Dynamics of *Larix* (Mill.) species in Siberia during the last 50,000 years inferred from sedimentary ancient DNA

Luise Schulte

Univ.-Diss.

zur Erlangung des akademischen Grades

"doctor rerum naturalium"

(Dr. rer. nat.)

in der Wissenschaftsdisziplin „Ökologie“

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

Institut für Biochemie und Biologie

der Universität Potsdam

angefertigt am

Alfred-Wegener-Institut

Helmholtz-Zentrum für Polar- und Meeresforschung

Datum der Einreichung: 22. Dezember 2021

Datum der Disputation: 20. Juni 2022

Hauptbetreuerin: Prof. Dr. Ulrike Herzschuh
(Universität Potsdam)
Weitere Gutachter: Prof. Dr. Peter Heintzman
(The Arctic University Museum, Tromsø, Norway)
Dr. Natalya A. Rudaya
(Russian Academy of Sciences, Novosibirsk, Russia)

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-55878
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-558782

Summary

The deciduous needle tree larch (*Larix* Mill.) covers more than 80% of the Asian boreal forests. Only a few *Larix* species constitute the vast forests and these species differ markedly in their ecological traits, most importantly in their ability to grow on and stabilize underlying permafrost. The pronounced dominance of the summergreen larches makes the Asian boreal forests unique, as the rest of the northern hemisphere boreal forests is almost exclusively dominated by evergreen needle-leaf forests. Global warming is impacting the whole world but is especially pronounced in the arctic and boreal regions. Although adapted to extreme climatic conditions, larch forests are sensitive to varying climatic conditions. By their sheer size, changes in Asian larch forests as range shifts or changes in species composition and the resulting vegetation-climate feedbacks are of global relevance. It is however still uncertain if larch forests will persist under the ongoing warming climate or if they will be replaced by evergreen forests. It is therefore of great importance to understand how these ecosystems will react to future climate warmings and if they will maintain their dominance. One step in the better understanding of larch dynamics is to study how the vast dominant forests developed and why they only established in northern Asia. A second step is to study how the species reacted to past changes in the climate.

The first objective of this thesis was to review and identify factors promoting Asian larch dominance. I achieved this by synthesizing and comparing reported larch occurrences and influencing components on the northern hemisphere continents in the present and in the past. The second objective was to find a possibility to directly study past *Larix* populations in Siberia and specifically their genetic variation, enabling the study of geographic movements. For this, I established chloroplast enrichment by hybridization capture from sedimentary ancient DNA (*sed*aDNA) isolated from lake sediment records. The third objective was to use the established method to track past larch populations, their glacial refugia during the Last Glacial Maximum (LGM) around 21,000 years before present (ka BP), and their post-glacial migration patterns.

To study larch promoting factors, I compared the present state of larch species ranges, areas of dominance, their bioclimatic niches, and the distribution on different extents and thaw depths of permafrost. The species comparison showed that the bioclimatic niches greatly overlap between the American and Asian species and that it is only in the extremely continental climates in which only the Asian larch species can persist. I revealed that the area of dominance is strongly connected to permafrost extent but less linked to permafrost seasonal thaw depths. Comparisons of the paleorecord of larch between the continents suggest differences in the recolonization history. Outside of northern Asia and Alaska, glacial refugial populations of larch were confined to the southern regions and thus recolonization could only occur as migration from south to north. Alaskan larch populations could not establish wide-range dominant forest which could be related to their own genetically depletion as separated refugial population. In Asia, it is still unclear whether or not the northern refugial populations contributed and enhanced the postglacial colonization or whether they were replaced by populations invading from the south in the course of climate warming. Asian larch dominance is thus promoted partly by adaptions to extremely continental climates and by adaptations to grow on continuous permafrost but could be also connected to differences in glacial survival and recolonization history of *Larix* species.

Except for extremely rare macrofossil findings of fossilized cones, traditional methods to study past vegetation are not able to distinguish between larch species or populations. Within the scope of this thesis, I therefore established a method to retrieve genetic information of past larch populations to

distinguish between species. Using the *Larix* chloroplast genome as target, I successfully applied the method of DNA target enrichment by hybridization capture on *sed*aDNA samples from lake records and showed that it is able to distinguish between larch species. I then used the method on samples from lake records from across Siberia dating back up to 50 ka BP. The results allowed me to address the question of glacial survival and post-glacial recolonization mode in Siberian larch species. The analyzed pattern showed that LGM refugia were almost exclusively constituted by *L. gmelinii,* even in sites of current *L. sibirica* distribution. For included study sites, *L. sibirica* migrated into its extant northern distribution area only in the Holocene. Consequently, the post-glacial recolonization of *L. sibirica* was not enhanced by northern glacial refugia. In case of sites in extant distribution area of *L. gmelinii,* the absence of a genetic turn-over point to a continuous population rather than an invasion of southern refugia. The results suggest that climate has a strong influence on the distribution of *Larix* species and that species may also respond differently to future climate warming. Because species differ in their ecological characteristics, species distribution is also relevant with respect to further feedbacks between vegetation and climate.

With this thesis, I give an overview of present and past larch occurrences and evaluate which factors promote their dominance. Furthermore, I provide the tools to study past *Larix* species and give first important insights into the glacial history of *Larix* populations.

Deutsche Zusammenfassung

Der sommergrüne Nadelbaum Lärche (*Larix* Mill.) bedeckt mehr als 80 % der Fläche der borealen Wälder Asiens. Nur wenige Lärchenarten bilden ausgedehnte Wälder und diese Arten unterscheiden sich deutlich in ihren ökologischen Eigenschaften, vor allem in ihrer Fähigkeit, auf Permafrost zu wachsen und diesen zu stabilisieren. Die ausgeprägte Dominanz der sommergrünen Lärchen macht die asiatischen borealen Wälder einzigartig, da der Rest der borealen Wälder der Nordhalbkugel fast ausschließlich von immergrünen Nadelwäldern dominiert wird. Die Klimaerwärmung wirkt sich auf die ganze Welt aus, ist aber in den arktischen und borealen Regionen besonders ausgeprägt. Obwohl die Lärchenwälder an extreme klimatische Bedingungen angepasst sind, reagieren sie empfindlich auf klimatische Schwankungen. Aufgrund ihrer schieren Größe sind Veränderungen in asiatischen Lärchenwäldern, wie z. B. Verschiebungen des Verbreitungsgebiets oder Veränderungen in der Artenzusammensetzung und die daraus resultierenden Rückkopplungen zwischen Vegetation und Klima, von globaler Bedeutung. Es ist jedoch noch ungewiss, ob die Lärchenwälder unter der fortschreitenden Klimaerwärmung bestehen bleiben oder durch immergrüne Wälder ersetzt werden. Es ist daher von großer Bedeutung zu verstehen, wie diese Ökosysteme auf die künftige Klimaerwärmung reagieren werden und ob sie ihre Dominanz behalten werden. Ein Schritt zum besseren Verständnis der Lärchendynamik besteht darin, zu untersuchen, wie die riesigen dominanten Wälder von heute entstanden sind und warum sie sich nur in Nordasien etabliert haben. In einem zweiten Schritt soll untersucht werden, wie die Art auf vergangene Klimaveränderungen reagiert hat.

Das erste Ziel dieser Arbeit bestand darin, die Faktoren zu ermitteln, die die Dominanz der asiatischen Lärche begünstigen. Dies erreichte ich, indem ich die dokumentierten Lärchenvorkommen und die sie beeinflussenden Komponenten auf den Kontinenten der nördlichen Hemisphäre in der Gegenwart und in der Vergangenheit gesammelt und verglichen habe. Das zweite Ziel bestand darin, eine Möglichkeit zu finden, frühere Lärchenpopulationen in Sibirien und insbesondere ihre genetische Variation direkt zu studieren, um geografische Bewegungen untersuchen zu können. Dafür etablierte ich die Methode der Anreicherung von Chloroplasten durch Hybridisierung von alter sedimentärer DNA (*sed*aDNA) isoliert aus Seesedimenten. Das dritte Ziel bestand darin, die etablierte Methode zu nutzen, um vergangene Lärchenpopulationen, ihre eiszeitlichen Refugien während des letzten glazialen Maximums (LGM) um ca. 21.000 Jahre vor der Gegenwart (ka BP) und ihre nacheiszeitlichen Migrationsmuster zu verfolgen.

Um die Faktoren zu untersuchen, die die Ausbreitung der Lärche begünstigen, verglich ich den gegenwärtigen Stand der Verbreitungsgebiete der Lärchenarten, die Gebiete, in denen sie vorherrschen, ihre bioklimatischen Nischen und die Verteilung auf verschiedene Ausdehnungen und Auftautiefen des Permafrosts. Der Artenvergleich zeigte, dass sich die bioklimatischen Nischen der amerikanischen und asiatischen Arten stark überschneiden und dass nur in den extrem kontinentalen Klimazonen ausschließlich die asiatischen Lärchenarten überleben können. Es zeigte sich, dass das Verbreitungsgebiet stark mit der Permafrostausdehnung zusammenhängt, aber weniger mit der saisonalen Auftautiefe des Permafrosts. Der Vergleich vergangener Lärchenvorkommen zwischen den Kontinenten deutet auf Unterschiede in der Rekolonisationsgeschichte hin. Außerhalb Nordasiens und Alaskas waren die eiszeitlichen Lärchenpopulationen auf die südlichen Regionen beschränkt, so dass die Wiederbesiedlung nur als Wanderung von Süden nach Norden erfolgen konnte. Die Lärchenpopulationen in Alaska konnten keinen weiträumig dominanten Wald etablieren, was mit ihrer eigenen genetischen Verarmung als abgeschiedene Refugialpopulation zusammenhängen könnte. In Asien ist noch unklar, ob die nördlichen Refugialpopulationen zur nacheiszeitlichen Besiedlung beigetragen und diese verstärkt haben oder ob sie im Zuge der Klimaerwärmung durch von Süden eindringende Populationen ersetzt wurden. Die Dominanz der asiatischen Lärche wird also zum Teil

durch Anpassungen an das extrem kontinentale Klima und durch Anpassungen an das Wachstum auf kontinuierlichem Permafrost begünstigt, könnte aber auch mit Unterschieden in der glazialen Überlebens- und Rekolonisationsgeschichte der *Larix*-Arten zusammenhängen.

Abgesehen von den äußerst seltenen Makrofossilienfunden versteinerter Zapfen sind die herkömmlichen Methoden zur Untersuchung der vergangenen Vegetation nicht in der Lage, zwischen Lärchenarten oder -populationen zu unterscheiden. Im Rahmen dieser Arbeit habe ich daher eine Methode zur Gewinnung genetischer Informationen früherer Lärchenpopulationen entwickelt, um zwischen den Arten zu unterscheiden. Unter Verwendung des *Larix*-Chloroplastengenoms habe ich die Methode der DNA-Anreicherung durch Hybridisierung erfolgreich auf *sed*aDNA-Proben aus See-sedimentbohrkernen angewandt und gezeigt, dass die Methode erlaubt zwischen Lärchenarten zu unterscheiden. Anschließend wendete ich die Methode auf Proben aus Seen in ganz Sibirien an, die bis zu 50 ka BP zurückreichen. Anhand der Ergebnisse konnte ich zur Beantwortung der Frage beitragen, welche sibirische Lärchenarten während des LGM überlebten und wie die postglaziale Wiederbesiedlung stattfand.

Das analysierte Muster zeigte, dass die LGM-Refugien fast ausschließlich von *L. gmelinii* gebildet wurden, selbst an Orten, an denen heute *L. sibirica* verbreitet ist. In den untersuchten Gebieten ist *L. sibirica* erst im Holozän in ihr heutiges nördliches Verbreitungsgebiet eingewandert. Folglich wurde die nacheiszeitliche Wiederbesiedlung von *L. sibirica* nicht durch nördliche eiszeitliche Refugien gefördert. Im Falle der Standorte im heutigen Verbreitungsgebiet von *L. gmelinii* deutet das Fehlen eines Wechsels genetischer Variation eher auf eine kontinuierliche Population als auf eine Invasion aus südlichen Refugien hin. Die Ergebnisse deuten darauf hin, dass das Klima einen starken Einfluss auf die Verbreitung von *Larix*-Arten hat und die Arten auch auf zukünftige Klimaerwärmung unterschiedlich reagieren könnten. Da die Arten sich in ihren ökologischen Eigenschaften unterscheiden, ist eine Änderung in der Verbreitung der Arten auch im Hinblick auf weitere Rückkopplungen zwischen Vegetation und Klima relevant.

In dieser Arbeit gebe ich einen Überblick über die heutigen und früheren Lärchenvorkommen und bewerte, welche Faktoren ihre Dominanz begünstigen. Darüber hinaus stelle ich eine Methode zur Untersuchung vergangener Lärchenarten bereit und gebe erste wichtige Einblicke in ihre glaziale Geschichte.

Table of Contents

1 Introduction

1.1 *Larix* forests in a changing climate

Climate warming affects the whole world, but is especially pronounced in the higher-latitudinal regions due to polar amplification (Miller et al., 2010). As such, the surface temperature in the Arctic has already increased at a rate of 50% greater than the average of the whole Northern Hemisphere (Meyer et al., 2015). Furthermore, by the end of this century a warming of 4 to 11°C, which is double the global average rate, is projected for boreal forest regions (IPCC, 2021; Scheffer et al., 2012). Boreal forests make up 30% of the global forest area and are the largest terrestrial ecosystem (Gauthier et al., 2015). Approximately 20% of the global boreal forests are dominated by the summergreen needle tree larch (*Larix* Mill.) (Osawa & Zyryanova, 2010). The majority of this ecosystem is situated in northern Eurasia on areas of continuous permafrost (Osawa et al., 2010b) which are strongly affected by global warming (Biskaborn et al., 2019).

Larch forests play an important role in regional and global climate patterns, and exert control on a variety of climate feedbacks such as carbon stocks (Kajimoto et al., 2010), or permafrost stability (Zhang et al., 2011), and have therefore been identified as a possible tipping element in the climate system (Seddon et al., 2016). Considering the large extent of the Siberian larch forests, comparatively little research has so far studied the ecosystem. Major questions still remain open, including how larch has become dominant in Siberia when everywhere else in the boreal forests, evergreens are predominant. Furthermore, studying the response of larch forests to changing climatic conditions is crucial to understand the potential global impacts of ecosystem changes. First of all, it is important to understand how the species has reacted to past climate changes, including their survival during glacial periods and their success in postglacial recolonization.

Here, I study factors promoting Asian larch dominance and study Asian larch species reacted to climate changes in the past 50,000 years using sedimentary ancient DNA (*sed*aDNA).

1.2 The genus *Larix*

Larch (*Larix* Mill.) is a summergreen needle-leaf tree in the family of Pinaceae and the only boreal conifer which is not evergreen (Taggart & Cross, 2009). Larch species grow best on well-drained nutrient rich soils but are more commonly found on nutrient-poor, often shallow or poorly drained soils where competition is less strong (Johnston, 1990; Neale & Wheeler, 2019; Osawa et al., 2010b). They efficiently use nitrogen and carbon and are a fast growing pioneer species on disturbed soils (Mamet et al., 2019). All species of the genus are very light-demanding and some are unable to become established in their own shade (Abaimov, 2010; Johnston, 1990). Their shade-intolerance typically hinders their competitive ability in late ecological succession which leads to their replacement by more shade-tolerant tree species, often evergreen conifers as *Picea* or *Pinus* (Warren et al., 2016). The deciduous habit of larch avoids winter desiccation and cold damage of the foliage (Gower & Richards, 1990), also less snow is collected on their branches making them less prone to wind breaks (Givnish, 2002). Larch stands often form altitudinal and latitudinal tree lines (Mamet et al., 2019) and constitute the northernmost forests at 72°N (Abaimov et al., 1999). Further, several species are considered as valuable timber trees for their original properties including density, mechanical properties and natural durability (Pâques et al., 2013).

The number of species in the genus *Larix* is still under discussion (Abaimov, 2010). The described taxonomic units are often not separated by geographic or reproductive barriers, and easily hybridize in nature (Polezhaeva 2010, Semerikov 2003). In Russia, most authors acknowledge three main boreal species: *L. sibirica* Ledeb., *L. gmelinii* (Rupr.) Rupr. and *L. cajanderi* Mayr (Abaimov, 2010). Some authors split *L. sibirica* into two species, *L. sibirica* to the east and *L. sukaczewii* Dylis to the west (Dylis, 1981), whereas *L. gmelinii* and *L. cajanderi* are often summarized under one species sometimes referred to as *L. gmelinii* (Rupr.) Kuzen. and sometimes as *L. dahurica* Turcz.ex Trautv (Geburek, 2014). The number of mountainous larch species at the southern distribution border in Eurasia is even more controversial (Guo et al., 2021). In Europe, only one larch species is described, the European larch, *L. decidua* (Mill) (Da Ronch et al., 2016). In North America, three species have their natural habitat. *L. laricina* (Du Roi) K. Koch., also known as tamarack, has a wide distribution and often forms the northern tree limit (Johnston, 1990). The other species, *L. occidentalis* Nutt. and *L. lyallii* Parl. are confined to small mountain habitats of North-Western USA and Canada (Betts, 1939; Carlson, 1998).

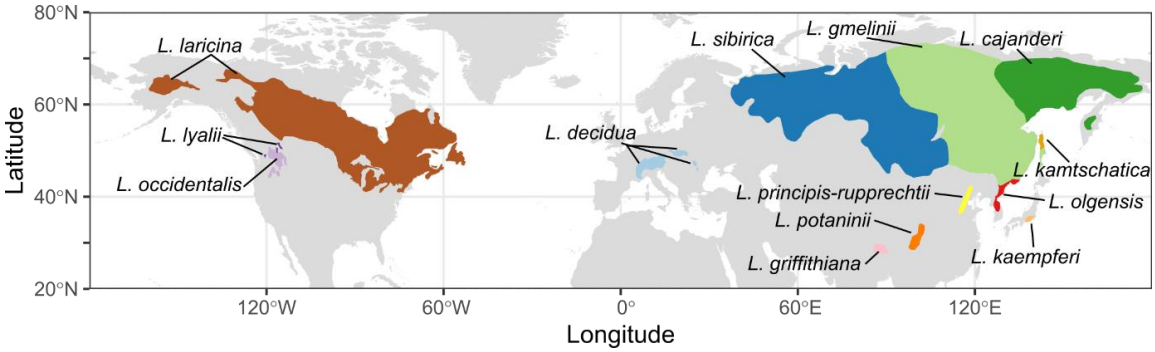

Fig. 1: Present-day natural distribution of *Larix* species (after Mamet et al., 2019)

The genus *Larix* is evolutionarily young. Its emergence is estimated to the late Cretaceous (Lepage, 2003; LePage & Basinger, 1995) and can be traced to high latitudes in both North America and north-eastern Eurasia (Taggart & Cross, 2009). Two distinct morphologies occur at the female cones of larches, with one group having long exserted bracts and the other short nonemergent bracts (Patschke, 1913). As LePage and Basinger (1995) observed, the group of short-bracted larches comprises wide spread boreal species which occur also at low altitudes (including American *L. laricina,* European *L. decidua,* and Asian *L. sibirica, L. gmelinii* and *L. cajanderi)*. On the other hand, long-bracted species are restricted to montane and subalpine regions (e.g. American *L. occidentalis* and *L. lyallii* and Asian *L. potanii, L. griffithiana*). However, genetic analysis revealed that *Larix* species are phylogenetically split into the Eurasian and the North American clade (Semerikov & Lascoux, 1999; Wei & Wang, 2003) with the two groups separating probably at the end of the Pliocene (Semerikov et al., 2003). Bract morphology thus is not defining a natural group but reflects more ancient history and, considering the different habitats of the morphologies, could be related to fitness (Semerikov & Lascoux, 1999).

1.3 *Larix* distribution in the world and their dominance in northern Asia

Siberian larch forests are unique in the world. Although species of the genus *Larix* are present on all northern hemisphere continents, their distribution and the degree to which they form the dominant tree species is completely different on each continent (Fig. 2). In Europe, the only *Larix* species, *L. decidua* Mill., is exclusively native in the central European mountain ranges, mostly as part of the subalpine forest belt or forming the tree limit, and absent from Scandinavia (Da Ronch et al., 2016; Pâques et al., 2013). In North America, three *Larix* species are native, but only *L. laricina* (Du Roi) K. Koch has a wide range distribution extending from the Atlantic to central

Alaska (Johnston, 1990). Despite of this wide range it rarely forms extensive pure stands and populations are typically small and occur on wet, poorly drained sites (Cheliak et al., 1988). Both in Europe and America, evergreen needle-leaf species are the dominant boreal forest forming species. It is only in northern Asia, where larches form extensive forests over vast areas as the only or dominant tree forming species (Abaimov et al., 1999). The reason for this tremendous difference in the distribution of larch forests on the northern hemisphere is still not well understood. Most research focus on a single *Larix* species or the *Larix* species on one continent. Wagner et al. (2015) conducted a thorough synthesis on the paleohistory of European larch, Warren et al. (2016) and Napier et al. (2020) tracked the post-glacial dynamics of the *L. laricina* in North America and Alaska, and with the book "Permafrost ecosystems – Siberian larch forests" (Osawa et al., 2010a) we have a great compendium of knowledge about the Asian larch species. However, few studies were undertaken that cover *Larix* across the northern hemisphere. Mamet et al. (2019) conducted a meta-study on *Larix* distribution shifts during the last 80 years and recently U. Herzschuh (2020) compared climate spaces of evergreen and summergreen needle-leaf trees between northern America and Asia. Although research on *Larix'* genetic diversity, population dynamics, glacial history and biogeography has been undertaken on all continents (e.g. Napier et al., 2020; Polezhaeva et al., 2010; Semerikov et al., 2013; Wagner et al., 2015; Warren et al., 2016) a synthesis of the available data on current distribution pattern and bioclimatic niches as well as past *Larix* occurrences was hitherto lacking.

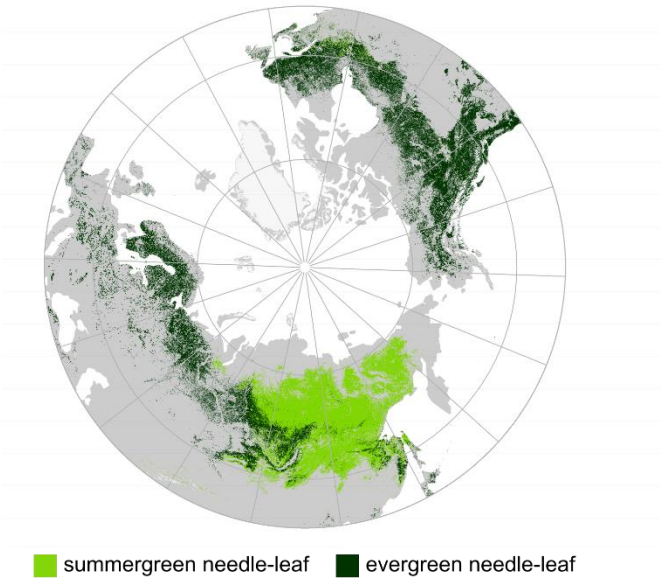

Fig. 2 Distribution of evergreen and summergreen needle-leaf forest on the northern hemisphere (data source: ESA, 2017; after Herzschuh, 2020)

In order to better understand factors leading to the current distribution of larches, also in regard to possible future changes in a warming climate, two things are highly desirable: firstly, a review and comparison of larch species across the continents to pinpoint the factors leading to larch dominance in Asia and secondly, directly studying past migration dynamics of larches to track their climatic responses. Especially the second objective is challenging as current methods are very limited in their ability to study past population dynamics as is explained in the next section.

1.4 Methods to study past species dynamics

1.4.1 Modern genetic marker studies

One way to study past population dynamics are genetic studies on individuals from modern populations. Numerous statistical models and tools help to decipher population structure and dynamics (Tsetsos et al., 2018). The analysis of genetic diversity patterns in tree populations across Europe revealed patterns of possible glacial refugia and postglacial migration patterns (Petit et al., 2003). Advances in sequencing techniques and the development of reduced representation methods which allow for the analysis of up to millions of marker positions at relatively low costs have drastically advanced studies of population genetics (Parchman et al., 2018). However, few studies were undertaken on Siberian larch populations, with a limited number of markers (Araki et al., 2008; Khatab et al., 2008; Polezhaeva et al., 2010; Semerikov et al., 2003, 2013). Inferring past population dynamics from modern genetic material is limited in that signals from events at different times may overlap and assumptions must be made about generation time and mutation rate (Semerikov et al., 2013). One way to circumvent these problems is to directly study ancient material which allows giving a time stamp to the occurrence of a taxon on a specific site.

1.4.2 Lake sediments as archives of the past

Lakes often exist over long periods of time and over the course of their existence sediments are deposited on their grounds. Lake sediments are highly influenced by their surrounding environment and store information on biological and geochemical processes outside of the lake (Sturm & Lotter, 1995). Layered lake sediments thus provide archives of the past which can be both highly resolved in time and long in duration, a combination which is rarely met in other records (Cohen 2003). With the help of radiometric dating using ^{14}C isotopes as well as other methods, the age of the sediment layers can be determined either through macrofossil remains or bulk sediment (Oswald et al., 2005). Within the lake sediments, plant remains from the surroundings of the lake are incorporated which can be used to study past vegetation composition, taxon abundance and their changes throughout the time (e.g. Andreev et al., 2004; Binney et al., 2009; Brubaker et al., 2005; Müller et al., 2010).

1.4.3 Pollen and macrofossils

As traditional proxies for vegetation in lake sediments, pollen, spores, and macrofossil plant remains are used (Birks, 1981; Van Geel, 2002). Pollen and spores contain high amounts of the resistant polymer sporopollenin and are therefore well preserved in the sediments (Parducci et al., 2017). Pollen and macrofossil records are a well-established proxy to reconstruct past vegetation composition and many fossil pollen records exist worldwide. Nonetheless, the vast and remote area of northern Asia is still underrepresented (Li et al., 2022). A major challenge in pollen analysis is the variation in pollen productivity for different plant taxa. Especially in the case of *Larix,* pollen productivity is extremely low and conservation poor (Niemeyer et al., 2015). The already featureless morphology is often damaged and folded, making it confusable with various algal, fungal, and animal non-pollen palynomorphs (de Klerk et al., 2014). This results in rare findings of *Larix* pollen in sediments and a possible misclassification of larch taiga as treeless tundra (Bigelow et al., 2003). Another important limitation is that *Larix* pollen does not allow for a taxonomic resolution to species level. This is only achievable by rare findings of fossilized cones (Kullmann, 1998). However, several studies using pollen and macrofossil records found hints towards glacial refugia of *Larix* in high northern latitudes (Binney et al., 2009; Brubaker et al., 2005; Lozhkin et al., 2018; MacDonald et al., 2008; Müller et al., 2010).

1.4.4 Metabarcoding of sedimentary ancient DNA

Lake sediments provide often good probability of DNA conservation, as the water column can become thermally stratified and the 4°C water at the lake bottom favors the development of anoxia (Parducci et al., 2017). Extracellular DNA from lysed cells can bind to charged minerogenic or organic particles and thus get protected from further degradation (Pietramellara et al., 2009). This environemtal DNA stored in lake sediments can be used to study past vegetation (Parducci et al., 2017).

A method termed "metabarcoding" uses a short, highly variable part of the genome flanked by very conserved regions which can robustly serve as primer binding sites in a PCR reaction (Taberlet et al., 2007). By comparing the amplified variable fragments with databases, sequences can be assigned to family, genus or even species level. Taberlet et al. (2007) developed metabarcoding primers for plant taxon detection based on the intron of the *trn*L gene on the chloroplast genome. The short version of it amplifies the P6 loop of the *trn*L, which is only 10-143 base pairs (bp) long and thus highly suitable for the use with fragmented ancient DNA. Multiple studies used the method to reconstruct ancient vegetation in lake sediments and permafrost (Clarke et al., 2019; Jørgensen et al., 2012; Pansu et al., 2015; Sonstebo et al., 2010; Willerslev et al., 2014; Zimmermann et al., 2017). It has been shown that metabarcoding often identifies more species at higher taxonomic resolution than pollen records (Liu et al., 2020; Niemeyer et al., 2015) and that the DNA signal is highly local compared to the more regional pollen signal (Alsos et al., 2018; Parducci et al., 2017). However, rare taxa are often not identified by metabarcoding (Alsos et al., 2018; Shirazi et al., 2021) and *Larix* cannot be distinguished up to species level. The method is also limited by the DNA fragment length (Pedersen et al., 2016) as amplification is not possible if not both primer binding sites are intact, and primers alone make up for 39 bp, not including the amplified insert (for *Larix* the insert size is 31 bp) (Taberlet et al., 2007).

