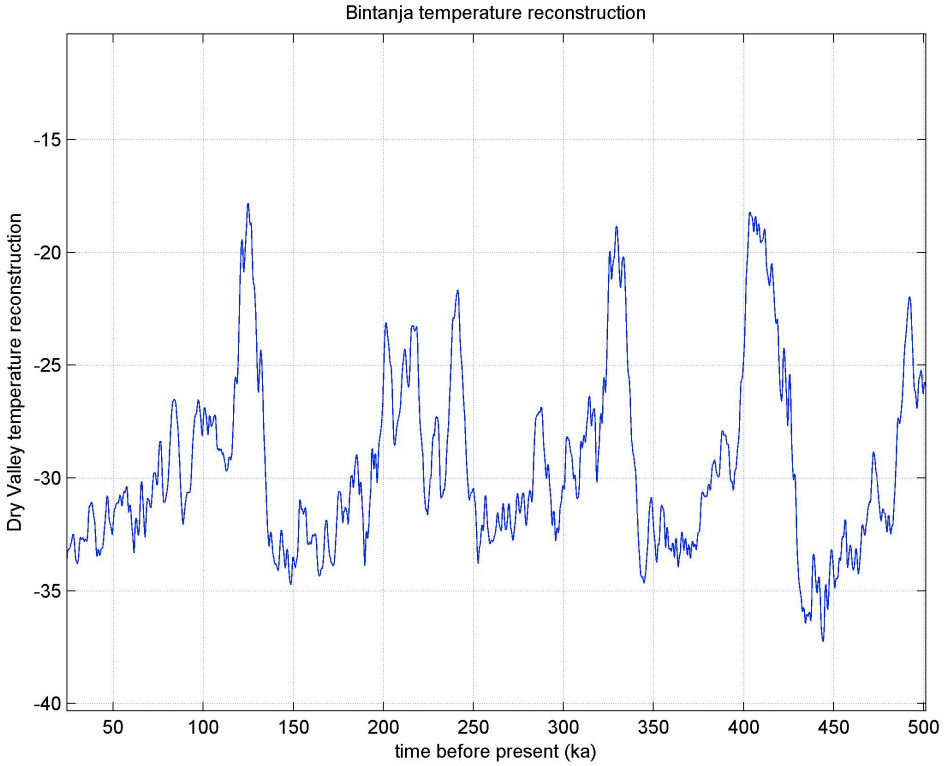


**Figure 2**. Plot of the temperature during the last 2.5MY at Camp Century when change in ice thickness and isostatic rebound of bedrock is implemented in the model. Red line is based on the hypothesis that ice melted away during the MIS11 (420-360KY) and black line is following the current temperature models temperature profile.