1.4.5 Metagenomic shotgun sequencing

Instead of PCR-amplifying only a short marker, it is also possible to sequence a fraction of the entire pool of extracted DNA, a method termed "metagenomic shotgun sequencing". The sequencing result is not limited to a specific marker and taxon group, but gives a snapshot of the entire biome that lived in and around the lake, including terrestrial and aquatic plants, but also bacteria, archaea, fungi, and mammals (Pedersen et al., 2016). The method is not per se limited by molecule length, but minimum length is set during the bioinformatic analysis of taxonomic assignment, this is often 30 bp (Ahmed et al., 2018; Pedersen et al., 2016). The typical damage patterns occurring in the process of DNA degradation, including deamination and DNA fragmentation, can be used to authenticate ancient DNA (Ginolhac et al., 2011). Up to date, only few studies used metagenomic shotgun sequencing on lake *sed*aDNA samples (Ahmed et al., 2018; Graham et al., 2016; Pedersen et al., 2016; Wang et al., 2017). A major disadvantage of the method is that only a small fraction of reads is represented in taxonomic sequence databases, for example Pedersen et al. (2016) taxonomically assigned less than 0.4% of the quality filtered reads. The majority of the assigned reads are microorganisms such as Bacteria and Archaea and the proportion of eukaryotic reads is typically low (Ahmed et al., 2018). Thus, an enormous sequencing effort is necessary to retrieve sequences of a specific taxon such as *Larix.*

1.4.6 Target enrichment by hybridization capture

A way to overcome the limitations of shotgun sequencing is to specifically enrich parts of a DNA sample prior to sequencing using hybridization capture (Mamanova et al., 2010). The first step of this method is to generate DNA fragments of a target genome region of the species of interest either by synthesis or by PCR amplification of the DNA extract of a modern individual (Maricic et

al., 2010). A DNA sequence adapter is ligated to the fragments to either immobilize the then so called “bait” to a glass surface (array-capture, Hodges et al., 2007) or to link it to tiny magnetic beads, which can be later immobilized by attraction of a magnet (in-solution capture, Maricic et al., 2010). The baits are mixed together with the aDNA sample and can bind where the sequence is complementary. Bound ancient DNA is immobilized together with the bait, whereas non-binding DNA is washed away. The remaining ancient sample contains much less DNA, but is enriched for the target region. The technique was originally developed for modern DNA (Mamanova et al., 2010), but has been also successfully applied in ancient DNA studies, particularly on samples from mammals, enriching their mitochondrial DNA (Carpenter et al., 2013; Dabney et al., 2013; Enk et al., 2016). Ancient plant DNA has been successfully enriched of ancient gourd rinds (Kistler et al., 2014) and subfossil needles collected in lake sediments (Schmid et al., 2017). Only few successful capture enrichments of *sed*aDNA have been reported so far. Slon et al. (2017) applied the technique on cave sediments and Murchie et al. (2020) on permafrost samples for mitochondrial DNA. An attempt to enrich mitochondrial DNA out of lake sediments failed (Moore et al., 2019).

1.5 Thesis Objectives

The main goal of this thesis was to contribute to the answer of why larch forests are dominant in wide ranges of the boreal forests of northern Asia and how larch recolonized northern Asia after the last glacial. The first objective was to review and synthesize known information of extant and past larch occurrences and to identify the main factors leading to differences in larch distribution between northern Asia, Europe, and America. The second objective was to establish a suitable method to trace *Larix* species back in time. For this, the method of target enrichment via hybridization capture was tested using the complete chloroplast genome of larch as target. The third objective was to use the established method to study *Larix* species history and the drivers of their distribution since the last glacial over a wide range in Siberia.

1.6 Thesis outline & author contributions

This thesis is designed as a cumulative dissertation and consists out of seven chapters. The first chapter gives a general introduction and scientific background to the topic. Chapters 2, 3 and 4 comprise scientific manuscripts which have been published or submitted to peer-reviewed scientific journals. An overall discussion of the thesis outcomes is presented in chapter 5. References used in the introduction and discussion chapter are given in chapter 6. Additional information to the manuscripts is given in the appendix (chapter 7). Figures and tables are numbered in each chapter separately.

Table 1 Overview on the three manuscripts included in this thesis.

Manuscript I	**Forest-permafrost feedback loops and glacial refugia are clues to explain the unequal distribution of *Larix* (Mill.) across the northern hemisphere**
Authors	Luise Schulte, Chenzi Li, Simeon Lisovski, Ulrike Herzschuh
Status	*Submitted* to *Journal of Biogeography* **update after submission of thesis**: revised version accepted for publication

Summary	In this manuscript, we comparatively analyze *Larix* distribution in northern Asia, Europe and America using a range of available datasets. These included *Larix* species distribution ranges, areas of larch as the dominant landcover type, climate data, permafrost extent and permafrost active layer depth. In addition, we synthesize all published paleo-datasets for *Larix* in the northern hemisphere, i.e. fossil pollen, macrofossils and ancient DNA metabarcoding data. Our analysis, in conjunction with a detailed literature review, showed that Asian larch dominance can only partly be attributed to their adaptation to extremely continental climates. To a great extent larch dominance is related to their adaptation and positive feedback mechanisms with permafrost extent and frequent wild fires. High latitude glacial refugia might have played an important role in conveying dominance after the last glacial.
Author contribution	LS did literature research to collect published macrofossil databases and ancient DNA metabarcoding datasets. CL collected and harmonized published pollen datasets. SL helped with data analyses of the permafrost datasets. LS did all other analyses of the study, prepared the figures and wrote the manuscript under the supervision of UH. All authors commented on the manuscript.

Manuscript II	**Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest**
Authors	Luise Schulte, Nadine Bernhardt, Kathleen Stoof- Leichsenring, Heike H. Zimmermann, Luidmila A. Pestryakova, Laura S. Epp, Ulrike Herzschuh
Status	*Published* in *Molecular Ecology Resources, 21*(3), 801-815
Summary	In this manuscript we establish the method of target enrichment via hybridization capture for ancient DNA extracted from lake sediment. As the target, we use the complete chloroplast genome of *L. gmelinii*. We used shotgun sequencing on the same samples to test the effectiveness of the enrichment. The results show that hybridization capture highly suitable for target enrichment. The enriched sequences allowed us to further distinguish between different *Larix* species. Therefore, the method is suitable to trace species history of *Larix* in time.
Author contributions	LS newly established the methods of library preparation for shotgun sequencing and hybridization capture in the laboratory. This included the selection and adaptation of laboratory protocols including the purchase of new laboratory equipment and chemicals. LS did all laboratory work except for the production of long-range PCR products for the hybridization probe set, which was done by LSE and Nick Mewes. LS performed all analyses, prepared the figures and wrote the manuscript under the supervision of UH and LSE. KS, NB and HZ advised in laboratory issues. LAP and UH organized the field work on which samples the study is based on. All authors commented on the manuscript.

Manuscript III	**Dynamics of larch species in Siberia since the Last Glacial captured from sedimentary ancient DNA**
Authors	Luise Schulte, Stefano Meucci, Kathleen Stoof-Leichsenring, Tony Heitkam, Nicola Schmidt, Barbara von Hippel, Andrej A. Andreev, Bernhard Diekmann, Boris K. Biskaborn, Bernd Wagner, Martin Melles, Lyudmila A. Pestryakova, Inger G. Alsos, Konstantin V. Krutovsky, Ulrike Herzschuh
Status	*Under review* at *communications biology* **update after submission of thesis**: revised version published Schulte, L., Meucci, S., Stoof-Leichsenring, K.R. *et al. Larix* species range dynamics in Siberia since the Last Glacial captured from sedimentary ancient DNA. *Commun Biol* **5,** 570 (2022). https://doi.org/10.1038/s42003-022-03455-0
Summary	In this manuscript we apply the previously established method of chloroplast hybridization capture to samples from 8 lakes from across Siberia and are thus able to track the population history of *Larix* in time and space. The results show that *L. sibirica* likely invaded its northern distribution area only in the Holocene and had a much wider distribution towards the east around 33,000 years ago. During the Last Glacial Maximum, samples consistently showed variants typical for *L. gmelinii*. With a second set of hybridization probes targeting nuclear genes, the analysis of off-target reads revealed excess of a main repetitive DNA sequence specific for *Larix*. We showed that the sequence of this repetitive element was highly conservative in time and space.
Author contributions	LS under the supervision of UH selected cores and samples for the study. LS did the core sampling and supervised the DNA extraction conducted by Svetlana Karachurina (student helper). BH and SM did the laboratory work of library built and hybridization capture. SM developed the nuclear gene capture probes with the help of KVK. KSL advised laboratory work. TH and NS did the analysis of repetitive elements in modern larch reference genomes. AAA counted pollen of Lake Kyutyunda samples. Field campaigns for lake sediment core drilling were organized by BD (Billyakh, Malaya Chabyda, Kytuyunda), BKB (Kyutyunda, Satagay), BW (Emanda), MM (Lama), IGA (Bolshoye Shchuchye), LAP (CH12) and UH (CH12). LS conducted all bioinformatic analyses, prepared figures and wrote the manuscript under the supervision of UH. All co-authors commented on the manuscript.

2 Manuscript I

Forest-permafrost feedback loops and glacial refugia help explain the unequal distribution of larch across continents

Status

Submitted to *Journal of Biogeography*
update after submission of thesis: revised version accepted for publication at Journal of Biogeography

Authors

Luise Schulte[1,2], Chenzhi Li[1,3], Simeon Lisovski[1], Ulrike Herzschuh[1,2,3]

Affiliations

[1] Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Polar Terrestrial Environmental Systems, Potsdam, Germany
[2] Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany
[3] Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany

Keywords
bioclimatic niche, glacial refugia, larch, *Larix*, permafrost, phylogeography, postglacial recolonization

2.1 Abstract

Aim: The deciduous needle-leaf tree larch (*Larix* Mill.) is a prominent component of boreal and montane forests. However, its distribution across the northern hemisphere differs markedly between the continents. Vast areas of northern Asia are covered by monospecific larch forests, whereas the species is scattered and confined to marginal spaces in North America, and mostly restricted to mountainous regions in central Europe. Our aim is to identify potential triggers and drivers of the current distribution patterns by comparing species' bioclimatic niches, glacial refugia, and postglacial recolonization patterns.
Location: Northern hemisphere
Taxon: Species of the genus *Larix* (Mill.)
Methods: We synthesize pollen, macrofossil, and ancient DNA paleo-evidence of past *Larix* occurrences of the last 60,000 years and track distribution patterns through time. Bioclimatic niches are analyzed using species ranges, sites of dominance, modern climate, modeled permafrost extent, and active layer thickness.
Results: Bioclimatic niches show large overlaps between Asian larch species and American *L. laricina.* The distribution across various degrees of permafrost extent is distinctly different for Asian *L. gmelinii* and *L. cajanderi* compared to the other species*,* whereas the distribution on different depths of active layer thickness is more similar among Asian and American species. Northern glacial refugia for *Larix* are only present in eastern Asia and Alaska.
Main Conclusion: Our results suggest that larch dominance in Asia is dependent on the interaction of several factors which allows *L. gmelinii* and *L. cajanderi* to dominate where these factors coincide. These factors include the early postglacial spread out of northern glacial refugia in the absence of competitors as well as a positive feedback mechanism between frozen ground and forest.

2.2 Introduction

The deciduous needle-leaf tree genus *Larix* (larch), is a prominent component within the northern hemisphere's boreal and montane forests. Larch-dominated forests form approximately 20% of the global boreal forests (Osawa & Zyryanova, 2010) and are important areas for biodiversity as well as acting as large global biospheric carbon sinks. High-latitude and high-elevation areas are especially sensitive to climate change (Seddon et al., 2016) and are most strongly affected by the recent warming. Boreal forests and the underlying permafrost are even discussed as an important component in tipping points of the global climate system (Lenton et al., 2019). Hence, understanding the history, function, and stability of larch forest is of global relevance.

The relative percentage of larch in northern hemisphere forests differs markedly between the continents. In northern Asia, *Larix* is the most common tree, forming over 80% of the boreal forests (Herzschuh, 2020). In contrast, the northern boreal forests of Europe are dominated by evergreen species such as pine (*Pinus)* and spruce (*Picea)* and deciduous broad-leaved tree taxa including birch (*Betula*), but no naturally occurring larches (Pividori et al., 2016). The natural distribution of *Larix* is restricted to central Europe, with disjunct populations in the major mountain ranges and a small population in the Polish lowlands (Geburek, 2014). In North America, *Larix* is widely distributed, ranging from the eastern coast of northern USA and Canada, to the Rocky Mountains in the west and north into Alaska (Cheliak et al., 1988). Despite its wide range, larch is mainly a companion species with the majority of forests dominated by spruce and fir (*Abies)* (Warren et al., 2016). Although larch is such a prominent conifer in the boreal forest ecosystems, the quantitative differences between the continents are little understood, which raises the question of whether this pattern will persist in the future.

Depending on a debated phylogeny, the genus *Larix* contains 10 to 15 species (Abaimov, 2010; Khatab et al., 2008; Wei & Wang, 2004). Recognized taxonomic units are often not separated by geographic barriers and *Larix* species easily hybridize in nature, resulting in conflicting classifications of full species, subspecies, or varieties (Polezhaeva 2010, Abaimov 2010). In Europe, only one species, the European larch *L. decidua* (Mill.), is recognized with a number of varieties or subspecies described in its disjunct populations (Pâques et al., 2013). In North America, three species are recognized, but only *L. laricina* (Du Roi) K. Koch has a wide distributional range from the Atlantic to Alaska (LePage & Basinger, 1995). In Asia, the phylogeny is particularly unsettled, but here we recognize three species – *L. sibirica* Ledeb. (Siberian larch), *L. gmelinii* (Rupr.) Rupr. (Dahurian larch), and *L. cajanderi* Mayr (Cajander's larch) – which form vast monospecific forests across Siberia ranging from European Russia to the Russian Far East (Abaimov, 2010).

Larix is light-demanding, efficient in its use of nitrogen and carbon, and can grow on poor soils (LePage & Basinger 1995): all features that make the species a pioneer on recently disturbed soils. However, its shade-intolerance results in its replacement during succession by more shade-tolerant species such as spruce, pine, or fir (Kharuk et al., 2007; Pluess, 2011; Uemura et al., 1997). Succession is consequently put forward as an explanation for the absence of *Larix* across wide parts of Europe and America (Wagner et al., 2015; Warren et al., 2016). But why is larch not replaced during succession in northern Asia?

Traditionally, it has been assumed that bioclimatic limits, that is the extreme cold hardiness of larches, explain the dominance of larch in northern Asia (Archibold, 1995; Kharuk et al., 2007;

Walter, 1985). The difference in dominance of larch between the continents could thus be related to the different bioclimatic niches of the larch species. The extent to which the bioclimatic niches of the individual species overlap has not yet been systematically investigated nor whether the different growth limits explain the differences in distribution between the continents. Increasing evidence suggests other factors are responsible for larch dominance in Asia such as the extensive permafrost, as both a water source and an impediment to evergreen invasion (Abis & Brovkin, 2019; Sugimoto et al., 2002; Zhang et al., 2011). Furthermore, frequent fire return intervals favor larch recruitment over evergreens (Schulze et al., 2012; Uemura et al., 1997). Recently, the additional factor of glacial history, where Asian larch dominance is considered as a legacy of the severe Last Glacial, was proposed (Herzschuh, 2020; Herzschuh et al., 2016). A systematic comparison of where *Larix* species survived the Last Glacial and how the postglacial recolonization took place has not yet been made for the two continents.

In this study, we aim to synthesize the available data on current larch distribution and dominance as well as data of the current climate and permafrost extent in order to compare bioclimatic niches of the most prominent larch species of the northern hemisphere. Furthermore, we collectively analyze paleo-evidence of larch in the northern hemisphere using pollen, macrofossil, and ancient DNA proxies, and compare glacial refugia and postglacial migration patterns. By integrating different data sources, the overall aim of the study is to provide a better understanding of why there are such strong differences in the *Larix* distribution across continents in the northern hemisphere. To this end, we ask the following questions: (1) Can we explain differences in modern *Larix* distribution by the divergence of bioclimatic limits in the *Larix* species? (2) Is past survival in northern glacial refugia a pre-requisite for *Larix* dominance today? Or alternatively, (3) are current patterns best explained by differences in postglacial migration dynamics?

2.3 Material and methods

2.3.1 Bioclimatic limits

We used three datasets to derive species bioclimatic limits. Modern dataset m1: the natural species distribution boundaries describing the area where a species can be found according to observations. Species distribution boundaries were derived from Mamet et al. (2019). Dataset m2: the ESA Land Cover CCI (ESA, 2017) indicates where larch is the dominant tree species. Dataset m3: climate variables from the WorldClim2 dataset with 30 arc-second spatial resolution (around 1 km^2 spatial resolution, depending on latitude) from which we use averaged monthly climate data from 1970 to 2000 for January and July temperatures and annual precipitation (Fick & Hijmans, 2017). We restricted the dataset to randomly generated sites north of 30°N (dataset m3). The sites were then confined to the natural species distribution boundaries (dataset m1), resulting in 1,449,036 sites. In a second analysis, the sites were further confined to sites with the land-cover classification “needle-leaved deciduous trees, closed to open (>15%)” from the ESA CCI land-cover map (dataset m2), resulting in 76,955 sites. Sites within the natural species distribution and with the particular land-cover attributes were treated as sites with dominance of *Larix*. Total area and proportion of dominance was calculated by comparing datasets m1 and m2. The species growth on permafrost was calculated both for the species distribution (dataset m1) and the areas of dominance (dataset m2), using the modeled extent of permafrost for 2019 indicating the proportion of ground ice that is perennially frozen (Obu et al., 2021b) and the modeled active-layer thickness of permafrost for 2019 indicating the depth of the seasonal thaw layer (Obu et al., 2021a). All calculations were done in QGIS (v. 3.10.14, QGIS Development Team,

2020) and R (R Core Team, 2013). Code used in the analysis is given in the supplementary information.

2.3.2 Pollen, macrofossil, and DNA datasets

Pollen data

The pollen datasets were obtained from Herzschuh et al. (2021), comprising 2831 global records with standardized chronologies (Li et al., 2021). Here, we only analyzed datasets containing *Larix (848 sites)*. Additionally, we used pollen data of Europe collected in Wagner et al. (2015). All data were binned into 9 time slices to capture general patterns: 0-3 thousand calibrated years before present (ka BP, where 'present' is 1950 CE), 3-6, 6-9, 9-12, 12-15, 15-18, 18-23, 23-28, and 28-60 ka BP. For the past occurrence analysis, each time slice with at least one occurrence of *Larix* was counted as presence.

Macrofossils

Larix macrofossil data were extracted from three databases: the macrofossil database compiled by Binney et al. (2009), downloaded from https://oxlel.zoo.ox.ac.uk/resources, the macrofossil database hosted by the National Oceanic and Atmospheric Administration (NOAA) (Gastaldo, 2020) downloaded from ftp://ftp.ncdc.noaa.gov/pub/data/paleo/plantmacros/napmd/, and macrofossil data from the Neotoma database (Williams et al., 2018), downloaded using the Neotoma Explorer and restricting the search to genus *Larix* and type macrofossil. Additionally, we used macrofossil data compiled in Wagner et al. (2015). For each site, each time slice with at least one occurrence of a macrofossil (needles, cones, twigs, or any other evidence of *Larix)* was counted as *Larix* presence.

DNA

To compile ancient DNA evidence of *Larix*, the literature was searched for publications using the metabarcoding primers amplifying a short fragment of the chloroplast *trn*L (UAA) intron 'g/h' published by Taberlet et al. (2007). This was done by searching the data repository https://datadryad.org/ separately with the keywords 'ancient DNA', 'metabarcoding', 'barcoding', 'plant', and 'vegetation', as well as searching all publications that cited Taberlet et al. (2007) on https://scholar.google.de/ and the additional key word 'sediment'. Included DNA datasets are lake sediment records of d1: the Omoloy region (Liu et al., 2020), d2: the Taymyr Peninsula (Epp et al., 2018), d3: Lake Billyakh (Herzschuh et al., in prep.), d4: Lake Illerney (Herzschuh et al., in prep.), d5: a dataset including samples from 7 lakes across Siberia (Herzschuh et al. in prep), d6: Lake Bolshoy Toko (Courtin et al., 2021), d7: Lake Bolshoye Shchuchye (Clarke et al., 2019), and samples from permafrost from d8: southern Chukotka (Sonstebo et al., 2010), d9: Bol'shoy Lyakhovsky Island (Zimmermann et al., 2017), d10: Buor Khaya Peninsula (Zimmermann et al., 2017), and d11: samples from around the Arctic (Willerslev et al., 2014).

For datasets d1-d6, raw data were analyzed with Obitools3 (Boyer et al., 2016, version 3.0.0b38) using the commands *alignpairedend* (to merge overlapping paired end sequences), *grep -a mode:alignment* (to remove unaligned sequences), *ngsfilter* (assign each sequence to the corresponding sample/marker combination), *unique* (dereplicate into unique sequences), and *clean -r 0.05* (clean sequence errors, -r defines ratio between the counts of two sequences so that the less abundant one can be considered a variant of the more abundant one). Reads were classified to a taxonomic level using two reference databases: the first database is based on the quality-checked and curated arctic and boreal vascular plant and bryophyte reference database

published by Sønstebø et al. (2010), Willerslev et al. (2014), and Soininen et al. (2015). The second database is based on the EMBL Nucleotide Database standard sequence release 143 (Kanz et al., 2005). Taxonomic assignments were filtered for a minimum identity of 0.97 and the percentages for *Larix* in each dataset calculated. The other datasets were incorporated in that workflow according to the state in which the data was published. For the analysis of past occurrences, a threshold of 0.01% *Larix* read counts per sample was applied. A mean percentage of *Larix* was calculated for the 9 selected time slices as described for the pollen dataset.

2.3.3 Ice sheets

Ice-sheet margins were adopted from Gowan et al. (2001). For the ten selected time slices (in parentheses), the following reconstructed ice margins were used: 30 kaBP (60-28 kaBP), 25 kaBP (28-23 kaBP), 20 kaBP (23-18 kaBP), 17.5 kaBP (18-15 kaBP), 12.5 kaBP (15-12 kaBP), 10 kaBP (12-9 kaBP), 7.5 kaBP (9-6 kaBP), 5 kaBP (6-3 kaBP), and 0 kaBP (3-0 kaBP).

2.4 Results

2.4.1 Bioclimatic limits of *Larix* and its distribution on permafrost

The degree of dominance of *Larix* species in their respective ranges differs markedly across the northern hemisphere. The European *L. decidua* dominates over around 0.5 thousand (K) km^2 which is 0.2% of its total range; American *L. laricina* dominates 11 K km^2 or 0.2% of its range; whereas the Asian larch species cover the greatest areas as dominant tree taxa with *L. sibirica* dominating 6.5 million (M) km^2 or 11%, *L. cajanderi* dominating 1.1 M km^2 or 54%, and *L. gmelinii* dominating 3 M km^2 or 55% of its total range (Fig. 1 and 2). Notwithstanding these distributional differences, our analyses show that the climate conditions of the American and Asian larch species have large overlaps (Fig 3).

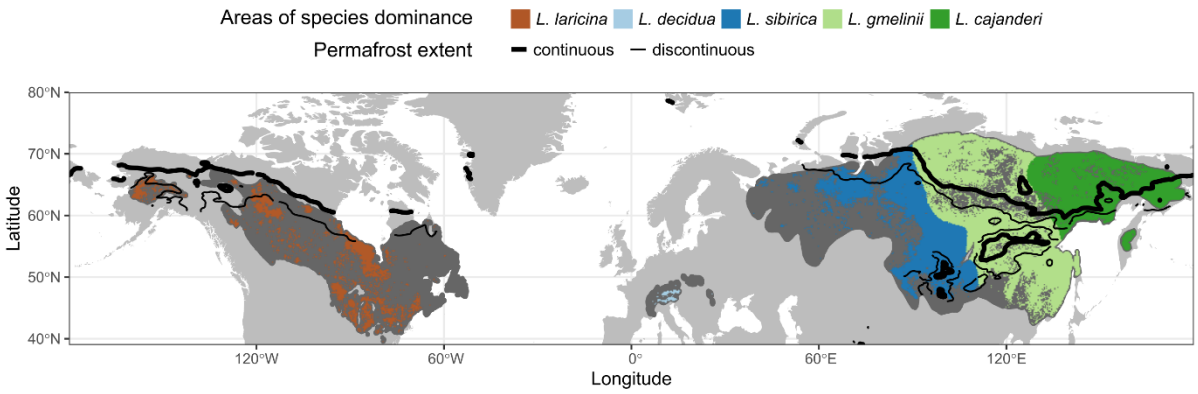

Fig. 1 Species distribution boundaries, dominance, and permafrost extent. Natural *Larix* species distributions according to the literature shaded in dark gray (Mamet et al. 2019), and dominance of *Larix* according to ESA CCI land-cover classification in colors (ESA, 2017); Black lines: margins of continuous (95%) and discontinuous (50%) permafrost extent (Obu et al., 2021b)

The three widespread boreal species, *L. sibirica, L. gmelinii,* and *L. cajanderi* have great overlaps in their bioclimatic niches, with *L. gmelinii* lying in between the two other species when considering January temperatures (Fig. 3A). All three species grow and dominate in extremely low January temperatures, with a gradient in the lower growing limit from *L. sibirica* (ca -32°C), over *L. gmelinii* (ca -40°C), to *L. cajanderi* (ca -45°C). The bioclimatic range of dominance in July temperatures is very similar between the three species at around 9-10°C, which is slightly lower than the described tree-line zone of where the average July temperature declines to 12.5-10°C

(MacDonald et al., 2008). The percentage of sites with dominance in relation to their whole range is shifted for all three species towards the higher end of annual temperature range with *L. gmelinii* and *L. cajanderi* dominating especially in annual climate variation of over 50 °C and low annual precipitation of below 300 mm per year (Fig. 3B and 3C). *Larix sibirica* dominates towards the more continental climate of its range, although this rapidly declines after 47 °C annual temperature amplitude with around 500 mm annual precipitation.

Comparing the occurrence of the species on the different degrees of permafrost, we see a gradient from *L. sibirica* mostly growing outside the permafrost zone, to *L. gmelinii* which has prominent parts of its range inside the discontinuous and continuous permafrost zone, to *L. cajanderi* with a range overlap of over 70% with the extent of continuous permafrost (Fig. 1 and 4A). The distribution of *L. cajanderi* is associated with a seasonal permafrost active-layer thickness (ALT) of mainly under 1 meter (Fig. 4B): *L. gmelinii* and *L. sibirica* show a similar distribution on ALT.

Larix decidua, the only species naturally occurring in Europe, forms dominant stands big enough to be visible in satellite images only in the Alps (Fig. 1). Compared to the other species, *L. decidua* has a distinct niche that overlaps little or not at all with the niches of the other species. *Larix decidua* occurs where January temperatures are warmer, and conditions are wetter (Fig. 1). However, regions with higher dominance are shifted towards lower July and January temperatures (Fig. 3).

In North America, dominance of the widespread boreal species *L. laricina* is scattered across its range (Fig. 1) with a very limited proportion of dominance compared to its range (Fig. 2). Minimum mean January temperature for *L. laricina* growth is around -30°C, minimum July temperature around 12.5°C, as described in literature (Johnston, 1990). *Larix laricina* dominance is limited to these cold and dry extremes of its range, while it loses its dominance at the warmer and wetter end of the January bioclimatic niche (Fig 2). When considering average July and January temperatures, we see a big overlap in the ranges of *L. laricina* and the three boreal species of Asia, *L. sibirica, L. gmelinii,* and *L. cajanderi*, as also reported by Mamet et al. (2019), except that *L. laricina* is unable to grow in as low January temperatures as the Asian species preferring higher January temperatures. The distribution range of *L. laricina* lies mostly in non-permafrost areas (Fig. 1 and 4A). Those populations occurring inside the permafrost zone grow on a relatively shallow ALT of around 1 meter which is less than for the main distribution of *L. sibirica* and *L. gmelinii* on permafrost (Fig. 4B).

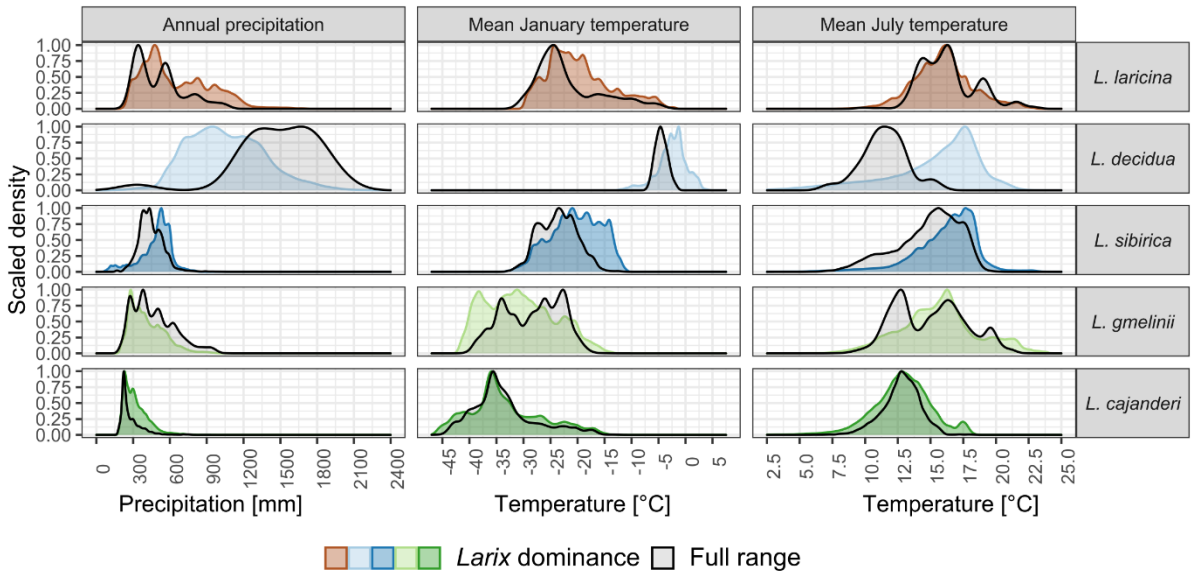

Fig. 2 Bioclimatic ranges of *Larix* species A: In gray shading below the black lines: scaled densities of climatic values for sites in the species distribution range; in colors, climatic values for sites with *Larix* dominance.

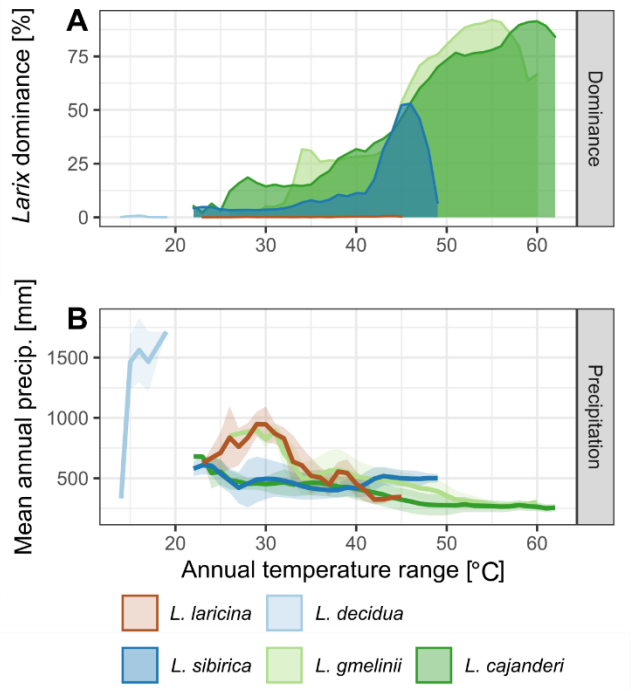

Fig. 3 Dominance of *Larix* in climate continentality A: Percentage of area dominated by *Larix* from the full distribution per annual temperature range in degree Celsius [°C]; B: mean annual precipitation and standard deviation per annual temperature range for sites of dominance.

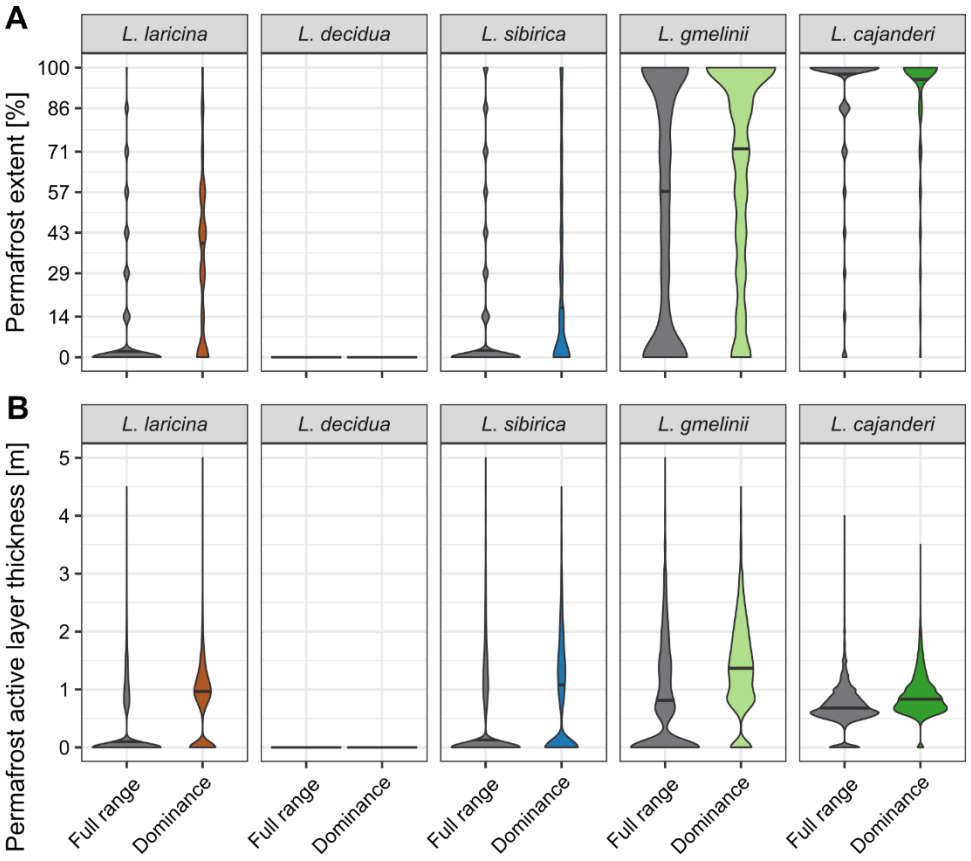

Fig. 4. Larch species dominance and distribution on permafrost extent and active-layer thickness A: Occurrence on permafrost (0% no permafrost, >60% discontinuous permafrost, 100% continuous permafrost). B: Distribution on permafrost active-layer thickness (0 meter for regions of no permafrost). In colors: areas of *Larix* dominance, in gray: area of species distribution range, horizontal line indicates median.

2.4.2 Glacial occurrence patterns of *Larix*

The collection of *Larix* data from pollen, macrofossil, and ancient DNA records shows the widespread occurrence of *Larix* in northern Asia during the LGM (Fig 5). In contrast, North American and European glacial refugia were locally confined towards the south of the ice shield and within unglaciated Alaska.

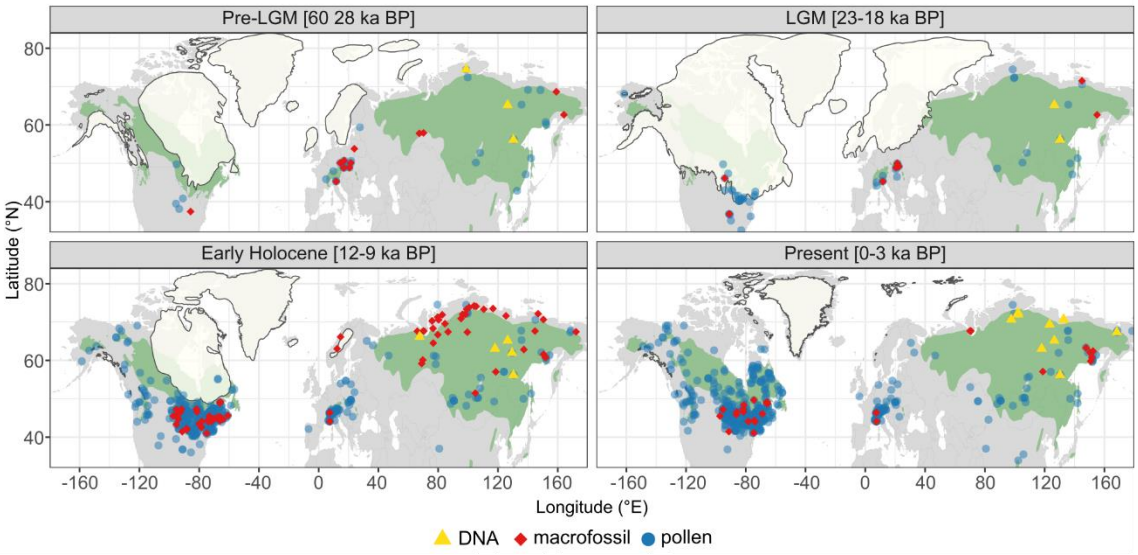

Fig. 5 Past *Larix* occurrences. In green, modern distribution of Larix (Mamet et al., 2019), in white, reconstructed ice margins (Gowan et al., 2021). Colors and shapes indicate the presence of Larix in a dataset of the respective proxy in at least one sample of the time period indicated.

The pollen data also indicate a widespread occurrence of *Larix* in Asia for the periods before and after the LGM (Fig. 5). Refugia are not restricted to the southern parts but occur as far north as 74° north. Occurrences of larch before and during the LGM are more concentrated in the central and eastern parts, in contrast to an absence within the western limits of the modern range.

In Europe, after range expansion between 60 and 28 ka BP, *Larix* retreated during the LGM. Europe was heavily glaciated with an ice sheet covering Fennoscandia, the Baltic Sea, and adjacent areas (Fig. 5). Thus, pollen and macrofossil remains of *Larix* are only found in regions far south of the Eurasian ice sheet, in the Carpathian Mountains, the southern Alps, and southern Italy (Fig. 5).

Pollen and macrofossil findings in North America show glacial survival of *Larix* farther south than its current range, with occurrences as far south as South Carolina. The largest refugia were east and west of the Great Lakes. Pollen findings in Alaska prior to the retreat of the Laurentide ice sheet show local glacial refugia in this region.

Occurrence patterns of additional time slices are given in the supplementary information.

2.5 Discussion

The distribution pattern of larch forests across the northern hemisphere is still not well understood. By comparing major factors influencing *Larix* distribution in North America, Europe, and Asia we hope to better understand where these differences come from.

2.5.1 Are differences in species bioclimatic limits responsible for disparity in *Larix* distribution across continents?

Temperature and precipitation play a subordinate role

Our results show that Asian and American larch species have great overlaps in their bioclimatic niches, with the exception of regions that have extremely low mean January temperatures and

low annual precipitation, where only Asian *L. gmelinii* and *L. cajanderi* grow. In these regions of ultra-continental climate, it is likely that the species' adaptation to this environment is responsible for the larch dominance found here. However, the majority of locations which share similar climatic conditions in Asia and North America are occupied by larches and evergreen needle-leaf trees, respectively (Herzschuh, 2020). It is therefore unlikely that, in these moderate continental regions, temperature and precipitation are the only factors responsible for the larch dominance found in Asia.

The bioclimatic niche of European *L. decidua* has little overlap with the other species and grows in much more oceanic climates. Still, with respect to its full range, *L. decidua* dominance is shifted towards lower July and January temperatures in the relatively warm central Europe (Fig. 3). The species increases in frequency on the southern slopes of the Alps where it forms the upper tree limit in pure forests, whereas on the more oceanic influenced north-facing slopes it is more often found in mixed stands with other alpine tree species such as *Pinus cembra* (Swiss stone pine) (Da Ronch et al., 2016; Geburek, 2014). However, compared to other *Larix* species, it is less cold tolerant which is why Asian larches are preferred for artificial plantations in Scandinavia (Karlman, 2010; Pâques et al., 2013).

The importance of permafrost

Comparisons of distribution ranges and areas of dominance on permafrost show that only the Asian *L. gmelinii* and *L. cajanderi* overlap considerably with areas of continuous permafrost (Fig. 4A). Both Asian *L. sibirica* and American *L. laricina* are distributed partly in discontinuous permafrost regions, but the area of dominance for *L. sibirica* is relatively more biased towards no permafrost. This corresponds to the description of *L. sibirica* being sensitive to growth on permafrost (Abaimov, 2010), whereas Alaskan *L. laricina* is reported to occur especially on bogs underlain by permafrost (Johnston, 1990). While both *L. gmelinii* and *L. cajanderi* are described as being able to grow on low ALT (Abaimov et al., 1999), only *L. cajanderi* has its main dominant distribution in areas with <1 meter ALT. In contrast, the distribution of *L. gmelinii* extends to relatively high ALT with a median ALT even higher than for American *L. laricina* (Fig. 4B). Larches are able to develop adventitious roots, allowing them to grow on poorly drained soils (Abaimov et al., 1999; Geburek, 2014; Islam & Macdonald, 2004). The active layer on permafrost is often poorly drained, as the frozen soil prevents rapid drainage and thus holds water from snow melt, summer precipitation, and permafrost thaw in the active layer zone (Schulze et al., 2002; Sugimoto et al., 2002; Zhang et al., 2011). Asian larches effectively use this water storage throughout summer drought which allows the establishment of forests where the vegetation would otherwise be semidesert (Krestov et al., 2009; Tchebakova et al., 2010). But also outside these extremely dry zones, research suggests a coupled system of permafrost and larch dominance, where permafrost supplies water, whereas larch maintains permafrost by intercepting solar radiation, effectively decoupling permafrost from climate (Zhang et al., 2011). Tree cover also decreases seasonal permafrost thaw by drying the soil through evapotranspiration and thus decreasing thermal conductivity of the soil (Fisher et al., 2016). *Larix* species are reported to have limited ability to control stomatal closure, which results in higher transpiration rates but also higher photosynthesis rates (Anfodillo et al., 1998; Eilmann & Rigling, 2012; Urban et al., 2019). This further increases the positive feedback mechanism with permafrost, increasing soil drying by respiration, while enabling high productivity in larch compared to other species (Sugimoto et al., 2002). Although American *L. laricina* also grows on (discontinuous) permafrost such a positive feedback mechanism has not been reported. In zones of discontinuous permafrost, the frozen soil often elevates above the ground, resulting in well-

drained permafrost 'plateaus' (due to the volumetric expansion of freezing water) (Dearborn et al., 2021). Whereas *L. laricina* often dominates the poorly drained wetlands (Cheliak et al., 1988; Johnston, 1990), on the well-drained elevated permafrost it is often outcompeted by *Picea mariana* (Dearborn et al., 2021). This suggests that although permafrost plays an important role in the dominance of Asian larch species, more factors are needed to explain their wide-ranging dominance.

2.5.2 Do high latitude glacial refugia guarantee larch dominance?

The continents of the northern hemisphere show pronounced differences in the occurrence and position of glacial refugia. Whereas high latitude glacial refugia are found in eastern and central parts of northern Asia and in Alaska, they are absent in the western part of northern Asia, the rest of northern America, and Europe (Fig. 5). As today *Larix* dominates mostly the central and eastern part of northern Asia, mainly in the distribution areas of *L. gmelinii* and *L. cajanderi,* we assume that the existence of northern glacial refugia played an important role in conveying this dominance. However, if the rapid spread out of these refugial populations would result in later dominance, we should also expect to see a dominance of larch in Alaska; yet this is not the case, as Alaskan forests are dominated by spruce intermingled with birch and aspen (Trugman et al., 2018). The situation of the refugial populations in Alaska, is however, quite distinct from the situation in Asia. Alaska is geographically isolated from the rest of northern America by the Canadian Rocky Mountains. Genetic studies show that gene flow between the populations from Alaska and the rest occurred only unidirectionally out of Alaska, probably by long-distance pollen transportation via the prevailing westerly winds (Napier et al., 2020). And although Alaska was connected with the Russian Far East during the Pleistocene glacial periods by the Bering Land Bridge (Elias & Crocker, 2008), Alaskan lineages could not apparently be genetically rescued by populations from the Russian Far East. American and Eurasian larches are genetically well separated groups (Gros-Louis et al., 2005; Semerikov & Lascoux, 1999; X. X. Wei & Wang, 2003) and the steppe-tundra dominated Beringian Land Bridge is unlikely to have served as a refugium for woody taxa during the LGM (Elias & Crocker, 2008; Wang et al., 2017). This resulted in genetic impoverishment and decreased fitness of the Alaskan population (Napier et al., 2020). In addition, evergreen tree species also survived in glacial refugia in Alaska and likewise could spread out of these populations after climate warming (Roberts & Hamann, 2015), whereas in Asia, northern glacial refugia of evergreens were mostly confined to the western parts and almost absent from central and eastern Asia (Binney et al., 2009). In contrast to the Alaskan population, current *Larix* populations of northern Asia are known to be genetically well connected (Kruse et al., 2018; Polezhaeva et al., 2010; Zimmermann et al., 2019).

2.5.3 What role does postglacial migration play in larch dominance?

Remarkable differences exist between the continents in the postglacial migration patterns. In Europe, *Larix* spread out of the refugial populations early after the LGM with prominent advances in the central European lowlands and rare occurrences by the Baltic Sea (Wagner et al., 2015). Despite this early spread, it remains unclear if *Larix* ever constituted a dominant tree species in European forests after the LGM. A study mapping Eurasian vegetation biomes of the last 21 ka in Eurasia, does not show a dominant cold-deciduous biome in Europe in the past (Binney et al., 2017). Instead, in Europe, steppe and tundra are directly replaced by temperate deciduous and evergreen needle-leaf forests from 14 ka BP on (Binney et al., 2017). In the course of the Holocene, *Larix* becomes restricted to mid- and high-elevations, as competing tree taxa become dominant in the lowlands (Wagner et al., 2015). The decrease in population under the mild

Holocene climate has been interpreted as a better adaptation of *L. decidua* to the severe conditions of the glacial epoch (Semerikov et al., 2013).

Larix decidua presumably never reached Scandinavia after the LGM. For broad-leaved European tree species, it has been shown that such a pattern can be best explained by postglacial dispersal limitations and that many species did not reach equilibrium with climate (Svenning & Skov, 2007). For *L. decidua,* it has been suggested that its distribution became restricted due to climate driven expansion of more competitive thermophilus taxa (Wagner et al., 2015). *Larix* macrofossils from 9 and 8 ka BP have been detected sporadically in Fennoscandia (Kullmann, 1998). These macrofossils were identified as *L. sibirica* and suggest an extremely rapid dispersal, probably with long-distance jumps without stepwise sequential migration (Kullmann, 1998). It has been suggested that the disappearance of *L. sibirica* coincided with the change to a more oceanic climate and that it was consequently outcompeted in Fennoscandia after 8 ka BP by *Pinus*, *Picea,* and *Betula* (Kullmann, 1998; Oksanen, 1995). Thus, for larches in Europe, both climate and competition are the main drivers of the current distribution range.

In North America, *Larix* closely tracked the northward retreat of the ice sheets after the LGM. Due to the Canadian Rocky Mountains forming a geographical barrier between the Alaskan population and the rest of the North American population, the Alaskan population most likely contributed to the recolonization only by wind-dispersed pollen transported over the mountain ranges (Napier et al., 2020). A study including range-wide comparative pollen and genetic analyses by Warren et al. (2016) finds evidence for two distinct recolonization scenarios for western and eastern Canada. In western Canada, *Larix* arrived around 10-9 ka BP with a 4000-year time lag to the evergreen species *Picea* and *Abies* and thus was confined to the ecologically marginal sites not yet occupied. In contrast, in eastern Canada, in the region of Labrador and Quebec, larches were already established by 13 ka BP, before the arrival of the competitors *Picea* and *Abies,* resulting in a more genetically connected population (Warren et al., 2016). However, in both the early and late successional scenarios, American larch could not establish wide-ranging dominant forests as exist in Asia. This implies that an early establishment of larch forests is not sufficient to explain the Asian dominance.

The paleo-record does not provide clear insights into the recolonization of *Larix* in northern Asia. Some evidence exists that *Larix* spread across the entire current range after the LGM and notably at the beginning of the Holocene (Fig. 4). This pattern has been frequently discussed in the literature as a glacial survival and rapid spread out of refugia with climate warming (Binney et al., 2009; Brubaker et al., 2005; Lozhkin et al., 2018; MacDonald et al., 2008; Tarasov et al., 2007). Spreading from high-latitude refugia would have given *Larix* a recolonization advantage over their evergreen competitors which had only a few and possibly genetically depleted northern glacial refugia (Binney et al., 2009; Herzschuh, 2020). In stark contrast, the advantage of *Larix* in North America would have been diminished, as recolonization could only follow the melting ice sheet. However, it remains speculative since neither pollen nor ancient DNA metabarcoding provide enough information on the origin of the *Larix* populations.

2.5.4 Fire as an additional factor

An additional factor discussed in the literature for promoting *Larix* dominance in Siberia is fire. In contrast to *Abies* and *Picea,* Asian larch species are fire-resistant, as they typically develop a thick bark and shed low-hanging branches which prevents a fire from reaching the crown (Schulze et al., 2012; Uemura et al., 1997; Wirth et al., 2002). Additionally, seeds are stored in serotinous cones and can remain in the tree canopy for several years, thus viable seeds are available after

surface fires (Osawa et al., 2010). The deciduous habit of *Larix* effectively accumulates fuel for frequent fires which are, at the same time, less severe due to the comparatively high leaf-water content (Rogers et al., 2015). Frequent low-intensity surface fires promote successful recruitment of *L. gmelinii* and *L. cajanderi* whereas they impede evergreen establishment (Schulze et al., 2012; Uemura et al., 1997). Interestingly, although *L. sibirica* is likewise fire-resistant, it is typically replaced after fire by *Betula* sp. and larch forest only recovers after several decades (Abaimov & Sofronov, 1996). In contrast to the Asian species, American *L. laricina* is fire-sensitive (Brown & Zobel, 1988). This species has only a thin bark and non-serotinous cones, so trees are easily killed and seeds consumed by fire (Busque & Arseneault, 2005; Johnston, 1990). Occasional fires may only be beneficial for *L. laricina* recruitment if abundant seeds are available from adjacent unburned sites, whereas the semi-serotinous seeds of *Picea mariana* (Black spruce) are readily available for postfire recolonization (Brown & Zobel, 1988). *Larix decidua* has a thick bark like the three Asian larch species and is reported to be a fire-resister, regenerating moderately well after fire disturbance (Moris et al., 2017). Periodic fire return intervals during the Holocene might have favored *Larix* establishment (Aurélie et al., 2009; Blarquez & Carcaillet, 2010; Moris et al., 2017), but, from around 4000 years ago, fire regimes in the Alps have been under a strong anthropogenic impact (Leys & Carcaillet, 2016).

2.5.5 Outlook

Although northern glacial refugial populations are visible in the fossil record, it is still uncertain to what extent modern *Larix* populations spread from northern refugia and to what extent they were replaced by the invasion of more southerly populations. A modern genetic study based on restriction site-associated DNA sequencing (RADseq) (Miller et al., 2007) or genotyping by sequencing (GBS) (Elshire et al., 2011) could analyze hundreds of thousands of loci at a low cost and give deep insights into their phylogeography as well as adaptation (Parchman et al., 2018). The methods are increasingly being applied to forest tree species but a widespread study of *Larix* has yet to be undertaken (Parchman et al., 2018).

A second approach to decipher the contribution of glacial refugia to modern populations could be an extended ancient DNA approach. Traditional paleoproxies such as fossilized pollen but also ancient DNA metabarcoding are only able to indicate larch occurrence and relative abundance. The method of hybridization capture, specifically enriching for a region or genome of interest, has been shown to work well also on degraded ancient DNA from diverse materials such as ancient bones (Dabney et al., 2013), ancient gourd rinds (Kistler et al., 2014), subfossil needles (Schmid et al., 2017), and cave sediments (Slon et al., 2017). Recently the method has also been applied to enrich ancient *Larix* chloroplast DNA from a lake record spanning the last 7,000 years (Schulte et al., 2021). The method has high potential ability and when applied to lake sediment samples from different ages and study sites could serve to reveal in detail the past population dynamics of *Larix.*

2.6 Conclusion

As our comparative analysis shows, the widespread dominance of Asian *L. gmelinii* and *L. cajanderi* is probably a result of several factors coinciding which have not occurred in North America and Europe. For a light-demanding species such as larch to establish, the first two pre-requisites seem to be a severe glacial resulting in widespread frozen ground and a forest-free landscape, and *Larix* species capable of growing on such cold and frozen soils. A second factor is probably a head-start of *Larix* in the postglacial recolonization compared to its competitors

conveyed by genetically fit northern glacial refugia. To then maintain *Larix* dominance despite a climate now suitable for evergreen competitors, a complex interaction of forest, fire, and permafrost seems to be necessary to impede the otherwise natural succession to an evergreen climax forest. Only the Asian species *L. gmelinii* and *L. cajanderi* meet all these requirements. They are well adapted to permafrost, being capable of growing in shallow thawing soils, they are good fire-resisters, and they survived in probably genetically well-connected glacial refugia in the absence of many evergreen competitors.

2.7 Acknowledgements

We gratefully acknowledge Thomas Böhmer for help with the pollen and climate data preparation and Stefan Kruse for the discussion of *Larix* traits. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 772852, ERC Consolidator Grant 'Glacial Legacy') and the Initiative and Networking Fund of the Helmholtz Association.

2.8 Author contributions

LS searched the literature to collect published macrofossil databases and ancient DNA metabarcoding datasets. CL collected and harmonized published pollen datasets. SL helped with data analyses of the permafrost datasets. LS did all other analyses of the study, prepared the figures and wrote the manuscript under the supervision of UH. All authors commented on the manuscript.

2.9 References

Abaimov, A. P., & Sofronov, M. A. (1996). The Main Trends of Post-Fire Succession in Near-Tundra Forests of Central Siberia. In J. G. Goldammer & V. v. Furyaev (Eds.), *Fire in Ecosystems of Boreal Eurasia* (1st ed., pp. 372–386). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-8737-2_33

Abaimov, Anatoly P. (2010). Geographical Distribution and Genetics of Siberian Larch Species. In *Permafrost ecosystems Siberian larch forests* (pp. 17–36).

Abaimov, Anatoly P., Lesinski, J. A., Martinsson, O., & Milyutin, L. (1999). *Variability and Ecology of Siberian Larch Species* (Vol. 43). Swedish University of Agricultural Sciences, Department of Silviculture.

Abis, B., & Brovkin, V. (2019). Alternative tree-cover states of the boreal ecosystem: A conceptual model. *Global Ecology and Biogeography*, *28*(5), 612–627. https://doi.org/10.1111/geb.12880

Anfodillo, T., Rento, S., Carraro, V., Furlanetto, L., Urbinati, C., & Carrer, M. (1998). Tree water relations and climatic variations at the alpine timberline: Seasonal changes of sap flux and xylem water potential in Larix decidua Miller, Picea abies (L.) Karst. and Pinus cembra L. *Annales Des Sciences Forestieres*, *55*(1–2), 159–172. https://doi.org/10.1051/forest:19980110

Archibold, O. W. (1995). *Ecology of world vegetation*.

Aurélie, G., Xavier, M., Sandrine, C., & Christopher, C. (2009). The function of surface fires in the dynamics and structure of a formerly grazed old subalpine forest. *Journal of Ecology*, *97*(4), 728–741. https://doi.org/10.1111/j.1365-2745.2009.01518.x

Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., Blaauw, M., Damblon, F., Haesaerts, P., Kienast, F., Kremenetski, K. V., Krivonogov, S. K., Lozhkin, A. V., MacDonald, G. M. G. M. G. M., Novenko, E. Y., Oksanen, P., Sapelko, T. V., Väliranta, M., & Vazhenina, L. (2009). The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews*, *28*(23–24), 2445–2464. https://doi.org/10.1016/j.quascirev.2009.04.016

Binney, H., Edwards, M., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J. O., Andreev, A., Bezrukova, E., Blyakharchuk, T., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K., Nield, J., Novenko, E., Ryabogina, N., Solovieva, N., Willis, K., & Zernitskaya, V. (2017). Vegetation of Eurasia from the last glacial maximum to present: Key biogeographic patterns. *Quaternary Science Reviews*, *157*, 80–97. https://doi.org/10.1016/j.quascirev.2016.11.022

Blarquez, O., & Carcaillet, C. (2010). Fire, fuel composition and resilience threshold in subalpine ecosystem. *PLoS ONE*, *5*(8). https://doi.org/10.1371/journal.pone.0012480

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS : a UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Brown, K. R., & Zobel, D. B. (1988). Seed dispersal, seedling emergence, and early survival of Larix larcina (DuRoi) K. Koch in the Tanana Valley, Alaska. *Canadian Journal of Forest Research*, *18*, 306–314.