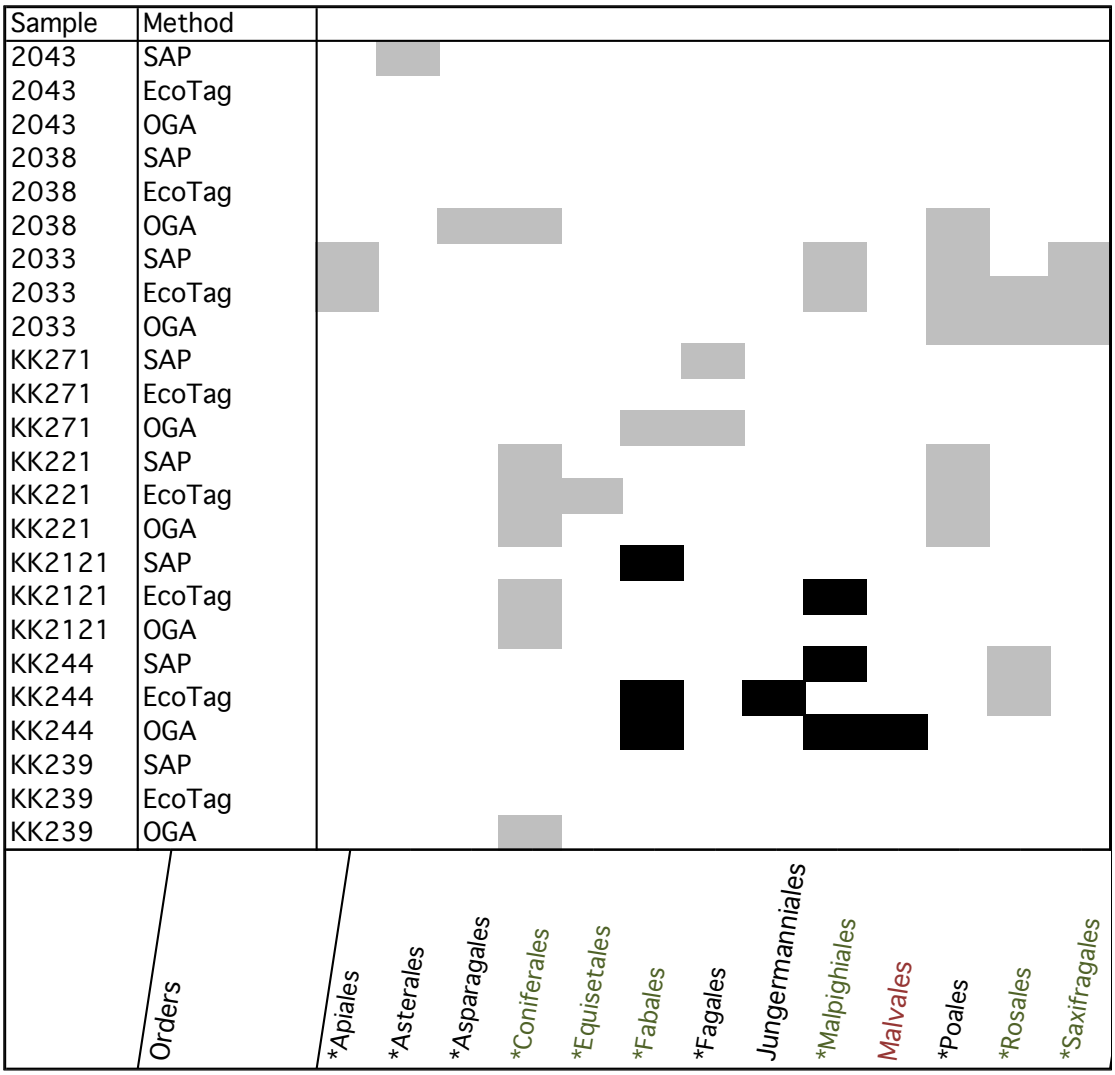



Bintanja R, van de Wal RSW & Oerlemans J (2005) Modelled atmospheric temperatures and global sea levels over the past million years. *Nature* **437**: 125-128.

de Boer B, van de Wal RSW, Bintanja R, Lourens LJ & Tuenter E (2010) Cenozoic global ice-volume and temperature simulations with 1-D ice-sheet models forced by benthic delta(18)O records. *Annals of glaciology* **51**: 23-33.