Brubaker, L. B., Anderson, P. M., Edwards, M. E., & Lozhkin, A. V. (2005). Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, *32*(5), 833–848. https://doi.org/10.1111/j.1365-2699.2004.01203.x

Busque, D., & Arseneault, D. (2005). Fire disturbance of larch woodlands in string fens in northern Québec. *Canadian Journal of Botany*, *83*(6), 599–609. https://doi.org/10.1139/b05-028

Cheliak, W. M., Wang, J., & Pitel, J. A. (1988). Population structure and genic diversity in tamarack, Larix laricina (Du Roi) K. Koch. *Canadian Journal of Forest Research*, *18*(10), 1318–1324. https://doi.org/10.1139/x88-203

Clarke, C. L., Edwards, M. E., Gielly, L., Ehrich, D., Hughes, P. D. M., Morozova, L. M., Haflidason, H., Mangerud, J., Svendsen, J. I., & Alsos, I. G. (2019). Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. *Scientific Reports*, *9*(1), 19613. https://doi.org/10.1038/s41598-019-55989-9

Courtin, J., Andreev, A. A., Raschke, E., Bala, S., Biskaborn, B. K., Liu, S., Zimmermann, H., Diekmann, B., Stoof-Leichsenring, K. R., Pestryakova, L. A., & Herzschuh, U. (2021). Vegetation Changes in Southeastern Siberia During the Late Pleistocene and the Holocene. *Frontiers in Ecology and Evolution*, *9*(April), 1–18. https://doi.org/10.3389/fevo.2021.625096

Da Ronch, F., Caudullo, G., Tinner, W., & de Rigo, D. (2016). Larix decidua and other larches in Europe : distribution , habitat , usage and threats. *San-Miguel- Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., Mauri, A. (Eds.), European Atlas of Forest Tree Species.*, 108–110.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763. https://doi.org/10.1073/pnas.1314445110

Dearborn, K. D., Wallace, C. A., Patankar, R., & Baltzer, J. L. (2021). Permafrost thaw in boreal peatlands is rapidly altering forest community composition. *Journal of Ecology*, *109*(3), 1452–1467. https://doi.org/10.1111/1365-2745.13569

Eilmann, B., & Rigling, A. (2012). Tree-growth analyses to estimate tree species' drought tolerance. *Tree Physiology*, *32*(2), 178–187. https://doi.org/10.1093/treephys/tps004

Elias, S. A., & Crocker, B. (2008). The Bering Land Bridge: a moisture barrier to the dispersal of steppe-tundra biota? *Quaternary Science Reviews*, *27*(27–28), 2473–2483. https://doi.org/10.1016/j.quascirev.2008.09.011

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, *6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

Epp, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A., & Herzschuh, U. (2018). Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. *Scientific Reports*, *8*(1), 17436. https://doi.org/10.1038/s41598-018-35550-w

ESA. (2017). *Land Cover CCI Product User Guide Version 2. Tech. Rep.* maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Fisher, J. P., Estop-Aragonés, C., Thierry, A., Charman, D. J., Wolfe, S. A., Hartley, I. P., Murton, J. B., Williams, M., & Phoenix, G. K. (2016). The influence of vegetation and soil characteristics on active-layer thickness of permafrost soils in boreal forest. *Global Change Biology*, *22*(9), 3127–3140. https://doi.org/10.1111/gcb.13248

Gastaldo, R. A. (2020). NOAA/WDS Paleoclimatology. *Boatyard (BOATYRD4) North American Plant Macrofossil Database*. https://www.ncdc.noaa.gov/paleo/study/7825

Geburek, T. (2014). Enzyklopädie der Holzgewächse: Handbuch und Atlas der Dendrologie. In B. Stimm, A. Roloff, U. M. Lang, & H. Weisgerber (Eds.), *Enzyklopädie der Holzgewächse: Handbuch und Atlas der Dendrologie*. Wiley. https://doi.org/10.1002/9783527678518

Gowan, E. J., Zhang, X., Khosravi, S., Rovere, A., Stocchi, P., Hughes, A. L. C., Gyllencreutz, R., Mangerud, J., Svendsen, J., & Lohmann, G. (2021). A new global ice sheet reconstruction for the past 80 000 years. *Nature Communications*, *12*(1), 1199. https://doi.org/10.1038/s41467-021-21469-w

Gros-Louis, M. C., Bousquet, J., Pâques, L. E., & Isabel, N. (2005). Species-diagnostic markers in Larix spp. based on RAPDs and nuclear, cpDNA, and mtDNA gene sequences, and their phylogenetic implications. *Tree Genetics and Genomes*, *1*(2), 50–63. https://doi.org/10.1007/s11295-005-0007-z

Herzschuh, U. (2020). Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography*, *29*(2), 198–206. https://doi.org/10.1111/geb.13018

Herzschuh, U., Birks, H. J. B., Laepple, T., Andreev, A., Melles, M., & Brigham-Grette, J. (2016). Glacial legacies on interglacial vegetation at the Pliocene-Pleistocene transition in NE Asia. *Nature Communications*, *7*(1), 11967. https://doi.org/10.1038/ncomms11967

Herzschuh, U., Böhmer, T., Chenzhi, L., Cao, X., Heim, B., & Wieczorek, M. (2021). *Global taxonomically harmonized pollen data collection with revised chronologies*. https://doi.org/doi.pangaea.de/10.1594/PANGAEA.929773

Islam, M. A., & Macdonald, S. E. (2004). Ecophysiological adaptations of black spruce (Picea mariana) and tamarack (Larix laricina) seedlings to flooding. *Trees - Structure and Function*, *18*(1), 35–42. https://doi.org/10.1007/s00468-003-0276-9

Johnston, W. F. (1990). Larix laricina (Du Roi) K. Koch tamarack. In R. M. Burns & B. H. Honkala (Eds.), *Silvics of North America* (pp. 141–151). Forest Service, United States Department of Agriculture.

Kanz, C., Aldebert, P., Althorpe, N., Baker, W., Baldwin, A., Bates, K., Browne, P., van den Broek, A., Castro, M., Cochrane, G., Duggan, K., Eberhardt, R., Faruque, N., Gamble, J., Diez, F. G., Harte, N., Kulikova, T., Lin, Q., Lombard, V., … Apweiler, R. (2005). The EMBL Nucleotide Sequence Database. *Nucleic Acids Research*, *33*(Database issue), D29-33. https://doi.org/10.1093/nar/gki098

Karlman, L. (2010). Genetic Variation in Frost Tolerance , Juvenile Growth and Timber Production in Russian Larches (Larix Mill .) - Implications for use in Sweden. In *Acta Universitatis Agriculturae Sueciae Sivestria*.

Kharuk, V. I., Ranson, K. J., & Dvinskaya, M. L. (2007). Evidence of Evergreen Conifers Invasion into Larch Dominated Forests During Recent Decades in Central Siberia. *Eurasian Journal of Forest Research*, *10*(2), 163–171. https://doi.org/10.1007/978-90-481-8641-9

Khatab, I. A., Ishiyama, H., Inomata, N., Wang, X.-R., & Szmidt, A. E. (2008). Phylogeography of Eurasian Larix species inferred from nucleotide variation in two nuclear genes. *Genes & Genetic Systems*, *83*(1), 55–66. https://doi.org/10.1266/ggs.83.55

Kistler, L., Montenegro, A., Smith, B. D., Gifford, J. A., Green, R. E., Newsom, L. A., & Shapiro, B. (2014). Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proceedings of the National Academy of Sciences*, *111*(8), 2937–2941. https://doi.org/10.1073/pnas.1318678111

Krestov, P. V., Ermakov, N. B., Osipov, S. V., & Nakamura, Y. (2009). Classification and phytogeography of larch forests of northeast Asia. *Folia Geobotanica*, *44*(4), 323–363. https://doi.org/10.1007/s12224-009-9049-6

Kruse, S., Epp, L. S., Wieczorek, M., Pestryakova, L. A., Stoof-Leichsenring, K. R., & Herzschuh, U. (2018). High gene flow and complex treeline dynamics of Larix Mill. stands on the Taymyr Peninsula (north-central Siberia) revealed by nuclear microsatellites. *Tree Genetics & Genomes*, *14*(2), 19. https://doi.org/10.1007/s11295-018-1235-3

Kullmann, L. (1998). Palaeoecological , Biogeographical and Palaeoclimatological Implications of Early Holocene Immigration of Larix sibirica Ledeb . into the Scandes Mountains , Sweden. *Global Ecology and Biogeography Letters*, *7*(3), 181–188. https://www.jstor.org/stable/2997373

Lenton, T. M., Rockström, J., Gaffney, O., Rahmstorf, S., Richardson, K., Steffen, W., & Schellnhuber, H. J. (2019). Climate tipping points — too risky to bet against. *Nature*, *575*(7784), 592–595. https://doi.org/10.1038/d41586-019-03595-0

LePage, B. A., & Basinger, J. F. (1995). The evolutionary history of the genus Larix (Pinaceae). *U.S. Dept. Agric., For Ser., Intermountain Res. Sta.*, *GTR-INT-31*(January 1995), 19–219.

Leys, B., & Carcaillet, C. (2016). Subalpine fires: the roles of vegetation, climate and, ultimately,

land uses. *Climatic Change*, *135*(3–4), 683–697. https://doi.org/10.1007/s10584-016-1594-4

Li, C., Postl, A. K., Böhmer, T., Cao, X., Dolman, A. M., & Herzschuh, U. (2021). Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0). *Earth System Science Data [Preprint]*. https://doi.org/10.5194/essd-2021-212

Liu, S., Stoof-Leichsenring, K. R., Kruse, S., Pestryakova, L. A., & Herzschuh, U. (2020). Holocene Vegetation and Plant Diversity Changes in the North-Eastern Siberian Treeline Region From Pollen and Sedimentary Ancient DNA. *Frontiers in Ecology and Evolution*, *8*(September), 1–17. https://doi.org/10.3389/fevo.2020.560243

Lozhkin, A., Anderson, P., Minyuk, P., Korzun, J., Brown, T., Pakhomov, A., Tsygankova, V., Burnatny, S., & Naumov, A. (2018). Implications for conifer glacial refugia and postglacial climatic variation in western Beringia from lake sediments of the Upper Indigirka basin. *Boreas*, *47*(3), 938–953. https://doi.org/10.1111/bor.12316

MacDonald, G. M., Kremenetski, K. V., & Beilman, D. W. (2008). Climate change and the northern Russian treeline zone. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1501), 2283–2299. https://doi.org/10.1098/rstb.2007.2200

Mamet, S. D., Brown, C. D., Trant, A. J., & Laroque, C. P. (2019). Shifting global Larix distributions: Northern expansion and southern retraction as species respond to changing climate. *Journal of Biogeography*, *46*(1), 30–44. https://doi.org/10.1111/jbi.13465

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, *17*(2), 240–248. https://doi.org/10.1101/gr.5681207

Moris, J. V., Vacchiano, G., Ravetto Enri, S., Lonati, M., Motta, R., & Ascoli, D. (2017). Resilience of European larch (Larix decidua Mill.) forests to wildfires in the western Alps. *New Forests*, *48*(5), 663–683. https://doi.org/10.1007/s11056-017-9591-7

Napier, J. D., Fernandez, M. C., Lafontaine, G., & Hu, F. S. (2020). Ice-age persistence and genetic isolation of the disjunct distribution of larch in Alaska. *Ecology and Evolution*, *10*(3), 1692–1702. https://doi.org/10.1002/ece3.6031

Obu, J., Westermann, S., Barboux, C., Bartsch, A., Delaloye, R., Grosse, G., Heim, B., Hugelius, G., Irrgang, A., Kääb, A. M., Kroisleitner, C., Matthes, H., Nitze, I., Pellet, C., Seifert, F. M., Strozzi, T., Wegmüller, U., Wieczorek, M., & Wiesmann, A. (2021a). ESA Permafrost Climate Change Initiative (Permafrost_cci): Permafrost active layer thickness for the Northern Hemisphere, v3.0. *NERC EDS Centre for Environmental Data Analysis*. https://doi.org/http://dx.doi.org/10.5285/29c4af5986ba4b9c8a3cfc33ca8d7c85

Obu, J., Westermann, S., Barboux, C., Bartsch, A., Delaloye, R., Grosse, G., Heim, B., Hugelius, G., Irrgang, A., Kääb, A. M., Kroisleitner, C., Matthes, H., Nitze, I., Pellet, C., Seifert, F. M., Strozzi, T., Wegmüller, U., Wieczorek, M., & Wiesmann, A. (2021b). ESA Permafrost Climate Change Initiative (Permafrost_cci): Permafrost extent for the Northern Hemisphere, v3.0. *NERC EDS Centre for Environmental Data Analysis*. https://doi.org/10.5285/6e2091cb0c8b4106921b63cd5357c97c

Oksanen, L. (1995). Isolated occurrences of spruce, Picea abies, in northernmost Fennoscandia in relation to the enigma of continental mountain birch forests. *Acta Botanica Fennica*, *153*(February), 81–92.

Osawa, A., & Zyryanova, O. A. (2010). Introduction. In *Permafrost ecosystems Siberian larch*

forests (pp. 3–13).

Osawa, A., Zyryanova, O. A., Matsuura, Y., Kajimoto, T., & Wein, R. W. (Eds.). (2010). *Permafrost Ecosystems* (Vol. 209). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9693-8

Pâques, L. E., Foffová, E., Heinze, B., Lelu-Walter, M.-A., Liesebach, M., & Philippe, G. (2013). Larches (Larix sp.). In L. E. Pâques (Ed.), *Forest Tree Breeding in Europe* (Vol. 25, Issue 1, pp. 13–122). Springer, Dortrecht. https://doi.org/10.1007/978-94-007-6146-9_2

Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M., & Eckert, A. J. (2018). RADseq approaches and applications for forest tree genetics. *Tree Genetics and Genomes*, *14*(3). https://doi.org/10.1007/s11295-018-1251-3

Pividori, M., Giannetti, F., Barbati, A., & Chirici, G. (2016). European Forest Types: tree species matrix. In J. San-Miguel-Ayanz, G. de Rigo, D. Caudullo, T. Houston Durrant, & A. Mauri (Eds.), *The European Atlas of Forest Tree Species* (pp. 34–35). Publication Office of the European Union, Luxembourg.

Pluess, A. R. (2011). Pursuing glacier retreat: Genetic structure of a rapidly expanding Larix decidua population. *Molecular Ecology*, *20*(3), 473–485. https://doi.org/10.1111/j.1365-294X.2010.04972.x

Polezhaeva, M. A., Lascoux, M., & Semerikov, V. L. (2010). Cytoplasmic DNA variation and biogeography of Larix Mill, in Northeast Asia. *Molecular Ecology*, *19*(6), 1239–1252. https://doi.org/10.1111/j.1365-294X.2010.04552.x

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org/

Roberts, D. R., & Hamann, A. (2015). Glacial refugia and modern genetic diversity of 22 western North American tree species. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1804). https://doi.org/10.1098/rspb.2014.2903

Rogers, B. M., Soja, A. J., Goulden, M. L., & Randerson, J. T. (2015). Influence of tree species on continental differences in boreal fires and climate feedbacks. *Nature Geoscience*, *8*(3), 228–234. https://doi.org/10.1038/ngeo2352

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017). HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods in Ecology and Evolution*, *8*(10), 1374–1388. https://doi.org/10.1111/2041-210X.12785

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S., & Herzschuh, U. (2021). Hybridization capture of larch (Larix Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Molecular Ecology Resources*, *21*(3), 801–815. https://doi.org/10.1111/1755-0998.13311

Schulze, E. D., Prokuschkin, A., Arneth, A., Knorre, N., & Vaganov, E. A. (2002). Net ecosystem productivity and peat accumulation in a Siberian Aapa mire. *Tellus B: Chemical and Physical Meteorology*, *54*(5), 531–536. https://doi.org/10.3402/tellusb.v54i5.16685

Schulze, E. D., Wirth, C., Mollicone, D., Von Lüpke, N., Ziegler, W., Achard, F., Mund, M., Prokushkin, A., & Scherbina, S. (2012). Factors promoting larch dominance in central Siberia: Fire versus growth performance and implications for carbon dynamics at the boundary of evergreen and deciduous conifers. *Biogeosciences*, *9*(4), 1405–1421. https://doi.org/10.5194/bg-9-1405-2012

Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D., & Willis, K. J. (2016). Sensitivity of global terrestrial ecosystems to climate variability. *Nature*, *531*(7593), 229–232. https://doi.org/10.1038/nature16986

Semerikov, V. L., & Lascoux, M. (1999). Genetic relationship among Eurasian and American Larix species based on allozymes. *Heredity*, *83*(1), 62–70. https://doi.org/10.1038/sj.hdy.6885310

Semerikov, V. L., Semerikova, S. A., Polezhaeva, M. A., Kosintsev, P. A., & Lascoux, M. (2013). Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (Larix sibirica): a range-wide analysis of cytoplasmic markers. *Molecular Ecology*, *22*(19), 4958–4971. https://doi.org/10.1111/mec.12433

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., … Meyer, M. (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, *356*(6338), 605–608. https://doi.org/10.1126/science.aam9695

Soininen, E. M., Gauthier, G., Bilodeau, F., Berteaux, D., Gielly, L., Taberlet, P., Gussarova, G., Bellemain, E., Hassel, K., Stenøien, H. K., Epp, L., Schrøder-Nielsen, A., Brochmann, C., & Yoccoz, N. G. (2015). Highly Overlapping Winter Diet in Two Sympatric Lemming Species Revealed by DNA Metabarcoding. *PLOS ONE*, *10*(1), e0115335. https://doi.org/10.1371/journal.pone.0115335

Sonstebo, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., Willerslev, E., Coissac, E., Rioux, D., Sannier, J., Taberlet, P., & Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, *10*(6), 1009–1018. https://doi.org/10.1111/j.1755-0998.2010.02855.x

Sugimoto, A., Yanagisawa, N., Naito, D., Fujita, N., & Maximov, T. C. (2002). Importance of permafrost as a source of water for plants in east Siberian taiga. *Ecological Research*, *17*(4), 493–503. https://doi.org/10.1046/j.1440-1703.2002.00506.x

Svenning, J.-C., & Skov, F. (2007). Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation? *Ecology Letters*, *10*(6), 453–460. https://doi.org/10.1111/j.1461-0248.2007.01038.x

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14. https://doi.org/10.1093/nar/gkl938

Tarasov, P., Williams, J. W., Andreev, A., Nakagawa, T., Bezrukova, E., Herzschuh, U., Igarashi, Y., Müller, S., Werner, K., & Zheng, Z. (2007). Satellite- and pollen-based quantitative woody cover reconstructions for northern Asia: Verification and application to late-Quaternary pollen data. *Earth and Planetary Science Letters*, *264*(1–2), 284–298. https://doi.org/10.1016/j.epsl.2007.10.007

Tchebakova, N. M., Rehfeldt, G. E., & Perfenova, E. I. (2010). From Vegetation Zones to Climatypes: Effects of Climate Warming on Siberian Ecosystems. In *Permafrost ecosystems Siberian larch forests* (Vol. 209, Issue 1925, pp. 427–446). https://doi.org/10.1007/978-1-4020-9693-8

Trugman, A. T., Medvigy, D., Anderegg, W. R. L., & Pacala, S. W. (2018). Differential declines in Alaskan boreal forest vitality related to climate and competition. *Global Change Biology*,

24(3), 1097–1107. https://doi.org/10.1111/gcb.13952

Uemura, S., Kanda, F., Isaev, A., & Tsuji, T. (1997). Forest structure and succession in southeastern Siberia. *Vegetation Science*, *14*(2), 119–127. https://doi.org/10.15031/vegsci.14.119

Urban, J., Rubtsov, A. V., Urban, A. V., Shashkin, A. V., & Benkova, V. E. (2019). Canopy transpiration of a Larix sibirica and Pinus sylvestris forest in Central Siberia. *Agricultural and Forest Meteorology*, *271*(January), 64–72. https://doi.org/10.1016/j.agrformet.2019.02.038

Wagner, S., Litt, T., Sánchez-Goñi, M.-F. F., & Petit, R. J. (2015). History of Larix decidua Mill. (European larch) since 130 ka. *Quaternary Science Reviews*, *124*, 224–247. https://doi.org/10.1016/j.quascirev.2015.07.002

Walter, H. (1985). *Vegetation of the earth and ecological systems of the geo-biosphere* (3rd editio). Springer.

Wang, Y., Heintzman, P. D., Newsom, L., Bigelow, N. H., Wooller, M. J., Shapiro, B., & Williams, J. W. (2017). The southern coastal Beringian land bridge: cryptic refugium or pseudorefugium for woody plants during the Last Glacial Maximum? *Journal of Biogeography*, *44*(7), 1559–1571. https://doi.org/10.1111/jbi.13010

Warren, E., de Lafontaine, G., Gérardi, S., Senneville, S., Beaulieu, J., Perron, M., Jaramillo-Correa, J. P., & Bousquet, J. (2016). Joint inferences from cytoplasmic DNA and fossil data provide evidence for glacial vicariance and contrasted post-glacial dynamics in tamarack, a transcontinental conifer. *Journal of Biogeography*, *43*(6), 1227–1241. https://doi.org/10.1111/jbi.12675

Wei, X.-X., & Wang, X.-Q. (2004). Recolonization and radiation in Larix (Pinaceae): evidence from nuclear ribosomal DNA paralogues. *Molecular Ecology*, *13*(10), 3115–3123. https://doi.org/10.1111/j.1365-294X.2004.02299.x

Wei, X. X., & Wang, X. Q. (2003). Phylogenetic split of \textit{Larix}: Evidence from paternally inherited cpDNA trnT-trnF region. *Plant Systematics and Evolution*, *239*(1–2), 67–77.

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Bråthen, K. A., Yoccoz, N., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*(7486), 47–51. https://doi.org/10.1038/nature12921

Williams, J. W., Grimm, E. C., Blois, J. L., Charles, D. F., Davis, E. B., Goring, S. J., Graham, R. W., Smith, A. J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A. C., Betancourt, J. L., Bills, B. W., Booth, R. K., Buckland, P. I., Curry, B. B., Giesecke, T., Jackson, S. T., Latorre, C., … Takahara, H. (2018). The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. *Quaternary Research*, *89*(1), 156–177. https://doi.org/10.1017/qua.2017.105

Wirth, C., Schulze, E. D., Lühker, B., Grigoriev, S., Siry, M., Hardes, G., Ziegler, W., Backor, M., Bauer, G., & Vygodskaya, N. N. (2002). Fire and site type effects on the long-term carbon and nitrogen balance in pristine Siberian Scots pine forests. *Plant and Soil*, *242*(1), 41–63. https://doi.org/10.1023/A:1020813505203

Zhang, N., Yasunari, T., & Ohta, T. (2011). Dynamics of the larch taiga-permafrost coupled system in Siberia under climate change. *Environmental Research Letters*, *6*(2). https://doi.org/10.1088/1748-9326/6/2/024003

Zimmermann, H., Raschke, E., Epp, L., Stoof-Leichsenring, K., Schirrmeister, L., Schwamborn, G., & Herzschuh, U. (2017). The History of Tree and Shrub Taxa on Bol'shoy Lyakhovsky Island (New Siberian Archipelago) since the Last Interglacial Uncovered by Sedimentary Ancient DNA and Pollen Data. *Genes*, *8*(10), 273. https://doi.org/10.3390/genes8100273

Zimmermann, Heike H., Harms, L., Epp, L. S., Mewes, N., Bernhardt, N., Kruse, S., Stoof-Leichsenring, K. R., Pestryakova, L. A., Wieczorek, M., Trense, D., & Herzschuh, U. (2019). Chloroplast and mitochondrial genetic variation of larches at the Siberian tundra-taiga ecotone revealed by de novo assembly. *PLOS ONE*, *14*(7), e0216966. https://doi.org/10.1371/journal.pone.0216966

Zimmermann, Heike Hildegard, Raschke, E., Epp, L. S., Stoof-Leichsenring, K. R., Schwamborn, G., Schirrmeister, L., Overduin, P. P., Herzschuh, U., Rosmarie Stoof-Leichsenring, K., Schwamborn, G., Schirrmeister, L., Paul Overduin, P., & Herzschuh, U. (2017). Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences*, *14*(3), 575–596. https://doi.org/10.5194/bg-14-575-2017

3 Manuscript II

Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest

Status

Published in *Molecular Ecology Resources,* 21(3), 801-815

Authors

Luise Schulte[1,2], Nadine Bernhardt[1], Kathleen Stoof-Leichsenring[1], Heike H. Zimmermann[1], Luidmila A. Pestryakova[3], Laura S. Ep[1], Ulrike Herzschuh[1,2,4]

Affiliations

[1]Alfred- Wegener- Institut, Helmholtz- Zentrum für Polar- und Meeresforschung, Forschungsstelle Potsdam, Potsdam, Germany
[2]Institut für Biochemie and Biologie, Universität Potsdam, Potsdam, Germany
[3]Institute of Natural Sciences, North- Eastern Federal University of Yakutsk, Yakutsk, Russia
[4]Institut für Geowissenschaften, Universität Potsdam, Potsdam, Germany

3.1 Abstract

Siberian larch (*Larix* Mill.) forests dominate vast areas of northern Russia and contribute important ecosystem services to the world. It is important to understand the past dynamics of larches in order to predict their likely response to a changing climate in the future. Sedimentary ancient DNA extracted from lake sediment cores can serve as archives to study past vegetation. However, the traditional method of studying sedimentary ancient DNA—metabarcoding—focuses on small fragments, which cannot resolve *Larix* to species level nor allow a detailed study of population dynamics. Here, we use shotgun sequencing and hybridization capture with long-range PCR-generated baits covering the complete *Larix* chloroplast genome to study *Larix* populations from a sediment core reaching back to 6700 years from the Taymyr region in northern Siberia. In comparison with shotgun sequencing, hybridization capture results in an increase in taxonomically classified reads by several orders of magnitude and the recovery of complete chloroplast genomes of *Larix*. Variation in the chloroplast reads corroborates an invasion of *Larix gmelinii* into the range of *Larix sibirica* before 6700 years ago. Since then, both species have been present at the site, although larch populations have decreased with only a few trees remaining in what was once a forested area. This study demonstrates for the first time that hybridization capture applied directly to ancient DNA of plants extracted from lake sediments can provide genome-scale information and is a viable tool for studying past genomic changes in populations of single species, irrespective of a preservation as macrofossil.

3.2 Introduction

Siberian forests are unique as they cover a vast area of about 263.2 million ha (Abaimov, 2010) dominated by a single genus of tree, the deciduous conifer larch (*Larix* Mill.). As the only extensive forest biome growing on continuous permafrost, it plays an important role for local communities and it provides critical ecosystem services in a global context including carbon stocks, climate feedbacks, permafrost stability, biodiversity and economic benefits (Herzschuh, 2019). It is

therefore important to understand how the genus and individual larch species have responded and will respond to changing climatic conditions.

Frequent natural hybridization between larch species makes it difficult to distinguish taxa, and the number of accepted species is still under discussion (Abaimov, 2010). This is one of the reasons why there is still little known about the population dynamics of Siberian larch species and the question remains of whether there have been migrations of larches in the current postglacial period.

Sedimentary ancient DNA (*sed*aDNA) from lakes can act as an archive of the past and has been demonstrated to be a valuable tool in the study of past vegetation history (Jørgensen et al., 2012; Parducci et al., 2017; Wang et al., 2017; Willerslev et al., 2003). Most *sed*aDNA studies focus on organellar DNA, as the higher copy number of organelles per cell compared with the nucleus allows a higher chance of retrieval. The metabarcoding approach (Taberlet et al., 2012) applied to DNA extracted from sediments is the most common, robust and fast technique to study past vegetation (Alsos et al., 2018; Niemeyer et al., 2017; Pansu et al., 2015). For ancient DNA of plants, a very short, but highly variable DNA fragment from the chloroplast genome is PCR-amplified out of the pool of DNA fragments and subsequently sequenced using high-throughput sequencing (Taberlet et al., 2007). However, the method is not suited to resolve population dynamics of single species, as metabarcoding markers used for ancient degraded samples must be very short while at the same time flanked by primers that are conserved across a larger taxonomic group. Therefore, their taxonomic resolution is, in most cases, insufficient to resolve closely related species (Sønstebø et al., 2010; Taberlet et al., 2007), let alone show subspecific variation.

Sequencing of the entire DNA extracted from ancient sediments, termed metagenomic shotgun sequencing, has been shown to provide information on the entire taxonomic composition of the sample (e.g. fungi, bacteria, archaea; Ahmed et al., 2018; Parducci et al., 2019; Pedersen et al., 2016). By sequencing complete DNA molecules, it is possible to authenticate ancient sequences versus modern contaminants by their specific postmortem DNA damage patterns towards the ends of the molecules (Ginolhac et al., 2011). As it is not restricted to a specific DNA fragment, it also allows the retrieval of many different loci belonging to single species provided they are sufficiently concentrated in the sample. A major drawback, however, is the immense sequencing effort that must be expended to achieve a sufficient overview of the DNA present in a sample. Most of the sequences retrieved from ancient environmental samples are not assignable to a specific taxon because available sequence databases are still limited, and most assigned sequences are assumed to be of noneukaryotic origin (Ahmed et al., 2018; Pedersen et al., 2016). Especially in the case of DNA extracted from lake sediments, the ratio of sequences assigned to terrestrial plants to total DNA-sequenced is expected to be extremely low (Parducci et al., 2019).

A way to overcome the limitations of shotgun sequencing is to enrich the DNA of the focal species in the samples via hybridization capture prior to sequencing. To do this, one can use short fragments of DNA of the species and target sites of interest as baits, to which the corresponding sites of interest in ancient DNA libraries are hybridized. This technique, originally developed for modern DNA, is commonly applied in ancient DNA studies, particularly for use on single specimens (Ávila-Arcos et al., 2011; Maricic et al., 2010) and with a focus on mammals, mostly using mitochondrial DNA (Carpenter et al., 2013; Dabney et al., 2013; Enk et al., 2016). Successful capture enrichment from sedimentary ancient DNA has been reported only a few times so far. Cave sediments (Slon et al., 2017) and permafrost samples (Murchie et al., 2019) were successfully enriched for a range of terrestrial organisms, while an attempt to capture ancient

mammalian DNA from lake sediments failed (Moore et al., 2019). Plants have received limited attention in ancient DNA studies (Parducci et al., 2017) but have also been targeted for enrichment. Kistler et al. (2014) captured chloroplast and nuclear DNA extracted from ancient gourd rinds, and Schmid et al. (2017) successfully applied hybridization capture on ancient DNA extracted from subfossil needles collected in lake sediments.

Ancient plant DNA recovered directly from lake sediments, which is commonly targeted in metabarcoding studies (Epp et al., 2015; Liu et al., 2020; Pansu et al., 2015), has, however, not yet been used as a target for the capture of larger genomic regions of specific target species. Beyond the retrieval of short fragments useful for species identification (as used in metabarcoding), it is not clear how complete the genomic record of plants in sediment cores is. This is also true for chloroplast DNA, which holds valuable ecological and adaptive information, through the genes for photosynthesis, and is widely used for taxonomic identification and phylogenetic analyses (CBOL Plant Working Group et al., 2009; Jansen et al., 2007; Shaw et al., 2007). In conifers as *Larix*, chloroplast DNA is paternally inherited via pollen (Szmidt et al., 1987), associated with a higher intraspecific gene flow and a lower rate of introgression than the maternally inherited mitochondrial DNA (Du et al., 2009). As a result, chloroplast DNA variation is more species-specific than mitochondrial DNA variation in this group (Du et al., 2009). Complete chloroplast genomes have, however, not yet been targeted for capture enrichment from lake sediment cores.

Here, we apply shotgun sequencing and a hybridization capture approach targeting the complete *Larix gmelinii* chloroplast genome to *sed*aDNA samples from a small lake in the Taymyr region of northeastern Siberia. The study site lies in the boundary zone of two larch species, *L. gmelinii* and *Larix sibirica*, with hybridization occurring between the boundary populations (Abaimov, 2010; Polezhaeva et al., 2010). It has been hypothesized for this region that a natural invasion of *L. gmelinii* into the range of *L. sibirica* occurred during the Holocene (Semerikov et al., 2013). The lake is situated in the treeline ecotone with scattered patches of *L. gmelinii* occurring in the area (Klemm et al., 2016). A sediment core of the lake has already been extensively studied using pollen analysis, DNA metabarcoding and mitochondrial variants (Epp et al., 2018; Klemm et al., 2016), making it an ideal site to study ancient larch population dynamics based on chloroplast DNA.

As a proof of concept, four samples were both shotgun-sequenced and enriched by hybridization capture for the chloroplast genome of *L. gmelinii*, to evaluate how well we can retrieve and assemble genome-scale data—here the complete chloroplast genome of *Larix*—from sedimentary ancient DNA. We demonstrate the successful enrichment by comparing taxonomically classified reads of the shotgun and hybridization capture data sets and evaluate the degree of coverage of the *Larix* chloroplast genome across the different annotated regions of the genome. This study presents the first successful recovery of complete chloroplast genomes from ancient lake sediments.

3.3 Methods

3.3.1 Sample material

Samples were obtained from a sediment core from lake CH12 (72.399°N, 102.289°E, 60 m a.s.l.) in the Khatanga region of the northern Siberian lowlands, located between the Taymyr Peninsula to the north and the Putorana Plateau to the south (Figure 1). The lake's position is in the northern part of the treeline ecotone and is currently surrounded by a vegetation of single-tree

tundra. Samples from this core (core ID 11-CH-12A) have already been analysed by Klemm et al. (2016) and Epp et al. (2018). Details of the chronology of the core are described in Klemm et al. (2016). Four new samples were chosen for the present study at depths/ages 121.5 cm/~6700 calibrated years before present (cal-BP), 87.5 cm/~5400 cal-BP, 46.5 cm/~1900 cal-BP and 2.5 cm/~60 cal-BP.

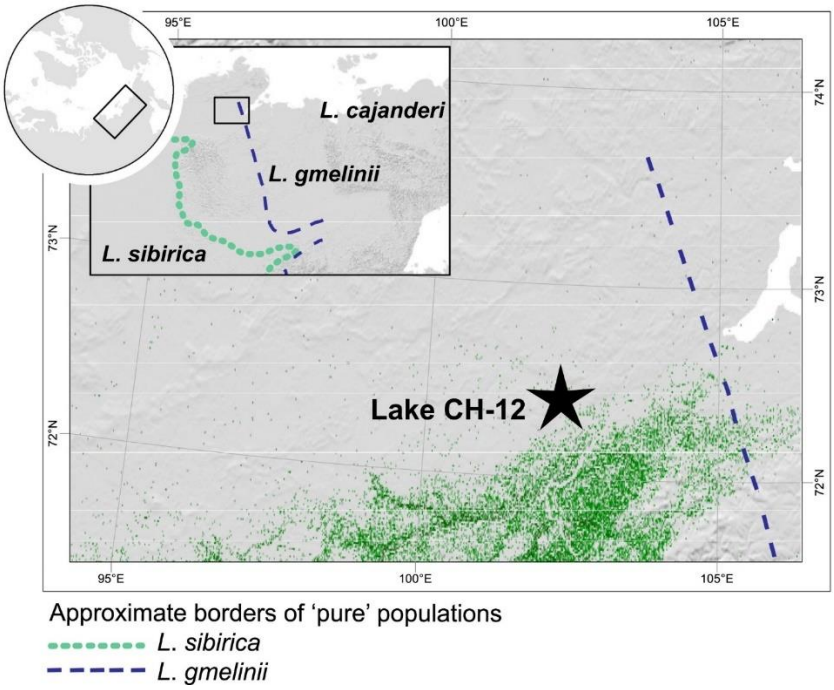

Fig. 1 Study area and position of lake analysed in this study (star). Broad-scale distribution of the different larch species *Larix sibirica*, *Larix gmelinii* and *Larix cajanderi* is shown in the upper left corner according to Abaimov (2010); Semerikov et al. (2013). The dotted and dashed lines indicate previously published boundaries between *L. sibirica* and *L. gmelinii* (Semerikov et al., 2013), with pure populations of *L. sibirica* occurring west of the dotted line and pure population of *L. gmelinii* occurring east of the dashed line. The green shading indicates the density of trees taller than 5 m in height, as published by (Hansen et al., 2013).

3.3.2 Laboratory work

Sampling, DNA extraction and library preparation

Core subsampling was performed as described in Epp et al. (2018). DNA extraction and library preparation were performed in a dedicated palaeogenetic laboratory at the Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research in Potsdam, Germany. DNA was extracted from 1.3–2.2 g of sediment using the DNeasy PowerMax Soil Kit (Qiagen, Germany) with the addition of 0.8 mg peqGOLD Proteinase K and 100 µl of 5 M dithiothreitol (VWR, Germany) in the initial lysis and homogenization step. After shaking the sample tubes horizontally for 10 min on a vortexer, they were incubated at 56°C overnight in a rotating incubator. The remaining protocol was conducted according to the manufacturer's instructions, and final extracts were eluted in 1.6 ml of solution C6. In total, six libraries with volumes of 50 µl each were prepared: four samples from the CH12 core, one extraction blank and one library blank. The extractions had DNA concentrations of 2.4 ng/µl (6700 cal-BP), 5.7 ng/µl (5400 cal-BP), 3.4 ng/µl (1900 cal-BP) and 6.7 ng/µl (60 cal-BP). 5 µl of each DNA extraction was used in the library preparation. Libraries were prepared following the single-stranded DNA library preparation protocol of Gansauge et al. (2017), which was specifically developed for ancient degraded DNA, with the following adjustment: as we had no access to a programmable shaking incubator in our ancient DNA laboratory, the ligation of the second adapter (CL53/CL73) was carried out in a

rotating incubator. The libraries were quantified with qPCR as described by Gansauge and Meyer (2013). We first prepared a standard for qPCR by amplifying a part of the pUC19 vector (New England BioLabs) with primers carrying P5 and P7 binding sites. The PCR contained 0.05 U Taq DNA Polymerase and 1x PCR buffer (Sigma-Aldrich), 0.25 µM CL105 and CL106, 1.25 mM dNTPs (Invitrogen) and 10 pg pUC19 DNA in a final volume of 100 µl and was carried out with the following cycling conditions: 5 min at 95°C, 30 cycles with 30 s at 95°C, 58°C and 72°C, each, followed by 5 min at 72°C. qPCR standards were purified using the MinElute PCR Purification Kit (Qiagen) following the manufacturer's recommendations. Standards were diluted in a series from 10^9 to 10^2 copies/µl. qPCR was carried out in 1× Maxima™ SYBR™ Green (Thermo Scientific), 0.2 µM IS7 and IS8 and 1 µl of the sample libraries diluted 1:20 with TET buffer in a total volume of 25 µl on a Rotor-GeneQ qPCR instrument (Qiagen). Cycler conditions were 10 min 95°C, 40 cycles of 30 s at 95, 60 and 72°C, each. Fluorescence was measured after each extension step.

The prepared libraries were used downstream for both shotgun sequencing and hybridization capture of chloroplast genomes.

Shotgun sequencing

Twenty-four µl of the prepared DNA library was amplified and indexed by PCR with 13 cycles as described in Gansauge and Meyer (2013) using the index primer sequences P5_1–P5_6 and P7_91–P7_96 with the following conditions: 1× AccuPrime™ Pfx reaction mix and 0.025 U AccuPrime™ Pfx polymerase (Invitrogen) and 0.4 µM forward and reverse primers in a volume of 100 µl with 2 min at 95°C followed by 13 cycles of 95°C for 15 s, 60°C for 30 s and 68°C for 1 min. PCR products were purified using the MinElute PCR Purification Kit (Qiagen) following the manufacturer's recommendations and eluting in 20 µL elution buffer. Fragment size distribution was checked on the 4200 TapeStation System (Agilent) using the D1000 ScreenTape Assay following the manufacturer's recommendations. Concentration of the libraries was estimated using the ds-DNA BR assay and the Qubit® 2.0 fluorometer (Invitrogen) using 1 µl of the purified libraries. The libraries were pooled in equimolar ratios to a final pool of 10 nM with the two blanks accounting for a molarity of 20% compared with the samples. The sequencing of the pool was performed by Fasteris SA Sequencing Service (Geneva, Switzerland) using a modified sequencing primer CL72 as described in Gansauge and Meyer (2013). The pool was sequenced on one lane of an Illumina HiSeq 2500 platform (2 × 125 base pairs (bp), high-output V4 mode).

Bait construction

Long-range PCR products covering the complete chloroplast genome of *Larix gmelinii* were generated using 18 primer pairs described in Zimmermann et al. (2019) with amplicon lengths of around 3000 to 10,000 base pairs. DNA was extracted from needles of an *L. gmelinii* individual collected in the Botanical Gardens of the University of Potsdam (Accession Number XX-0POTSD-3867, collected in East Asia, 1940). Approximately 40 needles were frozen using liquid nitrogen and homogenized using a FastPrep®-24 Homogenizer with 4 M/s for 50 s. The DNA was extracted using the Invisorb® Spin Plant Mini Kit following the manufacturer's recommendations. PCR amplification was conducted using the SequalPrep™ Long PCR Kit (Invitrogen), according to the manufacturer's cycling protocol instructions, and with specific annealing temperatures for each primer pair (see Zimmermann et al., 2019, Table S2). The PCR products were pooled in equimolar ratios in a volume of 130 µl and a final concentration of 10 ng/µl and sonicated using a Covaris M220 Focused-ultrasonicator (Covaris, USA) to a target peak of 350 bp with settings of peak incident power 50 W, duty factor 20%, cycles per burst 200 and treatment time 70 s. The fragment size and distribution were visualized with Agilent TapeStation (D1000 ScreenTape, Agilent Technologies). Fragment sizes ranged from 100 to 1000 bp with an average size of 370

bp. The complete sonicated pool was purified using the MinElute PCR Purification Kit (Qiagen), following the manufacturer's recommendations and eluted in 30 µl. The sheared pool was blunt-ended using the Fast DNA End Repair Kit (Thermo Scientific) with 1.3 µg DNA, 4 µl End Repair Enzyme Mix and 1× End Repair Reaction Mix in a volume of 100 µl for 10 min at room temperature. The blunt-ended DNA was purified using the MinElute PCR Purification Kit (Qiagen) following the manufacturer's recommendations and eluting in 15 µl elution buffer. Adapters Bio-T and B were hybridized as described in Maricic et al. (2010) and ligated to the blunt-ended DNA using the Rapid DNA Ligation Kit (Thermo Scientific) with 15 µl blunt-ended DNA, 1.25 µM hybridized adapters Bio-T/B, 1× rapid ligation buffer and 0.5 U T4 DNA ligase in a volume of 40 µl incubated for 15 min at room temperature. This reaction was again purified with the MinElute PCR Purification Kit (Qiagen) following the manufacturer's recommendations and eluting in 15 µl elution buffer. The concentration was estimated with the ds-DNA BR assay and the Qubit® 2.0 Fluorometer (Invitrogen) using 1 µl of the purified baits. Directly prior to the hybridization capture reaction, 500 ng of the purified baits was ligated to Dynal™ Dynabeads™ M-270 Streptavidin (Invitrogen) following the protocol described in Maricic et al. (2010).

Hybridization capture

The enrichment was done following the protocol of Maricic et al. (2010). Another 24 µl of prepared DNA libraries was PCR-amplified with 16 cycles with the same set of index primers, PCR products were purified, and fragment length and concentration were estimated in the same way as described for the shotgun samples. Libraries were pooled in equimolar amounts to a total of 2 µg, including the two blanks (extraction blank and library blank). The two blanks had a molarity of 20% compared with the samples. To prevent binding of library molecules to the adapter sequences, which would result in off-target capture, the adapter sequences were blocked prior to the capture experiment by blocking oligonucleotides. In addition to the blocking oligonucleotides BO3/4.P7.part1.F/R and BO5/6.P7.part2.F/R provided in the original protocol for single indexed libraries (Maricic et al., 2010), we split up the blocking oligonucleotides for the adapter sequences of P5 to account for the double indexed nature of the prepared libraries. The new blocking oligonucleotides had the following sequences: BO1.P5.part1.F: AATGATACGGCGACCACCGAGATCTACAC-phosphate, BO2.P5.part2.F: ACACTCTTTCCCTACACG-ACGCTCTTCCGATCT-phosphate, BO3.P5.part1.R: GTGTAGATCTCGGTGGTCGCCGTATCATT-phosphate and BO4.P5.part2.R: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-phosphate. The hybridization mixture of baits and adapter-blocked libraries was incubated at 65°C for 45 h in a rotating incubator. Subsequent to the hybridization capture experiment, 4 µl of the final elution was PCR-amplified with 18 cycles using the TruSeq DNA Nano Preparation Kit (Illumina Inc.), performed by Fasteris SA Sequencing Service (Geneva, Switzerland). The enriched library pool was sequenced by Fasteris SA Sequencing Service (Geneva, Switzerland) in the same way as described for shotgun sequencing.

3.3.3 Data analysis

Quality control, trimming and merging of reads

Quality control, trimming and merging were done for both data sets—shotgun and capture—in the same way. Demultiplexed fastq files, as obtained from the sequencing provider, were quality-checked using fastqc (v.0.7.11, Andrews, 2015) before and after trimming with trimmomatic (v.0.35, Bolger et al., 2014). The analyses performed by trimmomatic relied on a file containing the applied and common Illumina adapters, and the following parameter settings were used: remove adapters with maximum mismatch rate: 2, sliding window, window size: 4, average quality: 15, minimum quality to keep a base: 3 and minimum length to keep a sequence: 36

nucleotides. Unpaired reads were discarded because they contained only few reads with comparatively low base quality. Paired reads were merged using pear (v.0.9.10, Zhang et al., 2014). Merged and unmerged reads were treated separately in the following steps (Li et al., 2009).

Taxonomical classification

Reads from both the shotgun and capture data sets were classified using kraken2 (v. 2.0.7-beta, Wood et al., 2019) with a conservative confidence threshold (--confidence 0.8) against the nonredundant nucleotide database (nt) from NCBI (ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/-nt.gz; downloaded in May 2019) and the NCBI taxonomy (retrieved via the kraken2-build command in June 2019). This classification was used for description and comparison of the shotgun and capture data sets. Additionally, reads were classified using a custom database of 4,919 complete plant chloroplast genomes (downloaded from NCBI in July 2019; see list of accessions in Supporting Information). In this second classification, kraken2 was run with default parameters (--confidence 0) to allow for the retrieval of variation. Reads classified with the chloroplast database as genus *Larix* or below were extracted and used for further analysis (commands provided in Supporting Information).

Alignment

Alignments against an *L. gmelinii* chloroplast reference genome were made using three data sets: the complete capture data set, a subset of this data set containing only *Larix*-classified reads (as described above) and the same subset of the shotgun data set. As reference, the chloroplast genome of an *L. gmelinii* individual from the Taymyr region was used (NCBI Accession No.: MK468637.1). Reads were mapped using bwa aln algorithm (v.0.7.17-r1188, Li & Durbin, 2009) with the settings -l 1024 (disabling a seed region) -o 2 (maximum number of gap openings) -n 0.001 (fraction of missing alignments given 2% uniform base error rate) to ensure relaxed mapping. Duplicate reads were removed after mapping using picard markduplicates (v. 2.20.2-SNAPSHOT, Broad Institute, 2019) for merged reads and samtools markdup for unmerged reads. BAM files of merged and unmerged reads were combined to one file per sample with samtools' command 'merge' (v. 1.9, Li et al., 2009).

The comparison of alignments of the two *Larix*-classified subsets, shotgun and capture, was used to evaluate the degree of enrichment obtained by the hybridization capture experiment. The alignment of the subset of *Larix*-classified capture reads in comparison with the alignment of the complete capture data set was used to see whether gaps in the alignment are a result of missing sequences in the sample or a result of the impossibility of unambiguously assigning certain sequences to *Larix* using a lowest common ancestor approach.

Coverage of the Larix chloroplast genome at different annotated functions

We further explored which functional regions of the chloroplast genome are most affected by the use of reads classified with a lowest common ancestor approach. For this, the full capture data set and *Larix*-classified capture subset were compared to quantify the difference in coverage in different annotated regions of the genome. The annotation of the chloroplast genome was adopted from the published annotation file on NCBI for the used reference genome (Acc.: MK468637.1). Coverages for each sample were obtained using samtools depth (option –a) (Li et al., 2009) and summed up at each position across the samples. Box plots were made using R with ggplot2 (R Core Team, 2013).

Assessment of ancient DNA damage patterns

To authenticate the *Larix*-classified reads mapped against the *Larix* chloroplast genome as genuine ancient DNA, damage patterns were assessed using mapDamage (v. 2.0.8, Jónsson et al., 2013) with the options --rescale (to downscale quality score for misincorporations likely due to ancient DNA damage) and --single-stranded (for a single-stranded protocol). The alignments for merged paired reads and unmerged paired reads were processed individually. For damage pattern assessment of unmerged paired-end reads, the reads were mapped in two rounds as single-end reads to the reference (Acc.: MK468637.1) as described above and analysed with mapDamage to not confound the 5' and 3' damage patterns. To see whether the mapped reads are short, as expected for ancient DNA, read length distribution was assessed with mapDamage for overlapping merged reads and with geneious (v. 2019.2, Biomatters, 2019).

Assignment of Larix bases to L. sibirica or L. gmelinii

To distinguish between the two *Larix* species, *L. gmelinii* and *Larix sibirica* and determine their temporal occurrence, we performed multiple comparisons of the 12 complete chloroplast genomes of *L. gmelinii* (Accession Numbers. MK468630–39, MK468646 and MK468648) available on NCBI and the one available *L. sibirica* genome (Accession Number: MF795085.1) and classified species-specific single nucleotide polymorphisms (SNPs) and insertions and deletions (indels). The alignment was performed using bwa mem (Li & Durbin, 2009) with default parameters. SNPs and indels were called using bcftools mpileup (option -B) and call (option -mv) (v. 1.9, Li et al., 2009). Variable sites that differ between the two species were selected, but sites with variants occurring not only in the *L. sibirica* reference but also in one of the available *L. gmelinii* references were excluded. In total, 294 positions were determined as occurring only in *L. sibirica*. For each of these 294 positions, the above-produced alignments (ancient sample reads against *L. gmelinii* reference) were analysed with regard to whether the reads carried the same variant as *L. gmelinii* or *L. sibirica* or whether they had a different SNP or indel at each of these positions. First, a region file of the 294 positions was produced using vcf2bed (Neph et al., 2012) using the sorted vcf file of positions unique to *L. sibirica* as input. The analysis was carried out using bam-readcount (Larson & Abott, 2016) with the option -b 30 considering only bases with base quality above 30. The output of bam-readcount was then analysed with a custom python script (see Supporting Information), which classified the count of reads assigned to *L. gmelinii*, *L. sibirica* or 'other' for every position. The analysis was repeated in the same way with the *L. sibirica* chloroplast genome as reference.

3.4 Results

3.4.1 Overview of the shotgun and hybridization capture data sets

The shotgun data set

Shotgun sequencing of DNA extracted from ancient lake sediment samples resulted in about 424 million (M) read pairs for the four samples, and 24 M reads for the extraction and library blank (Table S1). After trimming and filtering, 62.6% of the sample reads remained for the analysis, of the blanks 0.2% remained. Comparable results were obtained by Ahmed et al. (2018) who retained 52% of shotgun sequenced sedimentary DNA after trimming and quality control. Eighty-two per cent of the sample reads overlapped and were merged.

Using kraken2 with the nt database and a confidence threshold of 0.8, 0.3% of the quality control passed (QC) shotgun reads could be classified. The majority of sample reads were classified as

Bacteria (62.6%) and Eukaryota (23.4%). Across all samples, 2.8 thousand (k) reads were classified as *Larix* (Figure 2, Table S2).

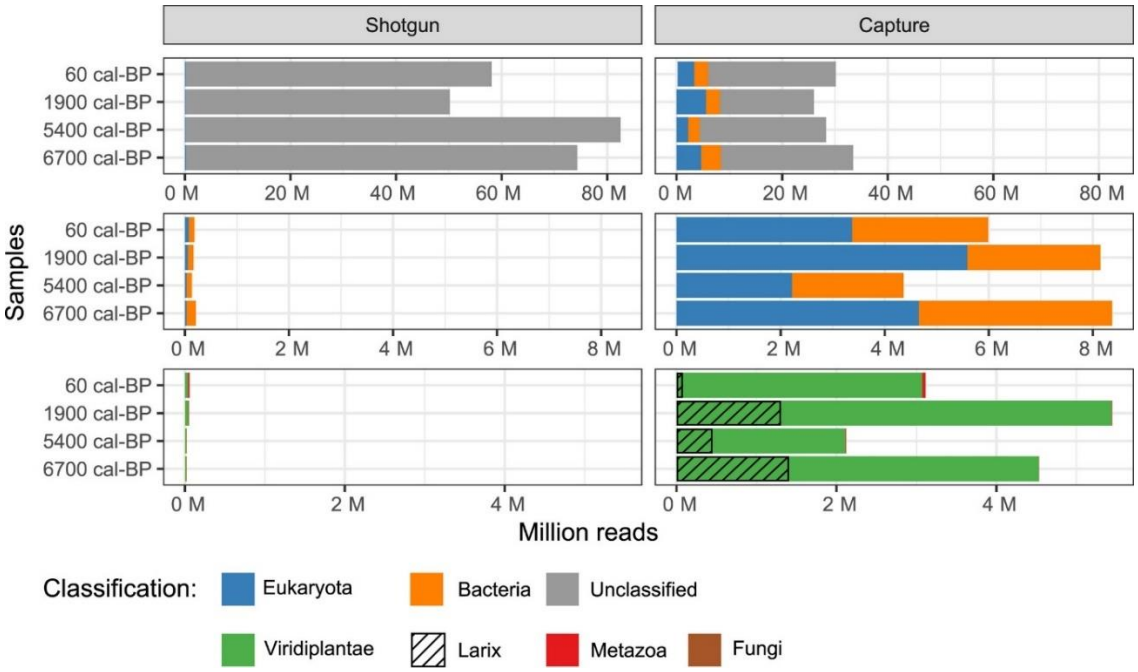

Fig. 2 Sequence reads classified with kraken2 at high confidence against NCBI nonredundant nucleotide database (nt). Left: shotgun data set, right: hybridization capture data set. Upper graphs: classification at domain level. Middle graphs: as above with unclassified reads excluded. Lower graphs: classification of Eukaryota at kingdom level and classified as *Larix*. Note differences in scale from upper to lower panels

When classifying against the custom chloroplast database using kraken2 with default confidence, 0.16% of the QC shotgun sample reads could be assigned to Viridiplantae (Table S3). In the samples, 3.6 k reads were classified as *Larix* and used in further analyses. In the blanks, no read was assigned to *Larix* with neither of the databases or thresholds. Therefore, they were not considered further in the analysis.

The hybridization capture data set

The sequencing of the hybridization capture experiment resulted in approximately 192 M paired-end reads for the four samples and 10 M reads for the blanks. After trimming and quality filtering, 66% were kept from the samples and 0.3% were kept from the blanks. About 91% of sample reads and 95% of blank reads overlapped and were merged (Table S1).

Classification with kraken2 using the nt database with a confidence threshold of 0.8 could classify 28% of the capture sample reads. Of the classified sample reads, the majority was classified as Eukaryota (44%) and more specifically Viridiplantae (43%). Three M reads were assigned to *Larix* (8.9% of classified reads, Figure 2).

Classification against the custom chloroplast database using kraken2 with default confidence resulted in 46 M (36.5%) of the capture sample reads assigned to Viridiplantae. Of the assigned reads, 9.2% were classified as *Larix* (4.2 M reads).

In the two blanks, six and eight reads (classification with nt and chloroplast database, respectively) were assigned to *Larix*. When factoring in that the blanks were sequenced only with one fifth of a share compared with the samples, the number of assigned reads to *Larix* is still many orders of magnitude smaller in the blanks than in the samples (the lowest number of *Larix*-

classified reads in a sample is 57 k reads, in the extraction blank it is 35 with the applied correction factor). Therefore, the blanks were not considered further in the analysis.

A comparison of the shotgun and capture data sets shows that 46.6- to 155.8-fold more reads were assigned to Eukaryota in the capture data set. Within the Viridiplantae, enrichment ranged from 77.8- to 236.9-fold enrichment of captured data in respect to shotgun data. The number of *Larix*-classified reads per sample corresponds to an increase of around 800- to 1160-fold compared with the shotgun data. These reads were filtered for PCR duplicates when aligning to the *Larix* chloroplast genome. Comparing the aligned, deduplicated reads of both shotgun and capture data sets, the enrichment ranged from 6.4- to 16.2-fold (Table S4).

3.4.2 Ancient DNA authenticity

A mapDamage analysis (Jónsson et al., 2013) was applied to the alignment files of *Larix*-classified reads aligned to the *Larix gmelinii* chloroplast genome. The overlapping merged reads for the three ancient samples (from 1900, 5400 and 6700 cal-BP) show a clear increase in C-to-T substitutions at both ends with a greater pronunciation at the 5’ ends. A clear increase in substitution rate with age is visible (Figure S1). The unmerged paired-end reads show comparable C-to-T substitution rates for the forward reads at the 5’ ends and for the reverse reads at the 3’ ends (Figure S2). The youngest sample (60 cal-BP) does not show any clear pattern of substitution rates for merged or unmerged paired-end reads. The length of sequencing reads was between 50 and 340 bp (mean 92 ± 36.7 bp) for all samples in the alignment of merged and unmerged paired reads against the *Larix* chloroplast genome, including the gap of unmerged reads (Figures S1 and S2).

3.4.3 Retrieval of the *Larix* chloroplast genome

To evaluate the retrieval of *Larix* chloroplast genome sequences, alignments against a reference were made with three data sets: the complete capture data set and the *Larix*-classified subsets of capture and shotgun data sets. The alignment of the *Larix*-classified capture reads resulted in a near-complete retrieval of the *Larix* chloroplast genomes for all samples except for the most recent one. The coverage of the chloroplast genome declined from the oldest to the most recent sample. At a minimum of onefold coverage, 91.4% of sample ‘6700 cal-BP’, 80.3% of sample ‘5400 cal-BP’, 85.1% of sample ‘1900 cal-BP’ and 14.3% of sample ‘60 cal-BP’ are covered. The mean coverage of the samples is 21.0x ± 14.2x (6700 cal-BP), 5.0x ± 4.5x (5400 cal-BP), 6.6x ± 5.9x (1900 cal-BP) and 0.3x ± 0.7x (60 cal-BP).

When aligning the complete QC-passed capture data set without prior taxonomic assignment, at onefold coverage the complete chloroplast genome (99.99–99.43%) could be retrieved from all samples except the youngest (60 cal-BP) where only 43% were covered at onefold (Table S5).

In the alignments, the coverage is not equal across the different annotated regions. When aligning the *Larix*-classified reads, the coverage is highest for inverted repeats and lowest for ribosomal RNA (Figure 3, dark shaded colours). In the same data set, the coverage is, on average, higher for intergenic regions, pseudogenes and conserved open reading frames (ORFs), than for protein-coding genes. When aligning the complete capture data set against the same reference, the coverages of the different annotated regions show a different pattern: highest coverage is at the ribosomal RNA, followed by the photosystem complex coding region and the inverted repeats.

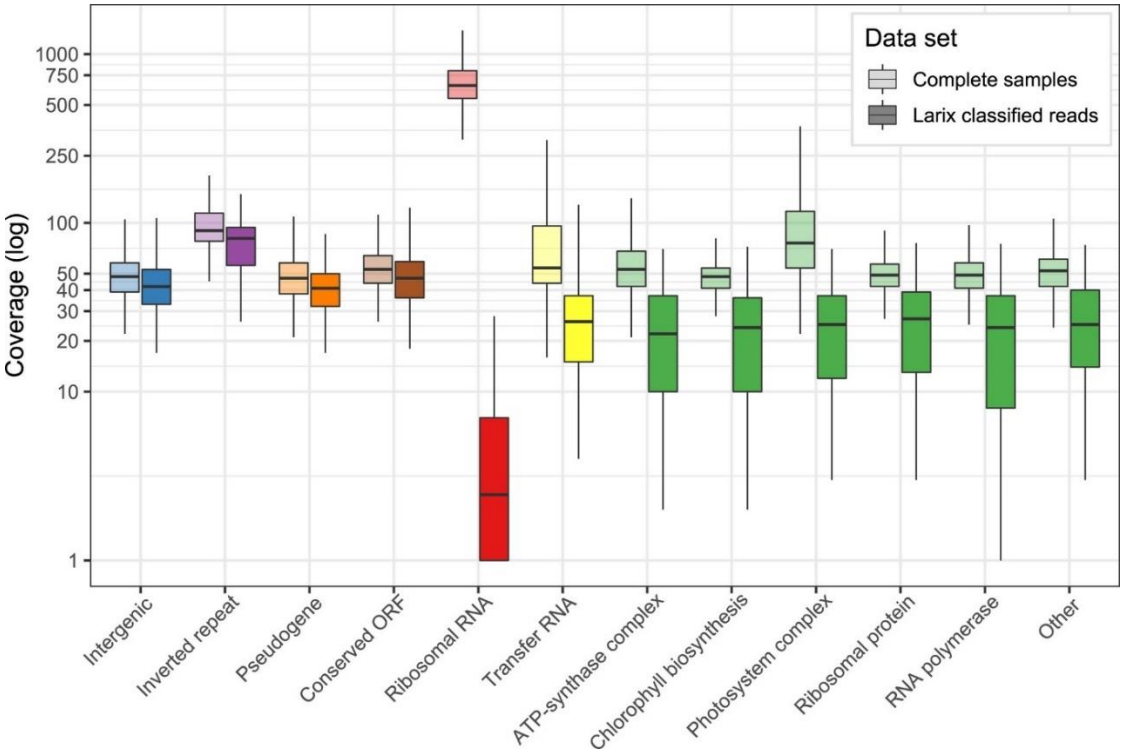

Fig. 3 Read coverage of all four samples in the alignment of the capture data set against the *Larix gmelinii* chloroplast genome. Coverage is shown according to the functional annotation. Light colour shades: complete quality-filtered read sets of the capture data set; dark colour shades: *Larix*-classified reads of the capture data set. Green shades: protein-coding genes. ORF, open reading frame of unknown function. Outliers not shown.