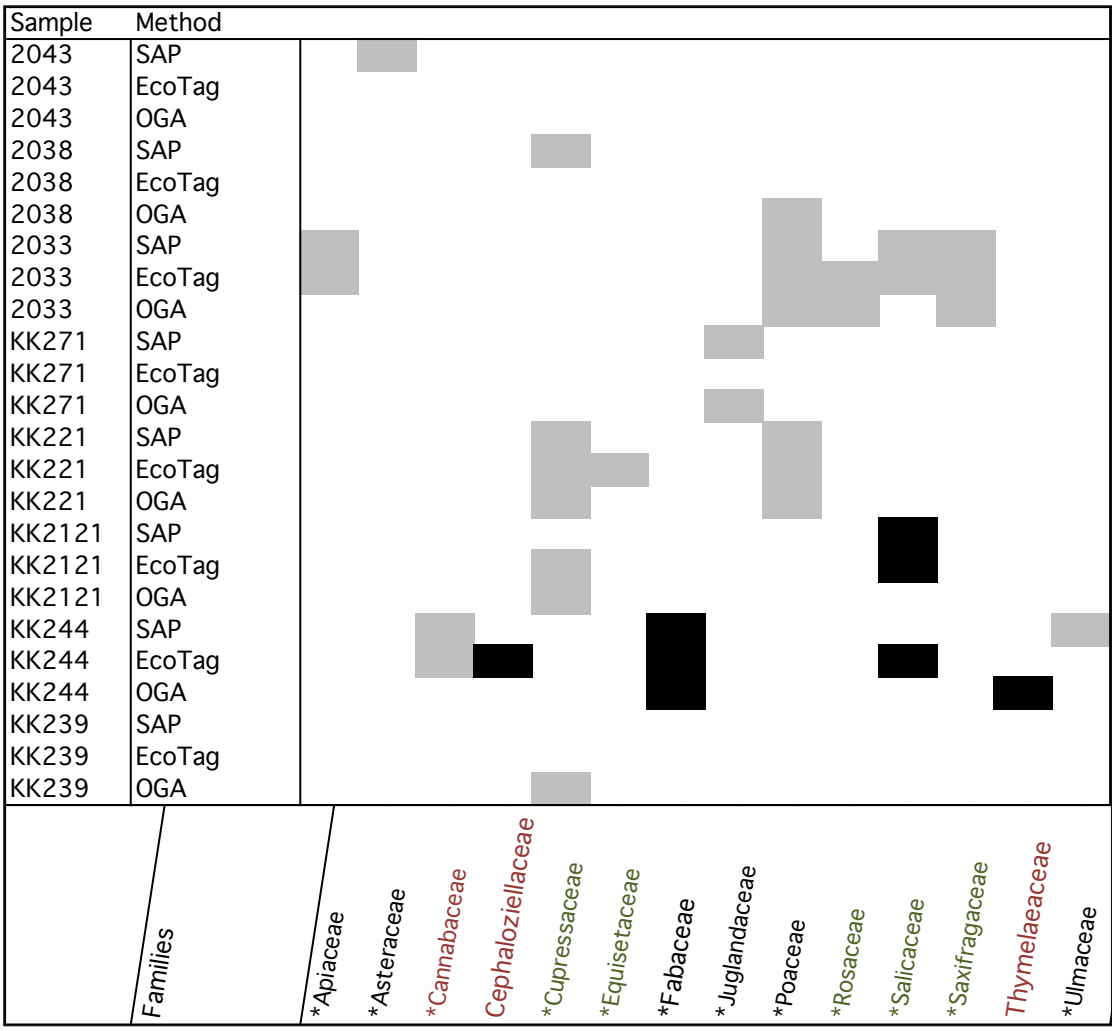
# Appendix F

**Table 1**. Presence-absence raw data at the order level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL and black by rbcL.