Considering the 294 sites that differ between the two reference genomes of *L. gmelinii* and *Larix sibirica*, 95.5% of all reads in all samples carry *L. gmelinii* specific variants, and 4% of the reads carry *L. sibirica* variants (Figure 4). Almost no reads (0.4%) carry neither of the two species-specific variants (‘other’).

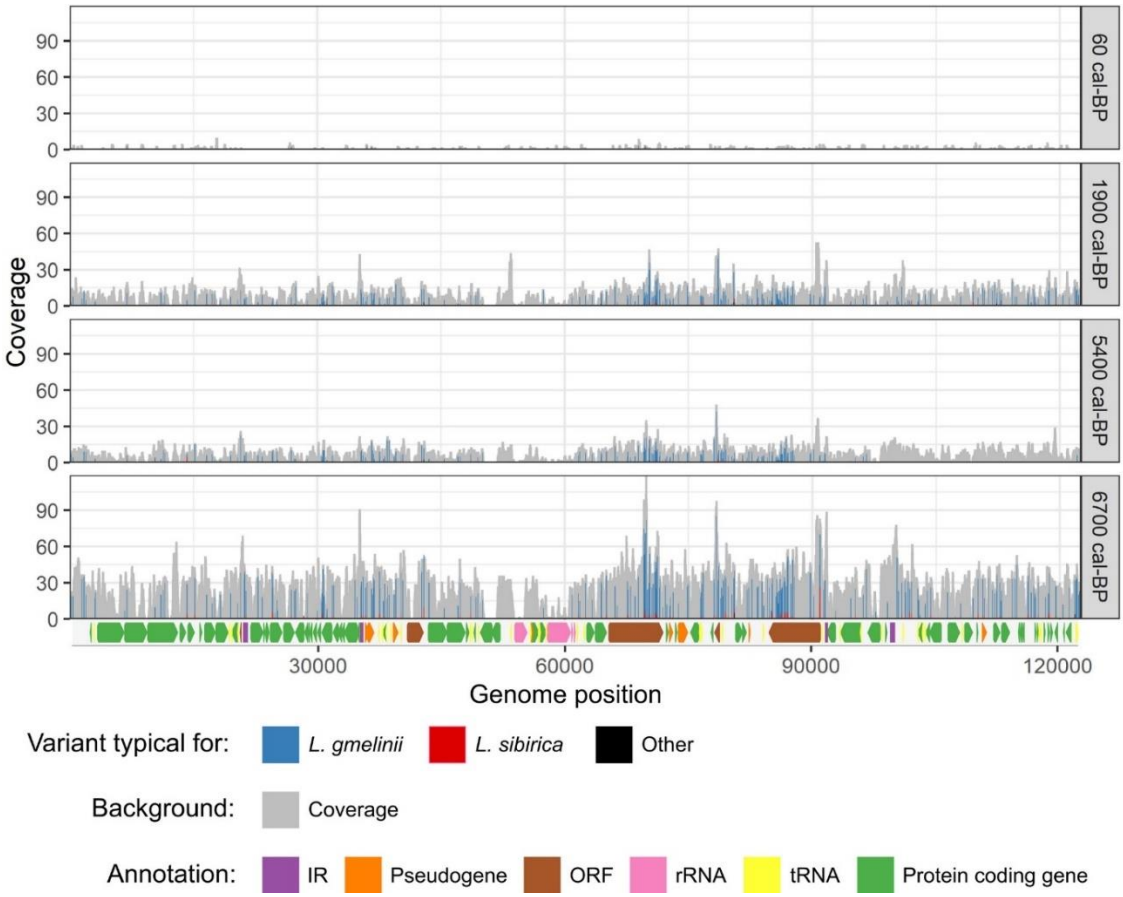

Fig. 4 Alignment of *Larix*-classified reads from hybridization capture data set against the *Larix* chloroplast reference genome. The coverage per position is depicted in grey. For the 294 sites, variable between the *Larix gmelinii* and *Larix sibirica* chloroplast genomes, colour indicates how many reads correspond to the variants found in each of the two species or if a read contained a variants found in neither of the two species ('other'). Key for the coloured arrows annotating the functional groups. IR, inverted repeat; ORF, open reading frame of unknown function; rRNA, ribosomal RNA; tRNA, transfer RNA.

Between 0.3% and 51% of the analysed positions contained at least one read which was classified as *L. sibirica*, with the highest percentage detected at 6700 cal-BP and the lowest percentage detected at 60 cal-BP (Table S6). The ratio of *L. sibirica* variants over all positions and reads varied from 5% (6700 cal-BP) to 1.6% (5400 cal-BP). Most of the variation between the two *Larix* species lies in the intergenic region and in conserved ORFs of unknown function (Figure 4). When using an *L. sibirica* chloroplast genome as reference, the vast majority of reads (79.6%) still carry variants assigned to *L. gmelinii*, while 19.1% of the reads are classified as *L. sibirica* and 1.3% as 'other' (Table S6).

3.5 Discussion

Ancient DNA from lake sediments constitutes a valuable resource to investigate the response of populations to past environmental changes. Previous studies using metabarcoding or shotgun sequencing have not yet explored the full potential of this resource. Here, we applied shotgun sequencing and hybridization capture using PCR-generated baits of the *Larix* chloroplast genome, to retrieve complete chloroplast genomes and study past changes in the population history of larches in northern Siberia.

3.5.1 Taxonomic classification—conservative approach results in low numbers of assignment

In the shotgun data set, only 0.3% of quality-filtered reads could be classified against the nt database. This is a very low number compared with other studies (Ahmed et al., 2018; Slon et al., 2017). In our analysis, the parameter setting in the bioinformatic approach had a high impact on the rate of classification. We used kraken2 (Wood et al., 2019), a new version of kraken, which is a particularly conservative tool compared with others, reporting less false-positive but also less true-positive hits than others tools, even with default values (Harbert, 2018). We used it with the very high-confidence threshold of 0.8, which calculates a score for each taxonomic level and can be set between 0 (most sensitive) and 1 (most specific). We decided upon this high-confidence setting, as we found it gives the best results in terms of vegetation composition based on our knowledge of the vegetation history (Epp et al., 2018), but with the consequence of very low overall assignment rates. Indeed, when we use the default confidence threshold of kraken2, we could assign 10%–16% of the reads. However, more lenient classification causes a reinforcement of the database bias: few deeply sequenced taxa are more likely to be assigned than the majority of shallowly or fragmentarily sequenced taxa (Parducci et al., 2017).

3.5.2 Target enrichment success—*Larix* reads increased by orders of magnitude along with other taxonomic groups

Hybridization capture resulted in an increase in taxonomically classified reads by orders of magnitude, especially with respect to the ratio of classified reads (0.3% to 28%), and also in absolute numbers of assigned reads (800 K reads to 35 M reads). These results show, for the first time, that DNA capturing of whole chloroplast genomes is effective even for DNA libraries that contain DNA from diverse origins and low on-target rates such as DNA from ancient lake sediments.

The number of reads classified as genus *Larix* using the nt database increased 800- to 1600-fold from shotgun to hybridization capture data set. However, in all samples there was a high level of PCR duplicates due to the high number of amplifications after capturing and possible oversequencing of reads. Future projects could reduce the number of PCR cycles and increase the number of samples pooled together to reduce duplicates.

The deduplication of reads when mapping against the reference genome reduced the enrichment to six- to 16-fold. This is in the range of results from enrichment studies performed on bone material (Ávila-Arcos et al., 2015; Carpenter et al., 2013), although higher enrichment ranges have also been reported (Mohandesan et al., 2017). Here, most likely the sample material plays a role, as the lake sediments contained a low content of the target *Larix* chloroplast DNA together with a high sequence diversity. Nevertheless, the target enrichment rate could possibly be increased, for example, by capturing at higher temperatures, by using a touch-down approach, by doing two consecutive rounds of capture or by using RNA instead of DNA for baits (Carpenter et al., 2013; Li et al., 2013; Paijmans et al., 2016; Peñalba et al., 2014).

Along with *Larix*, many other plant taxa were identified with a high number of reads classified as Viridiplantae (Figure 2). This can be explained by (a) chloroplast genomes of land plants and green algae share essentially the same set of protein-coding genes and ribosomal RNAs and differing mainly in the presence/absence of introns and repeats (Green, 2011), and (b) DNA libraries built from lake sediments contain a very complex mix of sequences of all domains of life. This complex mixture corresponds to a higher sequence divergence than mixtures from pooled individuals from one taxonomic order, which have previously been used to measure the capability of capturing sequences highly diverged from the baits (Paijmans et al., 2016; Peñalba et al., 2014).

This capability of baits capturing fragments from diverged taxa could be potentially refined and used to study wider taxonomic groups of interest in ancient lake sediments.

Apart from Viridiplantae sequences, the capture data set also contains considerable amounts of reads classified as bacteria encompassing different groups of bacteria. This can likewise be explained by the presence of highly conserved gene sequences in the chloroplast genome, which are also shared by bacteria. In particular, the chloroplast genome contains sequences coding for the 16S ribosomal RNA, which is widely used as a phylogenetic marker. Such marker genes are present in high amounts in the nt database so they are very likely to be taxonomically assigned.

3.5.3 Complete retrieval of ancient *Larix* chloroplast genomes

In the capture data set, complete chloroplast sequences could be retrieved from the three oldest samples (>99% with onefold coverage). Comparing the alignment of all reads to the chloroplast genome with the subset of only *Larix*-classified reads, coverages are most distinct for protein-coding genes, especially genes coding for the photosystem complex, transfer RNA and ribosomal RNA. The coverage of the complete data set in these regions is higher, in the case of ribosomal RNA, even orders of magnitude higher, whereas the same regions are low in coverage or even contain gaps in the alignment of *Larix*-classified reads. These coding regions are highly conserved across taxa (Green, 2011), and as the short reads can also be attributed to other organisms, they are classified to a higher taxonomic rank than *Larix*. The gaps in the alignment of *Larix*-classified reads can therefore be attributed to the conservative bioinformatic approach of only including unambiguously classified *Larix* reads and are not the result of missing sequences in the sample.

Analysis of DNA damage patterns in the *Larix* chloroplast alignment revealed C-to-T substitution rates typical for ancient DNA (Figures S1 and S2). Typical for the preparation of single-stranded libraries, these substitutions could be observed both at the 3’ end and at the 5’ end of the molecules (Gansauge & Meyer, 2013). C-to-T substitution rates increased with sample age, in line with previous observation (Pedersen et al., 2016; Sawyer et al., 2012). Mapped read and insert lengths ranged from 50 to 340 bp (mean 92 bp), showing the short fragment length typical for ancient DNA (Green et al., 2008).

3.5.4 *Larix sibirica* variants present over time

When comparing the ancient reads to chloroplast reference genomes from *Larix gmelinii* and *L. sibirica*, the great majority of reads carry *L. gmelinii* variants with a low frequency of *L. sibirica* variants in all four samples. In contrast, the analysis of one mitochondrial marker derived from the same core by Epp et al. (2018) showed a mixture of mitotypes typical for each of the respective species, with relatively high rates of the *L. sibirica* mitotype—except for the most recent sample, which showed clear dominance of the *L. gmelinii* mitotype—pointing to a co-occurrence of both species throughout most of the sediment core. In the genus *Larix*, chloroplasts are predominantly inherited paternally (Szmidt et al., 1987) whereas mitochondrial DNA is inherited maternally (DeVerno et al., 1993), a phenomenon which has been reported for almost all members of the conifers (Neale & Wheeler, 2019). This biparental inheritance results in different rates of gene flow and subsequently asymmetric introgression patterns (Du et al., 2009; Petit et al., 2004). Simulations (Currat et al., 2008) and molecular studies on a range of Pinaceae (Du et al., 2009, 2011; Godbout et al., 2012) showed that the seed-transmitted mitochondria, which experience little gene flow, introgress more rapidly than the pollen dispersed chloroplasts, which experience high gene flow. A second finding of these studies is that introgression occurs asymmetrically from the resident species into the invading species. An expected result of

introgression is therefore a population carrying mitotypes of the former local species and chlorotypes of the invader.

In the case of the population history of *L. sibirica* and *L. gmelinii* in their contact zone, Semerikov et al. (2013) found evidence for the asymmetric introgression of *L. sibirica* mitotypes in a population carrying only *L. gmelinii* chlorotypes, confirming the natural invasion of *L. gmelinii* into the range of *L. sibirica*. Here, we corroborate these findings with a distinct discrepancy between relatively high rates of *L. sibirica* mitotypes as reported before (Epp et al., 2018) and low rates of *L. sibirica* in the chloroplast reads found in this study.

This points to an invasion of *L. gmelinii* in a former population of *L. sibirica* prior to the date of our oldest sample (6700 cal-BP). Further evidence in support of this scenario is found in the results from a lake sediment core 250 km southwest of the study site (Epp et al., 2018), where samples reaching back to 9300 cal-BP show exclusively *L. sibirica* mitotypes, before they were gradually replaced by the *L. gmelinii* mitotype.

Our study shows that by capturing the complete chloroplast genome, we achieve a high resolution and can detect species-specific variants even at low frequencies. Further studies should also include mitochondrial sequences in the target enrichment to collect data from several markers or potentially the complete mitochondrial genome. By combining the two organelle genomes in a hybridization capture experiment, it would be possible to study hybridization and introgression events in detail, which would help to deepen our understanding of population dynamics over long time scales.

3.5.5 Larch forest decline over the last 7000 years

When looking at the overall retrieval of *Larix* reads among the samples, most reads could be recovered from the oldest sample of around 6700 cal-BP and the least number of reads in the most recent sample. This suggests a decline of *Larix* forests in the surroundings of the lake at the scale studied. However, a direct link between the amount of DNA recovered and the true vegetation at a time is difficult, as DNA preservation and recovery in lake sediments can be impacted by a variety of factors in the taphonomic and analytical processes (Giguet-Covex et al., 2019). Yet, our results are in line with the findings of Klemm et al. (2016) and Epp et al. (2018), who describe a general vegetation turnover from larch forest to an open tundra with only sparse *Larix* stands during the last 7000 years at this site. This gradual change in vegetation in response to late Holocene climate deterioration has also been inferred by various studies (Andreev et al., 2004; MacDonald et al., 2000, 2008; Pisaric et al., 2001) and is in correspondence with the reconstructed global cooling trend of the middle to late Holocene (Marcott et al., 2013).

3.6 Conclusion

Siberian larch forest covers vast areas of northern Asia with *Larix* as the only tree-forming species. Lake sediments containing ancient DNA constitute an archive to answer the question of how larch forests respond to changing climate, but the low amount of target DNA in combination with a complex mixture of sequences makes them challenging material to study the population dynamics of a specific species. Here, we have shown the success of hybridization capture of complete chloroplast genomes from 6700-year-old lake sediments originating from northern Siberia. Shotgun sequencing of *sed*aDNA prior to enrichment showed that, depending on the cautiousness of the bioinformatic approach, only very low rates of reads can be securely assigned to taxa even at the domain level. By using PCR-generated baits covering the whole chloroplast of *Larix* for hybridization capture, we could achieve increases by several orders of magnitude of

assignable reads. The enrichment of *Larix* reads was most distinct, but plant DNA in general was also enriched. With ancient DNA from lake sediments, hybridization capture thus offers the potential of not only analysing the target species in depth, but also studying the taxonomic diversity of the sample in a similar way to traditional molecular barcoding approaches. The method is more costly than the metabarcoding approach, and computationally more complex, but brings the advantage of not being restricted to a specific fragment length or by primer binding sites, and the possibility of authentication of ancient DNA. Similar experiments can be done for any species of interest for which PCR products are available and also on samples from complex environments with very low rates of target DNA. Future studies of plant biodiversity changes could focus on conserved coding regions of a set of diverged species to capture a more complete picture of the past vegetation. The analysed *Larix* reads confirm a general larch forest decline over the last 6700 years. Low rates of *Larix sibirica* variants in proportion to *Larix gmelinii* variants in the chloroplast could point to an invasion of *L. gmelinii* into *L. sibirica* populations before 6700 years ago. This study represents the first demonstration of hybridization capture from ancient DNA derived from lake sediments without macrofossils. Our results open the way for large-scale palaeogenomic analyses of ancient population dynamics using lake sediment cores.

3.7 Acknowledgments

We thank our Russian and German colleagues who helped in fieldwork in 2011 to obtain the samples. Nick Mewes is highly acknowledged for assistance in the laboratory. We also thank Cathy Jenks for English language proofreading and four anonymous reviewers for comments and suggestions that greatly improved the manuscript. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 772852, ERC Consolidator Grant 'Glacial Legacy') and the Initiative and Networking Fund of the Helmholtz Association. LSE was supported by the German Research Foundation through Grant EP98/2-1.

3.8 Author contributions

L.S.E. and U.H. designed the study; L.S. conducted library preparation and hybridization capture, and analysed the data with the help of N.B. and H.Z.; L.S.E. designed and supervised the preparation of the long-range PCR products; U.H., N.B., L.E., K.S. and H.Z. supervised the study; and L.S. wrote the manuscript under supervision of U.H. All co-authors commented on a first version of the text.

3.9 References

Abaimov, A. P. (2010). Geographical distribution and genetics of Siberian larch species. In *Permafrost ecosystems Siberian larch forests* (Vol. 209, pp. 41–58). https://doi.org/10.1007/978-1-4020-9693-8

Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J. E., … Wohlfarth, B. (2018). Archaeal community changes in Lateglacial lake sediments: Evidence from ancient DNA. *Quaternary Science Reviews*, *181*, 19–29. https://doi.org/10.1016/j.quascirev.2017.11.037

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. *PLoS ONE*, *13*(4), 1–23. https://doi.org/10.1371/journal.pone.0195403

Andreev, A. A., Tarasov, P. E., Klimanov, V. A., Melles, M., Lisitsyna, O. M., & Hubberten, H. W. (2004). Vegetation and climate changes around the Lama Lake, Taymyr Peninsula, Russia

during the Late Pleistocene and Holocene. *Quaternary International*, *122*, 69–84. https://doi.org/10.1016/j.quaint.2004.01.032

Andrews, S. (2015). FastQC: A quality control tool for high throughput sequence data. Babraham Bioinformatics.

Ávila-Arcos, M. C., Cappellini, E., Romero-Navarro, J. A., Wales, N., Moreno-Mayar, J. V., Rasmussen, M., … Gilbert, M. T. P. (2011). Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. *Scientific Reports*, *1*(74), 1–5. https://doi.org/10.1038/srep00074

Biomatters. (2019). Geneious version 2019.2.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Broad Institute. (2019). Picard toolkit. *GitHub-Repository*. GitHub. Retrieved from http://broadinstitute.github.io/picard/

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics*, *93*(5), 852–864. Retrieved from http://www.sciencedirect.com/science/article/pii/S000292971300459X

CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., … Little, D. P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794–12797. https://doi.org/10.1073/pnas.0905845106

Currat, M., Ruedi, M., Petit, R. J., & Excoffier, L. (2008). The hidden side of invasions: Massive introgression by local genes. *Evolution*, *62*(8), 1908–1920. https://doi.org/10.1111/j.1558-5646.2008.00413.x

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., … Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763. https://doi.org/10.1073/pnas.1314445110

DeVerno, L. L., Charest, P. J., & Bonen, L. (1993). Inheritance of mitochondrial DNA in the conifer *Larix*. *Theoretical and Applied Genetics*, *86*(2–3), 383–388. https://doi.org/10.1007/BF00222106

Du, F. K., Peng, X. L., Liu, J. Q., Lascoux, M., Hu, F. S., & Petit, R. J. (2011). Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan Plateau. *New Phytologist*, *192*(4), 1024–1033. https://doi.org/10.1111/j.1469-8137.2011.03853.x

Du, F. K., Petit, R. J., & Liu, J. Q. (2009). More introgression with less gene flow: Chloroplast vs. mitochondrial DNA in the *Picea asperata* complex in China, and comparison with other Conifers. *Molecular Ecology*, *18*(7), 1396–1407. https://doi.org/10.1111/j.1365-294X.2009.04107.x

Enk, J., Devault, A., Widga, C., Saunders, J., Szpak, P., Southon, J., … Poinar, H. (2016). Mammuthus population dynamics in late Pleistocene North America: Divergence, phylogeography, and introgression. *Frontiers in Ecology and Evolution*, *4*(April), 1–13. https://doi.org/10.3389/fevo.2016.00042

Epp, L.S., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schrøder-Nielsen, A., … Brochmann, C. (2015). Lake sediment multi-taxon DNA from North Greenland records early post-glacial appearance of vascular plants and accurately tracks environmental changes. *Quaternary Science Reviews*, *117*(0318), 152–163. https://doi.org/10.1016/j.quascirev.2015.03.027

Epp, Laura S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A., & Herzschuh, U. (2018). Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling.

Scientific Reports, *8*, 17436. https://doi.org/10.1038/s41598-018-35550-w

Gansauge, M.-T., Gerber, T., Glocke, I., Korlević, P., Lippik, L., Nagel, S., … Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using *T4* DNA ligase. *Nucleic Acids Research*, gkx033. https://doi.org/10.1093/nar/gkx033

Gansauge, M.-T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, *8*(4), 737–748. https://doi.org/10.1038/nprot.2013.038

Giguet-Covex, C., Ficetola, G. F., Walsh, K., Poulenard, J., Bajard, M., Fouinat, L., … Arnaud, F. (2019). New insights on lake sediment DNA from the catchment: importance of taphonomic and analytical issues on the record quality. *Scientific Reports*, *9*(1), 14676. https://doi.org/10.1038/s41598-019-50339-1

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics*, *27*(15), 2153–2155. https://doi.org/10.1093/bioinformatics/btr347

Godbout, J., Yeh, F. C., & Bousquet, J. (2012). Large-scale asymmetric introgression of cytoplasmic DNA reveals Holocene range displacement in a North American boreal pine complex. *Ecology and Evolution*, *2*(8), 1853–1866. https://doi.org/10.1002/ece3.294

Green, B. R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *Plant Journal*, *66*(1), 34–44. https://doi.org/10.1111/j.1365-313X.2011.04541.x

Green, R. E., Malaspinas, A. S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., … Pääbo, S. (2008). A complete Neandertal mitochondrial genome sequence determined by high-throughput sequencing. *Cell*, *134*(3), 416–426. https://doi.org/10.1016/j.cell.2008.06.021

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., … Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Harbert, R. S. (2018). Algorithms and strategies in short-read shotgun metagenomic reconstruction of plant communities. *Applications in Plant Sciences*, *6*(3), 1–7. https://doi.org/10.1002/aps3.1034

Herzschuh, U. (2019). Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography*, (July), 1–9. https://doi.org/10.1111/geb.13018

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., … Boore, J. L. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(49), 19369–19374. https://doi.org/10.1073/pnas.0709121104

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., … Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology*, *21*(8), 1989–2003. https://doi.org/10.1111/j.1365-294X.2011.05287.x

Kistler, L., Montenegro, A., Smith, B. D., Gifford, J. A., Green, R. E., Newsom, L. A., & Shapiro, B. (2014). Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proceedings of the National Academy of Sciences*, *111*(8), 2937–2941. https://doi.org/10.1073/pnas.1318678111

Klemm, J., Herzschuh, U., & Pestryakova, L. A. (2016). Vegetation, climate and lake changes over the last 7000 years at the boreal treeline in north-central Siberia. *Quaternary Science Reviews*, *147*, 422–434. https://doi.org/10.1016/j.quascirev.2015.08.015

Larson, D., & Abott, T. (2016). bam-readcount. Retrieved from https://github.com/genome/bam-

readcount

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., … Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(17), 4325–4333. https://doi.org/10.1073/pnas.1720115115

Li, C., Hofreiter, M., Straube, N., Corrigan, S., & Naylor, G. J. P. (2013). Capturing protein-coding genes across highly divergent species. *BioTechniques*, *54*(6), 321–326. https://doi.org/10.2144/000114039

Li, H. (2012). Seqtk. *GitHub-Repository*. GitHub. Retrieved from https://github.com/lh3/seqtk

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, S., Stoof-Leichsenring, K. R., Kruse, S., Pestryakova, L. A., & Herzschuh, U. (2020). Holocene Vegetation and Plant Diversity Changes in the North-Eastern Siberian Treeline Region From Pollen and Sedimentary Ancient DNA. *Frontiers in Ecology and Evolution*, *8*(September), 1–17. https://doi.org/10.3389/fevo.2020.560243

MacDonald, G. M., Gervais, B. R., Snyder, J. A., Tarasov, G. A., & Borisova, O. K. (2000). Radiocarbon dated *Pinus sylvestris* L. wood from beyond tree-line on the Kola Peninsula, Russia. *Holocene*, *10*(1), 143–147. https://doi.org/10.1191/095968300667807510

MacDonald, G. M., Kremenetski, K. V., & Beilman, D. W. (2008). Climate change and the northern Russian treeline zone. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1501), 2285–2299. https://doi.org/10.1098/rstb.2007.2200

Marcott, S. a., Shakun, J. D., Clark, P. U., & Mix, A. C. (2013). A reconstruction of regional and global temperature for the past 11,300 years. *Science (New York, N.Y.)*, *339*(6124), 1198–1201. https://doi.org/10.1126/science.1228026

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE*, *5*(11), e14004. https://doi.org/10.1371/journal.pone.0014004

Moore, C. R., Brooks, M. J., Goodyear, A. C., Ferguson, T. A., Perrotti, A. G., Mitra, S., … Pyne-O'Donnell, S. (2019). Sediment Cores from White Pond, South Carolina, contain a Platinum Anomaly, Pyrogenic Carbon Peak, and Coprophilous Spore Decline at 12.8 ka. *Scientific Reports*, *9*(1), 15121. https://doi.org/10.1038/s41598-019-51552-8

Murchie, T., Kuch, M., Duggan, A., Ledger, M. L., Roche, K., Klunk, J., … Poinar, H. (2019). PalaeoChip Arctic1.0: An optimised eDNA targeted enrichment approach to reconstructing past environments. *BioRxiv*, (August), 1–43.

Neale, D. B., & Wheeler, N. C. (2019). Genomes: Classical era. In D. B. Neale & N. C. Wheeler (Eds.), *The conifers: Genomes, variation and evolution* (pp. 25–42). Springer International Publishing. https://doi.org/10.1007/978-3-319-46807-5_2

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., … Stamatoyannopoulos, J. A. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics*, *28*(14), 1919–1920. https://doi.org/10.1093/bioinformatics/bts277

Niemeyer, B., Epp, L. S., Stoof-Leichsenring, K. R., Pestryakova, L. A., & Herzschuh, U. (2017). A comparison of sedimentary DNA and pollen from lake sediments in recording vegetation composition at the Siberian treeline. *Molecular Ecology Resources*, *17*(6), e46–e62. https://doi.org/10.1111/1755-0998.12689

Paijmans, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2016). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, *16*(1), 42–55. https://doi.org/10.1111/1755-0998.12420

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., … Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, *24*(7), 1485–1498. https://doi.org/10.1111/mec.13136

Parducci, L., Alsos Greve, I., Unneberg, P., Pedersen, M. W., Han, L., Lammers, Y., … Wohlfarth, B. (2019). Shotgun ancient DNA, pollen and macrofossil analysis of Lateglacial lake sediments from southern Sweden. *Frontiers in Ecology and Evolution*, *7*, 189. https://doi.org/10.3389/FEVO.2019.00189

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. *New Phytologist*, *214*(3), 924–942. https://doi.org/10.1111/nph.14470

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., … Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49. https://doi.org/10.1038/nature19085

Peñalba, J. V., Smith, L. L., Tonione, M. A., Sass, C., Hykin, S. M., Skipwith, P. L., … Moritz, C. (2014). Sequence capture using PCR-generated probes: A cost-effective method of targeted high-throughput sequencing for nonmodel organisms. *Molecular Ecology Resources*, *14*(5), 1000–1010. https://doi.org/10.1111/1755-0998.12249

Petit, R. J., Duminil, J., Fineschi, S., Hampe, A., Salvini, D., & Vendramin, G. G. (2004). INVITED REVIEW: Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, *14*(3), 689–701. https://doi.org/10.1111/j.1365-294X.2004.02410.x

Pisaric, M. F. J., MacDonald, G. M., Velichko, A. A., & Cwynar, L. C. (2001). The Lateglacial and Postglacial vegetation history of the northwestern limits of Beringia, based on pollen, stomate and tree stump evidence. *Quaternary Science Reviews*, *20*(1–3), 235–245. https://doi.org/10.1016/S0277-3791(00)00120-7

Polezhaeva, M. A., Lascoux, M., & Semerikov, V. L. (2010). Cytoplasmic DNA variation and biogeography of *Larix* Mill, in Northeast Asia. *Molecular Ecology*, *19*(6), 1239–1252.

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE*, *7*(3), e34131. https://doi.org/10.1371/journal.pone.0034131

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017). HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods in Ecology and Evolution*, *8*(10), 1374–1388. https://doi.org/10.1111/2041-210X.12785

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K. R., Zimmermann, H. H., Pestryakova, L. A., Epp, L. S., & Herzschuh, U. (2020). *PRJEB35838*. European Nucleotide Archive.

Semerikov, V. L., Semerikova, S. A., Polezhaeva, M. A., Kosintsev, P. A., & Lascoux, M. (2013). Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix* sibirica): a range-wide analysis of cytoplasmic markers. *Molecular Ecology*, *22*(19), 4958–4971. https://doi.org/10.1111/mec.12433

Shaw, J., Lickey, E. B., Schilling, E. E., & Small, R. L. (2007). Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The Tortoise and the hare III. *American Journal of Botany*, *94*(3), 275–288. https://doi.org/10.3732/ajb.94.3.275

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., … Meyer, M. (2017a). Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, *356*(6338), 605–608. https://doi.org/10.1126/science.aam9695

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., … Meyer, M.

(2017b). supplement - Neandertal and Denisovan DNA from Pleistocene sediments. *Science*, *356*(6338), 605–608. https://doi.org/10.1126/science.aam9695

Sønstebø, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., … Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, *10*(6), 1009–1018. https://doi.org/10.1111/j.1755-0998.2010.02855.x

Szmidt, A. E., Aldén, R., & Hällgren, J.-E. (1987). Paternal inheritance of chloroplast DNA in *Larix*. *Plant Molecular Biology*, *9*, 59–64.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793. https://doi.org/10.1111/j.1365-294X.2012.05542.x

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14. https://doi.org/10.1093/nar/gkl938

Wang, Y., Heintzman, P. D., Newsom, L., Bigelow, N. H., Wooller, M. J., Shapiro, B., & Williams, J. W. (2017). The southern coastal Beringian land bridge: cryptic refugium or pseudorefugium for woody plants during the Last Glacial Maximum? *Journal of Biogeography*, *44*(7), 1559–1571. https://doi.org/10.1111/jbi.13010

Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., … Cooper, A. (2003). Diverse plant and animal genetic records from holocene and pleistocene sediments. *Science*, *300*(5620), 791–795. https://doi.org/10.1126/science.1084114

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *BioRxiv*, 762302. https://doi.org/10.1101/762302

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zimmermann, H. H., Harms, L., Epp, L. S., Mewes, N., Bernhardt, N., Kruse, S., … Herzschuh, U. (2019). Chloroplast and mitochondrial genetic variation of larches at the Siberian tundra-taiga ecotone revealed by de novo assembly. PLOS ONE, 14(7), e0216966. https://doi.org/10.1371/journal.pone.0216966

4 Manuscript III

Dynamics of larch species in Siberia since the Last Glacial captured from sedimentary ancient DNA

Status

Under review at *communications biology*
update after submission of thesis: revised version published
Schulte, L., Meucci, S., Stoof-Leichsenring, K.R. *et al. Larix* species range dynamics in Siberia since the Last Glacial captured from sedimentary ancient DNA. *Commun Biol* **5,** 570 (2022). https://doi.org/10.1038/s42003-022-03455-0

Authors:

Luise Schulte[1,2], Stefano Meucci[1,2], Kathleen Stoof-Leichsenring[1], Tony Heitkam[3], Nicola Schmidt[3], Barbara von Hippel[1,2], Andrej A. Andreev[1], Bernhard Diekmann[1,4], Boris K. Biskaborn[1], Bernd Wagner[4], Martin Melles[4], Lyudmila A. Pestryakova[5], Inger G. Alsos[6], Charlotte Clarke[6,7], Konstantin V. Krutovsky[8,9,10,11,12], Ulrike Herzschuh[1,2,13]

Affiliations:

[1] Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Polar Terrestrial Environmental Systems, Potsdam, Germany
[2] Institute of Biochemistry and Biology, University of Potsdam, Potsdam, Germany
[3] Institute of Botany, Technische Universität Dresden, Dresden, Germany
[4] Institute of Geology and Mineralogy, University of Cologne, Cologne, Germany
[5] North-Eastern Federal University of Yakutsk, Institute of Natural Sciences, Yakutsk, Russia
[6] The Arctic University Museum of Norway, UiT - The Arctic University of Norway, Tromsø, Norway
[7] University of Southampton, School of Geography and Environmental Science, Southampton, UK
[8] Department of Forest Genetics and Forest Tree Breeding, George-August University of Göttingen, Göttingen, Germany
[9] Center for Integrated Breeding Research, Georg-August University of Göttingen, Göttingen, Germany
[10] Laboratory of Population Genetics, N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, 119333 Moscow, Russian Federation
[11] Laboratory of Forest Genomics, Genome Research and Education Center, Department of Genomics and Bio-informatics, Institute of Fundamental Biology and Biotechnology, Siberian Federal University, Krasnoyarsk, Russian Federation
[12] Forestry Faculty, G. F. Morozov Voronezh State University of Forestry and Technologies, Voronezh, Russian Federation
[13] Institute of Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany

4.1 Abstract

Climate change is expected to cause major shifts in Asian boreal forests which are in a vast area dominated by two species of the deciduous needle tree larch (*Larix*). The species differ markedly in their ecosystem functions, thus shifts in their respective ranges are of global relevance. However, drivers of species distribution are not well understood also because paleoecological data at species level are lacking. This study tracks *Larix* species history in time and space using target enrichment on 67 lake sediment samples from eight lakes across Siberia. We show that *Larix sibirica*, presently dominating in western Siberia, likely migrated to its northern distribution

area only in the Holocene, and had a much wider eastern distribution around 33,000 years before present. DNA samples dated to the Last Glacial Maximum, consistently show genotypes of *L. gmelinii*. Our results suggest climate as a strong determinant of species distribution in *Larix* and provide temporal and spatial data for species projection in a changing climate.

4.2 Introduction

Recent climate warming is strongly amplified in the high latitudes of the Northern Hemisphere[1], but strongest in the Russian Arctic[2], and is expected to cause major shifts in boreal ecosystems[3]. Today, the deciduous larch (*Larix*) dominates around 81% of the vast Siberian boreal forests[4]. It provides ecosystem services as a wildlife habitat, carbon stock and permafrost stabilizer, and renders strong economic benefits[4]. Understanding larch forest dynamics in response to climate changes in the past is critical for forecasting future ecosystem changes. Despite this importance we still lack knowledge and understanding of past distribution of *Larix* species and the major drivers constraining their current distribution.

A *Larix* species complex dominates the Siberian boreal forests with a biogeographical split in west and east (Fig. 1). West Siberia, from the western Urals to the West Siberian Plain, is dominated by *Larix sibirica*. Central and eastern Siberia is dominated by a species complex which is sometimes split into a western and eastern variety, subspecies or two separate species[5], but officially recognized only as one species, *L. gmelinii*[6]. Although outwardly very similar, the species differ in their ecological properties and ecosystem services[7]. *L. gmelinii* can grow in a severe continental climate and on continuous permafrost with shallow active-layer depth, and is reported to stabilize the soil, decrease permafrost thaw depth, and subsequently also the amount of carbon release[8,9]. In harsh ecological conditions it can still grow as dwarf or shrub-like forms[5]. In contrast, *L. sibirica* is more warmth-demanding and sensitive to frozen and swamped soils[7,10], grows faster[11] and produces more seeds[5]. It is still under discussion whether biogeography (such as the historical distribution of species) or environmental factors such as local hydrology and habitat are more restrictive for current species distribution. This knowledge however represents the basis for prediction of future forest cover.

Larix has been a well-established forest constituent in Eurasia since more than 25 million years ago[12]. Siberian larch species, in particular *L. gmelinii,* are considered to have formed due to climate changes in the Pleistocene resulting in adaptation to increasing climate continentality[5]. The adaptation to extreme cold prompts the question of whether the species could have survived the last glacial in general and more specifically the Last Glacial Maximum (LGM, 17-23 ka BP[13] in northern refugial populations or whether its distributional range retreated to the south like other tree species. Scarce paleobotanical data from a pre-LGM interstadial warm phase (Marine Isotope Stage (MIS) 3 or Karginskii Interstadial) showed that *Larix* forests were relatively common in northeast Siberia[14,15]. During the LGM, Siberia was dominated by open tundra and steppe vegetation[16], but scattered *Larix* pollen and macrofossil findings hint at the survival of refugial populations also in high latitudes[17–19]. However, our knowledge about these populations is limited, especially as *Larix* pollen production and preservation are poor[20]. Furthermore, it remains unclear to what extent these refugial populations contributed to a postglacial recolonization[11,21]. A second question is whether the mode of postglacial recolonization (a spread out of high latitude refugia or migrating from the south) differs among the different larch species.

Fig. 1 Map of study sites and current distribution of *Larix* species. Black dots indicate positions of lakes with studied sediment cores, colors indicate current species distribution (adapted from Semerikov and Lascoux[22], base map done with ggmap[23])

Traditional paleoecological approaches are based on pollen or metabarcoding using short diagnostic metabarcodes of ancient DNA samples isolated from lake sediments, peat or permafrost sections[17,24,25]. These methods are powerful in reconstructing the general past vegetation composition and species abundance[26], but do not allow taxonomic resolution of *Larix* into species. Studies using modern DNA can show distributions of current populations and also infer demographic scenarios. Nevertheless, tracing of populations in the past remains difficult as signals from different times can be superimposed and only a few studies have hitherto been available to serve as ground truth estimates for eastern Siberia[21]. Recently, Schulte et al.[27] developed target enrichment of ancient *Larix* DNA by using the complete chloroplast genome of *Larix* to develop a bait set (hybridization oligonucleotides) on sedimentary ancient DNA (*sed*aDNA). They showed that this method is capable of discriminating between different *Larix* species and potentially different populations over time. Chloroplasts are a promising target for capture enrichment as they are present in multiple copies in almost all plant cells. Another, yet unexplored multi-copy target for capture enrichment are repetitive DNA sequences in the nuclear plant genome. Occupying about 80% of the large genomes of conifers, they are believed to be the main source of genome expansion[28]. Satellite repeats are arrays of tandemly repeated non-coding DNA stretches that can affect important cell functions such as chromosome stability and cell division[29]. Among repetitive elements, satellite DNA repeats are among the fastest evolving and are hence often specific for a certain species or genus[30]. As they constitute up to 37% of some plant genomes[31], they often remain as off-target reads in target capture enrichment of single- or low-copy targets in modern species[32,33]. Despite their genomic abundance and their potential specificity to genus or species, enrichment of repetitive elements from *sed*aDNA has so far not been reported.

In this study, we applied chloroplast and nuclear target enrichment to a total of 67 *sed*aDNA samples, isolated from sediment cores from eight lakes distributed across Siberia covering the last 50,000 years (Fig. 1). The lakes were selected according to the availability of sediment records spanning the LGM and a known presence of *Larix* around the lakes. Via hybridization capture, we

successfully enriched the samples with *Larix* chloroplast genome and abundantly detected the *Larix* main satellite repeat in the off-targets. The analysis of genetic variants in the enriched sequences gave us unprecedented insights into ancient species distribution. We show that local habitat such as permafrost is a stronger determinant of species distribution than biogeography. We highlight that *L. sibirica* had a much wider distribution in the past; nevertheless, glacial refugial populations were dominated by *L. gmelinii,* potentially enhancing the post-glacial colonization in East Siberia.

To our knowledge, this study is the first to provide large-scale species information using hybridization capture data of *sed*aDNA isolated from lake archives and the first to report the enrichment of ancient repetitive DNA.

4.3 Results & Discussion

4.3.1 Chloroplast and repetitive DNA enrichment in the *sed*aDNA samples

We generated the first large-scale hybridization capture dataset using sedimentary ancient DNA. Sequencing of two datasets produced 325.5 million (M) quality-filtered paired-end reads. The first hybridization capture dataset, targeting both the chloroplast and a set of nuclear genes of *Larix* on 63 samples and 19 negative controls from seven lake sediment records resulted in 324 M quality-filtered paired-end reads. The second hybridization capture dataset, targeting only the set of nuclear genes of *Larix* on 4 samples and 2 negative controls from an additional lake (Lake CH12) resulted in 1.5 M reads. Quality-filtering of an additional published hybridization capture dataset[27], targeting the *Larix* chloroplast genome on the same CH12 samples as applied for the second dataset, added another 54 M reads.

We are the first to report *sed*aDNA enrichment for both chloroplast and nuclear sequences. For the chloroplast enrichment, 390 thousand (K) reads (1%) were classified to *Larix* at the genus or species level. Average coverage of bait regions was 19% at a mean read depth of 0.8. Sequencing of 19 library and extraction blank (negative control) samples resulted in 597 K paired-end reads, of which 58% quality-filtered and deduplicated reads remained. Of these, 38% were classified, with 0.03% of them (463 reads) corresponding to the genus *Larix.* Negative controls from library preparation resulted in no to very few (0 to 5) reads mapping to the *Larix* chloroplast reference genome. Negative controls from DNA extractions, which were in several cases pooled to one library, showed a low number of reads mapped to *Larix* (0 to 94 reads, except 237 reads in one case). Excluding all reads in negative controls from the sample analysis had no impact on the patterns resulting from the analysis of sample data. Detailed results and evaluation of negative controls are included in the Supplementary Information and Supplementary Tables 1 and 2. Samples of all lake records with sufficient read coverage showed damage patterns typical of ancient DNA (see Supplementary Table 3).

These results are comparable to the results obtained by Schulte et al.[27], where 36% of quality-filtered reads were classified as Viridiplantae with 9% assigned to *Larix.* In contrast to[27], we raised the confidence threshold of taxonomic classification, which drastically reduced the number of classified reads, but increased the confidence in the analysis.

To analyze the enrichment obtained by the gene bait set, taxonomic classification was repeated using a plant genome database including available Pinaceae genomes. The classification resulted

in 716 K reads assigned to *Larix*, increasing the previous results by 325 K reads. However, almost no reads were mapped against the targeting baits (a maximum of 5 reads for some samples). A closer inspection of unmapped reads assigned to *Larix* revealed a high content of repetitive DNA, in particular EulaSat1, the most abundant satellite repeat of *Larix*[30,34]. This short repeat of 173 bp is arranged in large arrays of tandem repeats and is exclusively present in larches. Analysis of modern *L. sibirica*, and *L. gmelinii* (western and eastern range) genomes reveals that EulaSat1 occurs in all species, contributing to 0.62% (*L. sibirica)*, 2.52% (western range *L. gmelinii)*, and 2.39% (eastern range *L. gmelinii)*, of the genomes, respectively (Supplementary Fig. 2). A comparison of the read proportions mapping to the repeat sequence in the different datasets of Lake CH12 showed a specific enrichment of the sequence by the nuclear gene hybridization probe set (Supplementary Fig. 3).

In total, 17 K reads mapped to the sequence of EulaSat1. The abundance of all reads mapped per sample is in agreement with the abundance of reads mapped to the chloroplast genome, confirming the general history of forest development (Fig. 2). Analysis of the repeat nucleotide frequencies showed a high constancy over all samples (Supplementary Fig. 4). This suggests a high conservation of the EulaSat1 sequence in Siberian larches over time and space. Although satellite repeats are reported to have a high sequence turnover, for larches it has been shown that repeat profiles between two geographically well separated species – the European larch (*L. decidua)* and the Japanese larch (*L. kaempferi*) – are very similar[30]. The main satellite in all larches, EulaSat1, is believed to have greatly multiplied after the split of *Larix* from *Pseudotsuga*[30]. Given the ongoing hybridization between the three Siberian larch species, it is not surprising to find a consistent pattern of nucleotide frequencies in all samples.

Off-target reads of hybridization capture enrichment sequencing have already been demonstrated to be useful for the analysis of high-copy DNA such as ribosomal DNA or plastomes[32,35,36]. A recent study on five modern sedges showed that hybridization capture sequencing data originally targeting a set of gene exons can also be used to study the repetitive sequence fraction and even infer phylogenetic relationships based on repetitive sequence abundance[33]. Even more promising than using repetitive sequence abundance in off-target capture enrichment datasets is to use sequence variation in repeats[37,38].

Reads mapping to both the chloroplast and to the repeat sequence, show similar patterns of abundance (see Fig. 2). Compared with published metabarcoding and pollen data from the same locations, the *Larix* abundance patterns can be globally reproduced, underpinning the notion that read abundances in capture data can be used as good estimates of plant abundances. For older parts of the lake records, capture data show *Larix* where metabarcoding data were unable to detect a clear signal (see Fig. 2, lakes Billyakh, Bolshoye Shchuchye, Kyutyunda and Lama). This shows that target capture is superior to metabarcoding when analyzing one taxonomic group in depth, as it is less prone to errors by DNA degradation, which can impede primer binding if the molecule becomes too short. Also, independent from age, rare taxa mostly need multiple PCR replicates to be detected by metabarcoding[39,40]. Target capture however, is more sensitive in identifying one focal taxon group, as the total target length can be much larger (e.g. a complete organellar genome) than for metabarcoding and the DNA damage patterns are put to use to authenticate ancient DNA. Also, it is limited by molecule length only by the applied threshold in the bioinformatic analysis, for which we used 30 base pairs (bp) as opposed to a minimum of 85 bp molecule length for the *Larix* metabarcoding marker (for the plant-specific trnL g/h marker[41]). Similarly, compared to traditional pollen analysis, target capture is more accurate at tracing a

specific target group, as it is not dependent on pollen productivity. Especially in the case of *Larix,* pollen productivity is low and preservation poor, resulting in rare findings of its pollen in the sediments[20,42]. This could explain why for Lake Bolshoye Shchuchye, only a single *Larix* pollen grain was retrieved throughout the core, whereas capture and metabarcoding show a strong signal in the Holocene sediments. Capture data also records signals in MIS 2 sediments, however, read counts are extremely low, and as it is the only record, where both of the other proxies fail to report a signal, it should be interpreted with caution.

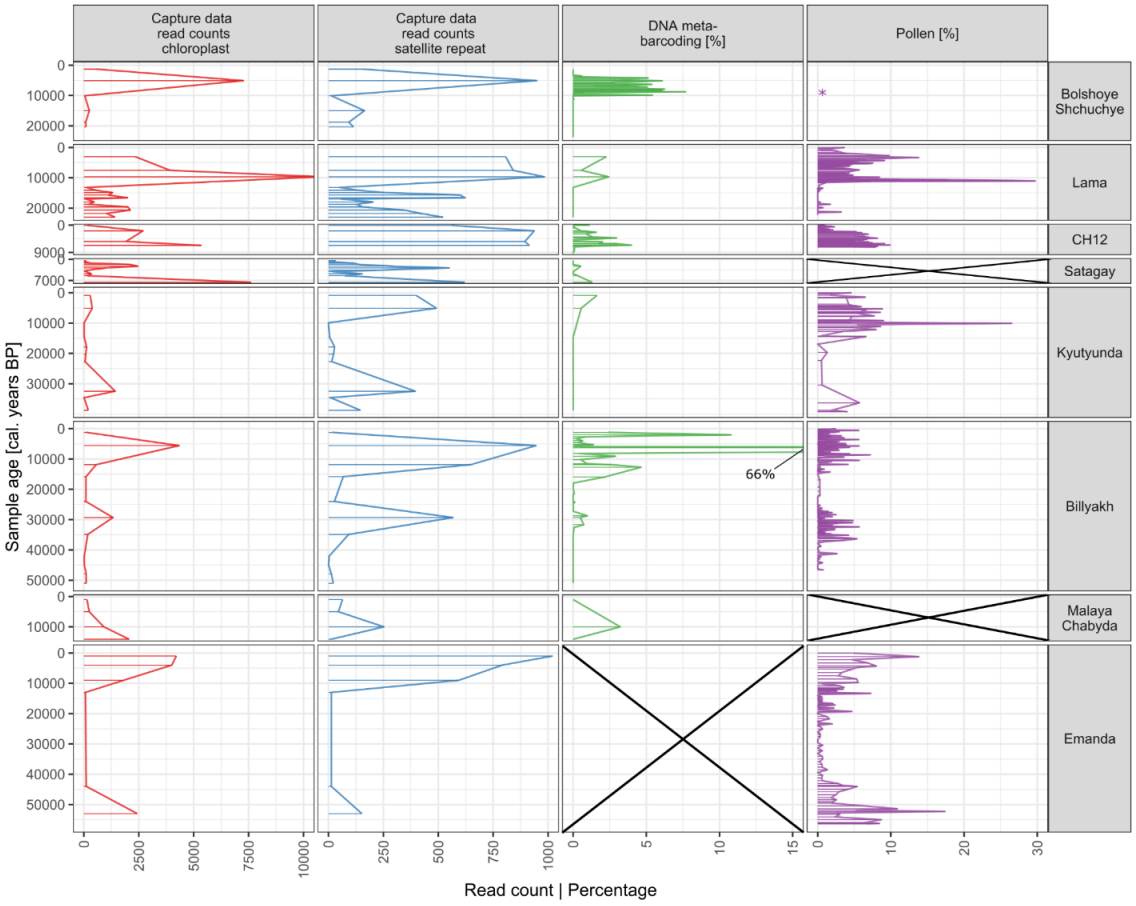

Fig. 2 Comparison of capture enrichment with available DNA metabarcoding and pollen datasets. From left to right: *Larix*-classified read counts mapping to (1) the *Larix* chloroplast and (2) the EulaSat1 satellite repeat sequence, (3) percentage of *Larix* counts in metabarcoding data, (4) percentage of *Larix* pollen in pollen assemblages. All data from this study, except metabarcoding data from lakes CH12[11] and Bolshoye Shchuchye[43] and all pollen data except for several samples of Lake Kyutyunda which were produced in this study[44–46]. Pollen data of Lake Lama and Holocene part of Lake Kyutyunda are based on parallel sediment cores PG1111 and PG2022, respectively. No available data are marked with crosses, asterisk marks a single *Larix* pollen grain found in the Bolshoye Shchuchye sediments.

4.3.2 A wider pre-glacial distribution of *L. sibirica*

Chloroplast genomes of *L. gmelinii* and *L. sibirica* differ at 157 positions, which can be used to differentiate species in target capture enriched *seda*DNA[27]. Here, we applied this approach to lake sediment records, which are distributed across Siberia (Fig. 1) and have time ranges back to MIS 3, and thereby were able to track species composition in space and time for wide parts of the species ranges.

In lakes Billyakh and Kyutyunda, ca. 1,500 km east of *L. sibirica*'s current range (Fig. 1), we found evidence for a wider distribution of the species around 32 and 34 ka BP in MIS 3 (Fig. 3). Billyakh

is situated in the western part of the Verkhoyansk Mountains, and Kyutyunda on the Central Siberian Plateau. Both lakes have low counts of *Larix* reads in their oldest samples dated to 51 ka BP (Billyakh) and 38 ka BP (Kyutyunda) with variants of *L. gmelinii*, but there is a sudden rise in variants attributed to *L. sibirica* at 34 ka BP (Billyakh) and 32 ka BP (Kyutyunda), which persists in the following samples, but strongly decreases in younger samples (Fig. 3). The rise in the *L. sibirica* variants coincides with a peak in read counts for Lake Kyutyunda. These signals suggest a rapid invasion of *L. sibirica* into the ranges of *L. gmelinii* in climatically favorable times and a local depletion or extinction of *L. sibirica* during the following harsher climates. Lake Billyakh pollen data suggest a moister and warmer climate around 50-30 ka BP than in the later part of the last glacial associated with the MIS3 Interstadial in Siberia[47].

Strong support for a wider pre-glacial distribution of *L. sibirica* comes from genetic analyses which show that it is genetically close to *L. olgensis*, today occurring on the Korean Peninsula and adjacent areas of China and Russia[48,49]. It is assumed that the *L. sibirica-olgensis* complex used to share a common range, which was disrupted and displaced when the better cold-adapted *L. gmelinii* expanded south and southwest during the more continental climatic conditions of the Pleistocene[48,50]. Furthermore, modern and ancient genetic studies suggest that the *L. sibirica* zone was recently invaded by *L. gmelinii* from the east in the hybridization zone of the species, as climate cooled after the mid-Holocene thermal maximum[11,21]. Today, pure stands of *L. sibirica* do not form a continuous habitat, but occur in netted islands[5] and morphological features of *L. sibirica* can be found in populations of *L. gmelinii* located at least a hundred kilometers east of the closest *L. sibirica* populations[51]. Macrofossil findings of *L. sibirica* in Scandinavia dated to the early Holocene, point to the capability of rapid long-distance jump dispersal of this species[52]. Fossil *L. sibirica* cones dated to the end of the Pliocene and in the Pleistocene have also been found far east of its current range in several river valleys including Kolyma, Aldan and Omolon and even in the basin of the Sea of Okhotsk[7]. These indicate long-distance seed dispersal by rivers which may also have assisted in successful establishment since the active-layer depth is deeper close to rivers[53,54]. As mentioned earlier, *L. sibirica* is sensitive to permafrost and waterlogged soils. A warmer phase with a deeper thawed layer above the permafrost could have enabled *L. sibirica* to spread and establish in regions that today are part of the geographic range of *L. gmelinii*, as *L. sibirica* is reported to have higher growth rates than *L. gmelinii*[11].

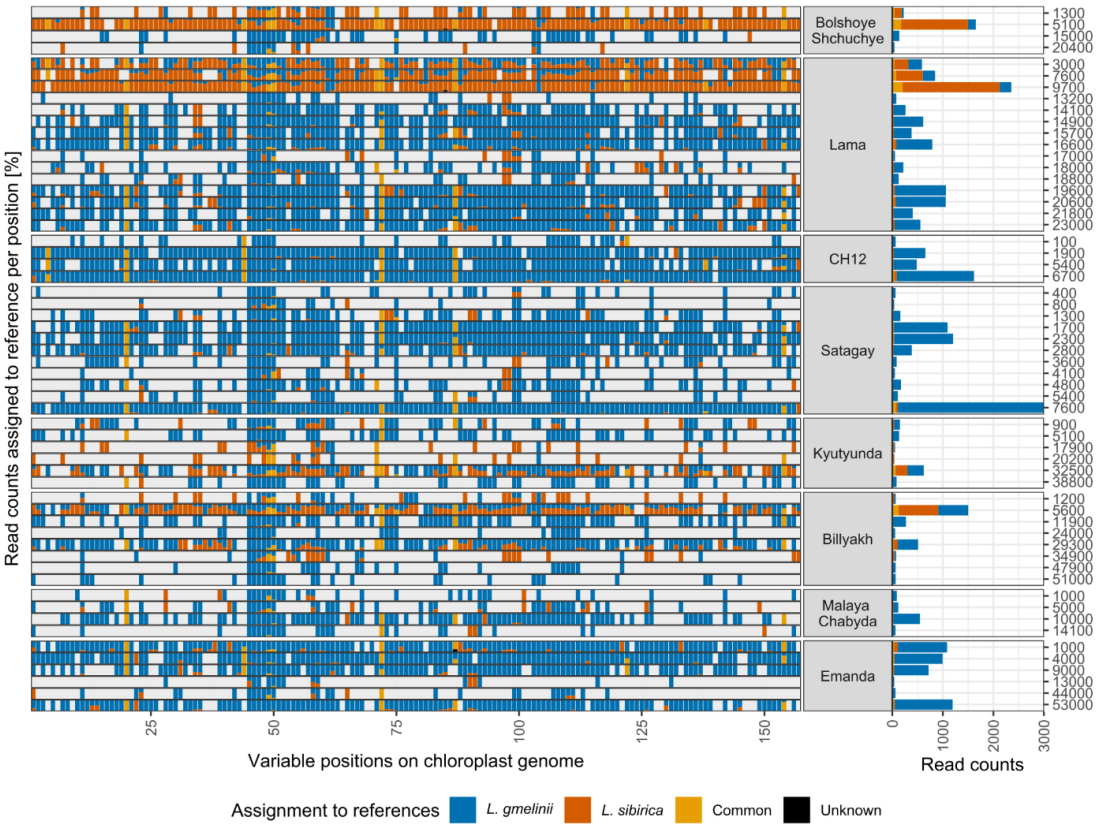

Fig. 3 Percentage and counts of variable positions along *Larix* chloroplast genome assigned to species. Left: Alignment of *Larix*-classified reads against the chloroplast genome at the 157 variable positions between the species. For each position, the percentage of reads assigned to the single species is displayed. Each row represents one sample named according to the calibrated age before present. Gray background indicates no coverage at the respective position. Right: Total number of reads assigned to each of the species per sample.

4.3.3 *Larix gmelinii* formed northern LGM refugia across Siberia

The possible survival of *Larix* in high latitude glacial refugia during the LGM is still under discussion[4,55] although more and more evidence is reported in favor of the existence of such refugia[15,18,19]. The question of which of the *Larix* species formed these populations has hitherto been unanswered, as both pollen and established metabarcoding markers are not sufficiently variable in the genus *Larix,* and findings of fossilized cones identifiable to species are rare. By enriching the samples with chloroplast genome sequences, we are, for the first time, able to distinguish between *L. sibirica* and *L. gmelinii* in glacial refugial populations.

From Lake Lama, located at the western margin of the Putorana Plateau (Taymyr Peninsula), we obtained a continuous record extending from 23 ka BP to today with varying read counts with minima around 18-17 ka BP and 13 ka BP. All samples prior to the Holocene show variants predominantly assigned to *L. gmelinii* (Fig. 3). Our results suggest a local survival of *L. gmelinii* at Lake Lama throughout the LGM, which is supported by low numbers of *Larix* pollen detected through this period. Both capture data and pollen indicate an increase from ca. 11 ka BP[56]. Sparse *Larix* pollen in the bottom part of the record could be an indication of a possible refugial population (Fig. 2;[56]).