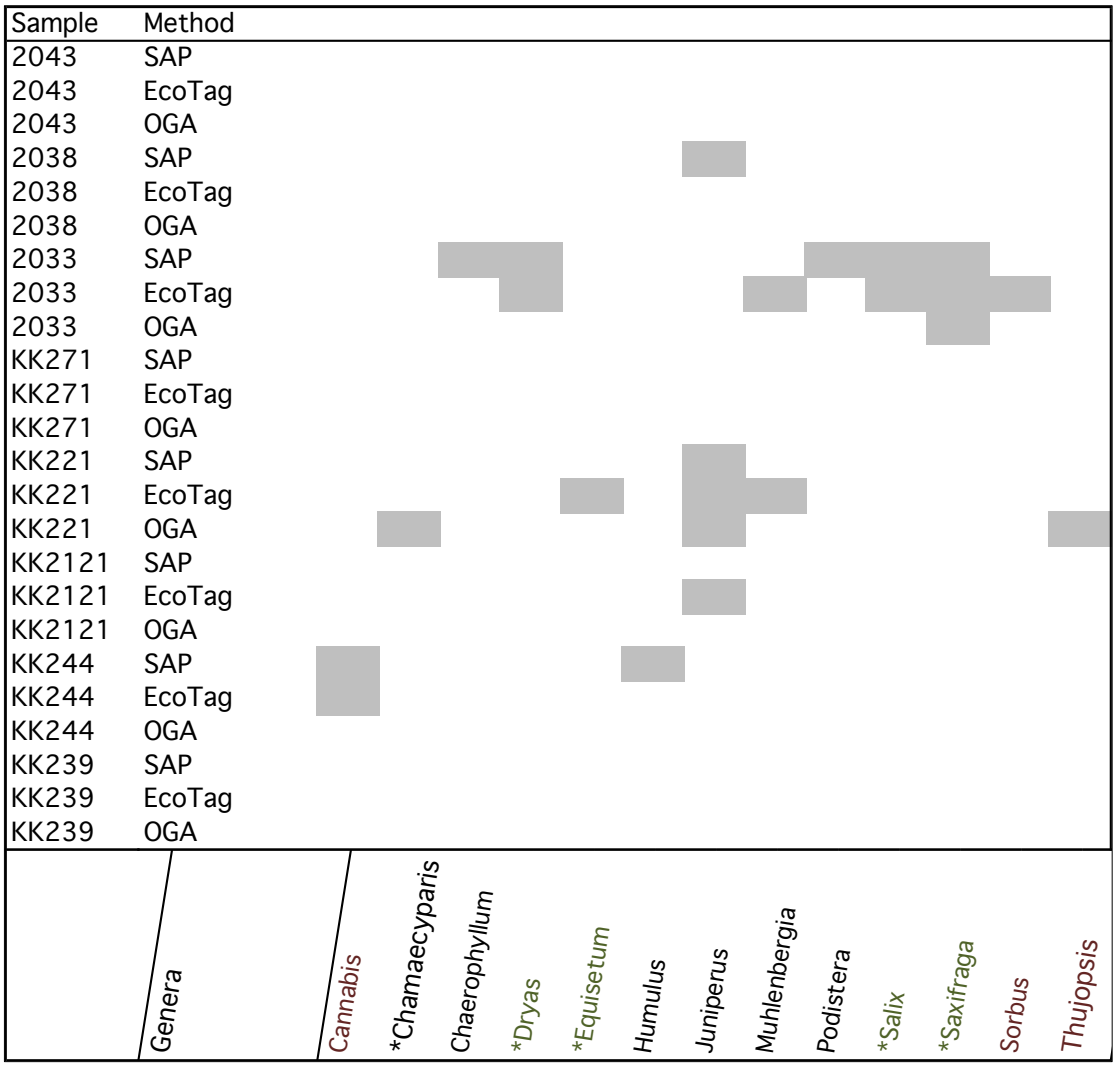

Green writing indicates the taxa that were found in the fossil record.
Black writing indicates the taxa that are unique found by the aDNA. Red writing indicate taxons likely to be mis-assignments or exotic food contaminants
* Illustrate that the taxon was also found in Camp Century.

**Table 2**. Presence-absence raw data at the family level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL and black by rbcL




Green writing indicates the taxa that were found in the fossil record. Black writing indicates the taxa that are unique found by the aDNA. Red writing indicate taxons likely to be mis-assignments or exotic food contaminants.
* Illustrate that the taxon was also found in Camp Century.

**Table 3**. Presence-absence raw data at the genus level of assigned taxa for each of the three assignment methods. Colours indicate the marker that detected the taxon listed for the different samples and methods. Grey is detected by trnL and black by rbcL




Green writing indicates the taxa that were found in the fossil record. Black writing indicates the taxa that are unique found by the aDNA. . Red writing indicate taxons likely to be mis-assignments or exotic food contaminants.

* Illustrate that the taxon was also found in Camp Century.