In Bolshoye Shchuchye, the westernmost lake of the study, situated in the Polar Ural Mountains, all Pleistocene samples are similarly dominated by *L. gmelinii* variants (Fig. 3). However, read counts for some samples are extremely low and samples from 18 and 10 ka BP had so low counts of mapped reads that none of the variable positions between the species was covered. Although

reads mapped to the satellite repeat of *Larix* also showed a Pleistocene signal, this was not repeated in pollen or metabarcoding (Fig. 2) which instead indicate a treeless arctic-alpine flora for the late Pleistocene[43,45]. Especially for the sample of 20.4 ka, *Larix* read counts are extremely low and new investigations would be needed to confirm a local presence of *Larix* during the LGM.

The record of Lake Billyakh situated in the western Verkhoyansk Mountain Range, likewise shows extremely low counts of reads mapped to the reference for a range of samples with no reads covering the studied variable sites (45, 42 and 15 ka BP, 11-56 reads mapped to non-variable sites). However, the pollen record for the same core shows a quasi-continuous record of *Larix* with a gap only occurring during the early LGM[47] (25-22 ka BP, Fig. 2). Considering the known short-distance dispersal ability and poor preservation of *Larix* pollen, this strongly supports the presumed existence of a local glacial refugium at Lake Billyakh during that time[18]. Our samples also show a low but steady presence of *Larix* throughout the rest of the record, thus making glacial survival probable. The sample closest to the LGM (24 ka BP) indicates a clear dominance of *L. gmelinii* type variants.

The only exception to this general pattern is the record from Lake Kyutyunda, which is located on the Central Siberian Plateau west of the Verkhoyansk Mountain Range. In this record, LGM samples have extremely low counts, but show variants assigned to *L. sibirica* and not to *L. gmelinii* as in the other lakes. In addition, the preceding sample dated to the MIS3 interstadial shows *L. sibirica* variants. A possible explanation is that relics of *L. sibirica* survived during the LGM, but were unable to spread after climate warming, possibly due to genetic depletion or later local extinction. The presence of reworked sediment material can also not be excluded, as suggested by reworked pollen in the record[46].

In conclusion, our data show almost exclusively *L. gmelinii* variants for samples covering the most severe LGM climate conditions. This is in agreement with the ecological characteristics describing the species as adapted to extreme cold. In contrast to *L. sibirica*, both can grow in dwarf forms and propagate clonally and potentially survive thousands of years of adverse climatic conditions[57].

4.3.4 Postglacial colonization history - differences among larch species

Of great interest in the *Larix* history is not only the location and extent of possible high latitude glacial refugia, but also if and to what extent these refugia contributed to the recolonization of Siberia after the LGM. Northern refugial populations could have functioned as kernels of post-glacial population spread and recolonization, or spreading could have been driven by populations that survived in southern refugia. There are only a few studies on modern populations that report evidence for possible recolonization scenarios of *Larix*[21,49,58]. Here, we show that patterns differ between *L. sibirica* and *L. gmelinii*.

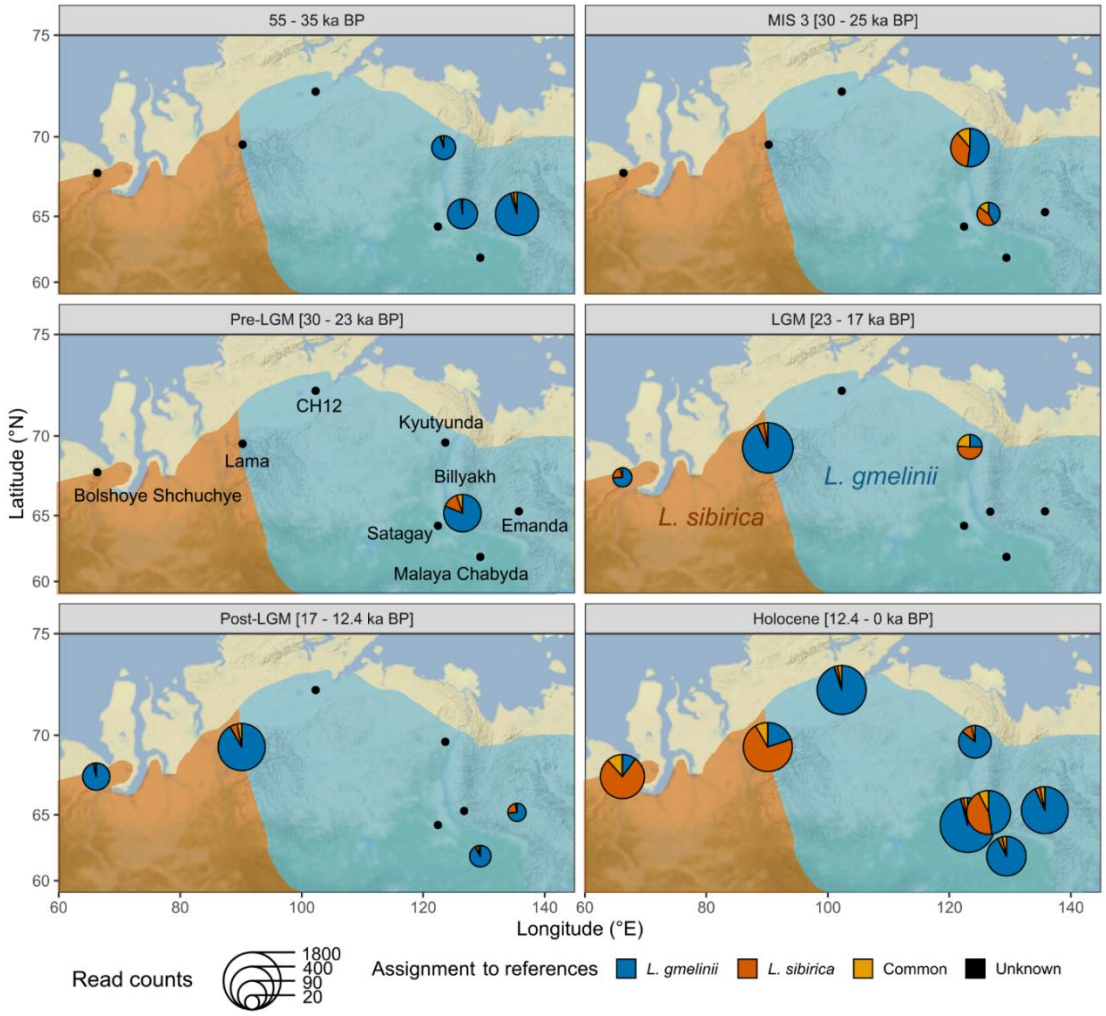

Fig. 4 Percentage of reads assigned to references displayed on the geographical locations of the lakes investigated. Samples in the same time frame are averaged. Lake names and current species ranges are annotated in the middle plots.

In the western part of our study region, two lakes are situated in the current distribution range of *L. sibirica* (Figs. 1 and 4): Lake Bolshoye Shchuchye in the Polar Ural Mountains and Lake Lama on the Taymyr Peninsula. Despite this, both lakes show *L. gmelinii* for all Pleistocene samples, and a strong signal of *L. sibirica* variants only in the Holocene samples, with ages of 5.1 ka BP in Lake Bolshoye Shchuchye and 9.7 ka BP in Lake Lama (Fig. 3). The peak in *L. sibirica* also coincides with a peak of read counts in the respective sample, with a *Larix* pollen peak in Lake Lama sediments[56], and metabarcoding for Lake Bolshoye Shchuchye[43]. This points to a migration of *L. sibirica* in its current northern area of distribution in the course of climate warming during the early Holocene, whereas glacial refugial populations were consisting of *L. gmelinii*. Although the local survival of *L. gmelinii* around Lake Bolshoye Shchuchye remains uncertain due to extremely low read counts, it is clear that *L. sibirica* did not form a refugial population at this site.

A range-wide genetic study of *L. sibirica* analyzing chlorotypes and mitotypes of individuals[21] found strong indications for a rapid colonization of the West Siberian Plains from populations originating from the foothills of the Sayan Mountains in the south, close to the border of Mongolia, with only limited contribution from local populations. According to our results, the

local populations could have been *L. gmelinii* populations, while the rapid invasion could have been *L. sibirica*.

At the eastern range of the study region, in the current range of *L. gmelinii,* namely at lakes Emanda, Satagay and Malaya Chabyda, genetic variations throughout the records are less pronounced. Of the three lake records, only that from Lake Emanda reaches back beyond the LGM, but with a sampling gap for the time of the LGM. Therefore, it remains uncertain whether populations survived the LGM locally, or whether they were invaded or replaced by populations coming from the south with Holocene warming. The restricted variations throughout the record, however, hint at stable populations, which is supported by scarce pollen findings (Fig. 2).

Our data suggest that postglacial recolonization of *L. sibirica* was not started from high latitude glacial refugia, but from southern populations. In contrast, northern glacial populations of *L. gmelinii* could have potentially enhanced rapid dispersal after the LGM in their current area of distribution.

4.3.5 Environment likely plays a more important role than biogeography

The current boundaries of boreal *Larix* species arranged from west to east suggest a possible strong influence of biogeography as a driver of distribution, whereas the gradient of increasing continental climate towards the east assumes a strong influence of environment. By tracking species distribution in the past, spanning the time of the strongly adverse climate of the LGM, we can give hitherto unprecedented insights into species distribution history.

Several lines of evidence suggest a strong influence of the environment on species distribution: 1) Signals for *L. sibirica* appeared in its current area of distribution as late as the Holocene warming, whereas cold Pleistocene samples are dominated by *L. gmelinii* type variants; 2) in lakes far east of its modern range, signals of variants typical for *L. sibirica* coincide with peaks in read counts (29 ka BP, Lake Billyakh; 32 ka BP Lake Kyutyunda), which point to a more forested vegetation around the lakes and consequently a more favorable climate at that time; and 3) samples dated to the severely cold LGM are dominated by variants of the *L. gmelinii* type.

This is in accordance with the different ecological characteristics described for the species. *L. sibirica* is sensitive to permafrost and only occurs outside of the zone of continuous permafrost[5]. In addition, *L. sibirica* achieves substantially higher growth rates and longer growth periods than *L. gmelinii*[7,11] and can also produce more than twice as many seeds[5]. This potentially gives *L. sibirica* the ability to quickly react to climate change and outcompete the other species when climate becomes more favorable.

In contrast, *L. gmelinii* is adapted to extremely low soil and air temperatures and is able to grow on permafrost with very shallow thaw depths. It's distribution almost completely coincides with continuous permafrost[5], and even a restriction to permafrost areas is discussed as it does not grow well in field trials on warmer soils or where there is a small temperature gradient between air and soil[7]. Due to this ecology, *L. gmelinii* is more likely to survive in a high latitude refugium, even during the severe continental climate of the LGM, which was most probably connected to a continuous permafrost of low active-layer depths.

A study combining mitochondrial barcoding on *sed*aDNA and a modelling approach on *Larix* distribution in the Taymyr region around Lake CH12 concluded that the distributions of *L. gmelinii*

and *L. sibirica* are most strongly influenced by stand density and thus by competition between the species, with *L. gmelinii* outcompeting *L. sibirica* at high stand densities[11]. As our study includes sediment cores reaching further back in time, we see a different trend. Instead of *L. gmelinii*, it was *L. sibirica,* which dominated samples with high read counts, suggesting high stand density and more favorable climate. A possible explanation for the different outcomes is the use of different organelle genomes. Epp et al.[11] used a marker representing the mitochondrial genome, which is known to introgress more rapidly and as a consequence might show a long past species history[59,60].

Our findings have potentially important implications for the projections of vegetation-climate feedback. A warming climate in conjunction with a greater permafrost thaw depth could enable the replacement of *L. gmelinii* by *L. sibirica.* In contrast to *L. gmelinii, L. sibirica* is not known to stabilize permafrost thus potentially further promoting permafrost thaw and with it the release of greenhouse gases, creating positive feedback on global warming[9]. On the other hand, the substantially higher growth rates of *L. sibirica* in comparison to *L. gmelinii* would increase carbon sequestration, thus mitigating global warming[11]. This shows the importance of understanding species specific reactions to climate change, which can result in great shifts of distribution. Target capture applied on *sed*aDNA is able to reveal the impact of past climate change on populations and with the increasing availability of modern reference genomes will further enhance its value of information.

4.4 Conclusion

Our data demonstrate that Siberian larches have formed the dominant forest biomes across vast areas of Siberia since the last glacial. This is the first time that a combination of nuclear repeats and plastid DNAs have been tracked backwards in time to temporally resolve geospatial species distributions. Enrichment of the nuclear satellite repeat of *Larix* clearly indicated presence of larch DNA in the *sed*aDNA sample, whereas the captured plastid DNA also allowed differentiation of the *Larix* species. The observed genetic variation in the target-enriched *sed*aDNA gave us unprecedented insights into ancient *Larix* species distribution and led to four main conclusions: 1) environment (likely permafrost) plays a more important role than biogeography for *L. sibirica*; 2) *L. sibirica* was more widely distributed during the MIS3; 3) refugial populations existed and were almost exclusively composed of *L. gmelinii*; 4) *L. sibirica* did not recolonize its current distribution area out of northern refugial populations, but migrated from more southern refuges. Our high-resolution *Larix* projections highlight the importance of considering species traits in vegetation-climate-feedbacks.

4.5 Material & methods

4.5.1 Sample material

Samples from eight lake sediment records were included in the analysis (Supplementary Table 4). Samples from seven lake records were included in the capture enrichment targeting both the chloroplast genome and a set of nuclear genes of *Larix*, namely lakes Bolshoye Shchuchye (67.53°N, 66.18°E), Lama (69.32°N, 90.12°E), Kytuyunda (69.38°N,123.38°E), Satagay (64.10°N, 122.15°E), Billyakh (65.17°N, 126.47°E), Malaya Chabyda (61.96°N, 129.41°E) and Emanda (65.17°N, 135.45°E). DNA samples from Bolshoye Shchuchye were kindly provided by Inger Alsos and the ancient DNA laboratory of The Arctic University of Norway. In a second capture enrichment, targeting only the set of *Larix* nuclear genes, only samples from Lake CH12 (72.39°N,

102,29°E) were included. DNA data of Lake CH12 samples of a capture enrichment targeting the *Larix* chloroplast genome were obtained from the study of Schulte et al.[27]. Age models for the sediment records were adopted from[43,46,47,61,62]. The age model for Lake Satagay was established in this study.

Extracted DNA was used for single-stranded DNA library preparation following the protocol of [63,64] with slight adjustments as described in[27]. For the first dataset, two hybridization probe (bait) sets were used for target enrichment: 1) a set of 16 long range PCR products covering the chloroplast genome of *Larix* excluding the ribosomal RNA regions, and 2) a set of 65 candidate adaptive genes harboring in the nuclear genome of *Larix* as PCR products with published primers by[65,66]. Negative controls were included in DNA extraction batches and library preparation batches. Controls of DNA extractions were partly pooled together for library preparation. Sample libraries were equimolarly pooled by lake, both bait sets were jointly captured in two consecutive rounds of hybridization capture, following the procedure as described in[67]. The captured pools were pooled for paired-end (2x250 bp) sequencing on a NovaSeq6000.

For the second dataset of Lake CH12 samples, only the nuclear gene probe set was used in one round of hybridization capture, performed as described in[27] without adjustments. The second experiment was sequenced on one lane of an Illumina MiSeq with 2x150 bp paired-end sequencing.

To compare enrichment of both chloroplast and repeat data, published pollen and metabarcoding records for our study sites were used. Additional metabarcoding datasets were produced for lake Billyakh and in coarse resolution for lakes Kyutyunda, Lama, Malaya Chabyda and Satagay (Table S9). Pollen datasets from lakes Billyakh, CH12, Emanda and Lama were retrieved from the taxonomically harmonized pollen database collected by[44]. Pollen data from Lake Bolshoye Shchuchye were obtained from[45]. Pollen data from Lake Kyutyunda were obtained from[46] (parallel core PG2022) and additional samples were counted in this study from core PG2023.

A more detailed description of sample material and laboratory procedures is given in the supplementary information.

4.5.2 Sequence data analysis

Quality controlled reads were first classified against a plastid database from NCBI refseq[68] using KRAKEN2[69] with the very conservative confidence threshold of 0.8 and only reads classified to *Larix* at genus or species level were used for further analysis. Reads were aligned against an *L. gmelinii* chloroplast genome (NCBI GenBank accession number MK468637.1) and called variants compared to variants from all publicly available complete chloroplast genomes of *L. gmelinii, L. cajanderi* (NCBI GenBank accession numbers MK468630.1-36.1 & 38.1-48.1) and *L. sibirica* (NCBI GenBank accession number NC_036811.1). Reads from the hybridization capture study of Lake CH12[27] were included in the full analysis.

In a second line of analysis, reads were classified with kraken2 against the plant database from NCBI RefSeq, prepared with the built-in function of kraken2, with a manually added *L. sibirica* genome (NCBI GenBank accession number NWUY0000000000.1) as well as available genomes of Pinaceae (*Picea abies, P. glauca, Pinus taeda,* NCBI GenBank accession numbers GCA_900067695.1, GCA_000411955.6 and GCA_000404065.3). Classification was done using a confidence threshold of 0.8. Reads classified as *Larix* were mapped both to the targeted bait set

and the published *Larix* satellite DNAs[30] (EulaSat1-5). Nucleotide frequency per position was analyzed for the most abundant satellite repeat EuLaSat1[30]. A detailed description of the bioinformatic analysis is given in the supplementary information along with detailed results.

4.6 Data availability

The Illumina sequence data of the hybridization capture dataset, targeting both the chloroplast and a set of nuclear genes of *Larix* on 64 samples and 19 negative controls from seven lake sediment records and the hybridization capture dataset, targeting only the set of nuclear genes of *Larix* on 4 samples and 2 negative controls from Lake CH12 are submitted to the European Nucleotide Archive under the project number PRJEB47872.

4.7 Acknowledgments

We gratefully acknowledge John-Inge Svendsen, Jan Mangerud and Haflidi Haflidason for providing the Bolshoye Shchuchye sediment core samples. We thank Svetlana Karachurina for DNA extraction and Janine Klimke for help with DNA library preparation. We also thank Cathy Jenks for English language proofreading. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No. 772852, ERC Consolidator Grant 'Glacial Legacy') and the Initiative and Networking Fund of the Helmholtz Association.

4.8 Author contributions

L.S. did core sampling, planned the lab work, prepared the chloroplast baits for capture and conducted bioinformatic data analysis supervised by U.H., S.M. prepared nuclear gene baits and did the capture lab work, K.S. advised lab work, T.H and N.S. analyzed repetitive elements in modern *Larix*, B.v.H. prepared libraries in the lab, A.A.A. counted Kyutyunda pollen samples, B.D. coordinated field work for the Billykah sediment record, B.K.B. coordinated field work for Kyutyunda, Malaya Chabyda and Satagay sediment records and did the age model for the Satagay record, B.W. coordinated field work for the Emanda sediment record, M.M. coordinated field work for the Lama sediment record, L.A.P. and U.H. coordinated filed work for the CH12 sediment record, I.G.A. and C.C. contributed Bolshoye Shchuchye *sed*aDNA samples, K.V.K. advised nuclear gene bait design. L.S. wrote the manuscript supervised by U.H. that all co-authors commented on.

4.9 References

1. Miller, G. H. *et al.* Arctic amplification: Can the past constrain the future? *Quat. Sci. Rev.* **29**, 1779–1790 (2010).

2. Biskaborn, B. K. *et al.* Permafrost is warming at a global scale. *Nat. Commun.* **10**, 1–11 (2019).

3. Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D. & Willis, K. J. Sensitivity of global terrestrial ecosystems to climate variability. *Nature* **531**, 229–232 (2016).

4. Herzschuh, U. Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Glob. Ecol. Biogeogr.* **29**, 198–206 (2020).

5. Abaimov, A. P. Geographical Distribution and Genetics of Siberian Larch Species. in *Permafrost ecosystems Siberian larch forests* 17–36 (2010).

6. World Flora Online. *WFO* http://www.worldfloraonline.org (2021).

7. Dylis, N. V. *Listvennitsa (Larch)*. (Lesnaya Promyshlennost, 1981).

8. Zhang, N., Yasunari, T. & Ohta, T. Dynamics of the larch taiga-permafrost coupled system in Siberia under climate change. *Environ. Res. Lett.* **6**, (2011).

9. Fisher, J. P. *et al.* The influence of vegetation and soil characteristics on active-layer thickness of permafrost soils in boreal forest. *Glob. Chang. Biol.* **22**, 3127–3140 (2016).

10. Juřička, D. *et al.* Large-scale permafrost degradation as a primary factor in *Larix* sibirica forest dieback in the Khentii massif, northern Mongolia. *J. For. Res.* **31**, 197–208 (2018).

11. Epp, L. S. *et al.* Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. *Sci. Rep.* **8**, 17436 (2018).

12. LePage, B. A. & Basinger, J. F. The evolutionary history of the genus *Larix* (Pinaceae). *U.S. Dept. Agric., Ser., Intermt. Res. Sta.* **GTR-INT-31**, 19–219 (1995).

13. Monegato, G., Scardia, G., Hajdas, I., Rizzini, F. & Piccin, A. The Alpine LGM in the boreal ice-sheets game. *Sci. Rep.* **7**, 1–8 (2017).

14. Anderson, P. M. & V. Lozhkin, A. The stage 3 interstadial complex (Karginskii/middle Wisconian interval) of Beringia: Variations in paleoenvironments and implications for paleoclimatic interpretations. *Quat. Sci. Rev.* **20**, 93–125 (2001).

15. Binney, H. A. *et al.* The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quat. Sci. Rev.* **28**, 2445–2464 (2009).

16. Cao, X. *et al.* Pollen-based quantitative land-cover reconstruction for northern Asia covering the last 40 ka cal BP. *Clim. Past* **15**, 1503–1536 (2019).

17. Brubaker, L. B., Anderson, P. M., Edwards, M. E. & Lozhkin, A. V. Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *J. Biogeogr.* **32**, 833–848 (2005).

18. Tarasov, P., Müller, S., Andreev, A., Werner, K. & Diekmann, B. Younger Dryas *Larix* in eastern Siberia: A migrant or survivor? *PAGES news* **17**, 122–123 (2009).

19. Lozhkin, A. *et al.* Implications for conifer glacial refugia and postglacial climatic variation in western Beringia from lake sediments of the Upper Indigirka basin. *Boreas* **47**, 938–953 (2018).

20. Niemeyer, B., Klemm, J., Pestryakova, L. A. & Herzschuh, U. Relative pollen productivity estimates for common taxa of the northern Siberian Arctic. *Rev. Palaeobot. Palynol.* **221**, 71–82 (2015).

21. Semerikov, V. L., Semerikova, S. A., Polezhaeva, M. A., Kosintsev, P. A. & Lascoux, M. Southern montane populations did not contribute to the recolonization of West Siberian

Plain by Siberian larch (*Larix* sibirica): a range-wide analysis of cytoplasmic markers. *Mol. Ecol.* **22**, 4958–4971 (2013).

22. Semerikov, V. L. & Lascoux, M. Genetic relationship among Eurasian and American *Larix* species based on allozymes. *Heredity (Edinb).* **83**, 62–70 (1999).

23. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2. *R J.* **5**, 144–161 (2013).

24. Jørgensen, T. *et al.* A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Mol. Ecol.* **21**, 1989–2003 (2012).

25. Zimmermann, H. *et al.* The History of Tree and Shrub Taxa on Bol'shoy Lyakhovsky Island (New Siberian Archipelago) since the Last Interglacial Uncovered by Sedimentary Ancient DNA and Pollen Data. *Genes (Basel).* **8**, 273 (2017).

26. Liu, S. *et al.* Sedimentary ancient DNA reveals a threat of warming-induced alpine habitat loss to Tibetan Plateau plant diversity. *Nat. Commun.* **12**, 1–9 (2021).

27. Schulte, L. *et al.* Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Mol. Ecol. Resour.* **21**, 801–815 (2021).

28. Nystedt, B. *et al.* The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**, 579–584 (2013).

29. Jagannathan, M., Cummings, R. & Yamashita, Y. M. A conserved function for pericentromeric satellite DNA. *Elife* **7**, 1–19 (2018).

30. Heitkam, T. *et al.* Comparative Repeat Profiling of Two Closely Related Conifers (*Larix* decidua and *Larix* kaempferi) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs. *Front. Genet.* **12**, (2021).

31. Garrido-Ramos, M. A. Satellite DNA: An evolving topic. *Genes (Basel).* **8**, (2017).

32. Weitemier, K. *et al.* Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant Phylogenomics. *Appl. Plant Sci.* **2**, 1400042 (2014).

33. Costa, L. *et al.* Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA. *Ann. Bot.* 1–14 (2021) doi:10.1093/aob/mcab063.

34. Hizume, M. *et al.* Tandem repeat DNA localizing on the proximal DAPI bands of chromosomes in \textit{*Larix*, Pinaceae}. *Genome* **45**, 777–783 (2002).

35. Schmickl, R. *et al.* Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). *Mol. Ecol. Resour.* **16**, 1124–1135 (2016).

36. Sproul, J. S., Barton, L. M. & Maddison, D. R. Repetitive DNA profiles reveal evidence of rapid genome evolution and reflect species boundaries in ground beetles. *Syst. Biol.* **69**, 1137–1148 (2020).

37. Vitales, D., Garcia, S. & Dodsworth, S. Reconstructing phylogenetic relationships based on

repeat sequence similarities. *Mol. Phylogenet. Evol.* **147**, 106766 (2020).

38. Heitkam, T. & Garcia, S. Can we have it all? Repurposing target capture for repeat genomics. A commentary on: 'Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA.' *Ann. Bot.* **102**, 1–2 (2021).

39. Alsos, I. G. *et al.* Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. *PLoS One* **13**, 1–23 (2018).

40. Shirazi, S., Meyer, R. & Shapiro, B. Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Authorea Prepr.* 1–22 (2021) doi:10.22541/au.159309876.62184178/v2.

41. Taberlet, P. *et al.* Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14–e14 (2007).

42. de Klerk, P., Teltewskoi, A., Theuerkauf, M. & Joosten, H. Vegetation patterns, pollen deposition and distribution of non-pollen palynomorphs in an ice-wedge polygon near Kytalyk (NE Siberia), with some remarks on Arctic pollen morphology. *Polar Biol.* **37**, 1393–1412 (2014).

43. Clarke, C. L. *et al.* Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. *Sci. Rep.* **9**, 19613 (2019).

44. Herzschuh, U. *et al. Global taxonomically harmonized pollen data collection with revised chronologies*. https://doi.pangaea.de/10.1594/PANGAEA.929773 (2021) doi:doi.pangaea.de/10.1594/PANGAEA.929773.

45. Clarke, C. L. *et al.* A 24,000-year ancient DNA and pollen record from the Polar Urals reveals temporal dynamics of arctic and boreal plant communities. *Quat. Sci. Rev.* **247**, 106564 (2020).

46. Biskaborn, B. K. *et al.* Late Quaternary vegetation and lake system dynamics in north-eastern Siberia: Implications for seasonal climate variability. *Quat. Sci. Rev.* **147**, 406–421 (2016).

47. Müller, S. *et al.* Late Quaternary vegetation and environments in the Verkhoyansk Mountains region (NE Asia) reconstructed from a 50-kyr fossil pollen record from Lake Billyakh. *Quat. Sci. Rev.* **29**, 2071–2086 (2010).

48. Semerikov, V. L. & Lascoux, M. Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae) species. *Am. J. Bot.* **90**, 1113–23 (2003).

49. Polezhaeva, M. A., Lascoux, M. & Semerikov, V. L. Cytoplasmic DNA variation and biogeography of *Larix* Mill, in Northeast Asia. *Mol. Ecol.* **19**, 1239–1252 (2010).

50. Dylis, N. V. *Larch of Eastern Siberia and Far East*. (Academy of Sciences of USSR, 1961).

51. Dylis, N. V. *Siberian larch. Materials on taxonomy, geography, and history*. (Odshch Issled Prir, Series "Botany," 1947).

52. Kullmann, L. Palaeoecological , Biogeographical and Palaeoclimatological Implications of Early Holocene Immigration of *Larix* sibirica Ledeb . into the Scandes Mountains , Sweden.

Glob. Ecol. Biogeogr. Lett. **7**, 181–188 (1998).

53. Wieczorek, M. *et al.* Dissimilar responses of larch stands in northern Siberia to increasing temperatures—a field and simulation based study. *Ecology* **98**, 2343–2355 (2017).

54. Neilson, R. P. *et al.* Forecasting regional to global plant migration in response to climate change. *Bioscience* **55**, 749–759 (2005).

55. Semerikov, V. L., Semerikova, S. A., Putintseva, Y. A., Oreshkova, N. V. & Krutovsky, K. V. Mitochondrial DNA in Siberian conifers indicates multiple postglacial colonization centers. *Can. J. For. Res.* **49**, 875–883 (2019).

56. Andreev, A. A. *et al.* Vegetation and climate changes around the Lama Lake, Taymyr Peninsula, Russia during the Late Pleistocene and Holocene. *Quat. Int.* **122**, 69–84 (2004).

57. Kruse, S., Kolmogorov, A. I., Pestryakova, L. A. & Herzschuh, U. Long-lived larch clones may conserve adaptations that could restrict treeline migration in northern Siberia. *Ecol. Evol.* **10**, 10017–10030 (2020).

58. Khatab, I. A., Ishiyama, H., Inomata, N., Wang, X.-R. & Szmidt, A. E. Phylogeography of Eurasian *Larix* species inferred from nucleotide variation in two nuclear genes. *Genes Genet. Syst.* **83**, 55–66 (2008).

59. Currat, M., Ruedi, M., Petit, R. J. & Excoffier, L. The hidden side of invasions: Massive introgression by local genes. *Evolution (N. Y).* **62**, 1908–1920 (2008).

60. Neale, D. B. & Wheeler, N. C. *The Conifers: Genomes, Variation and Evolution*. *The Conifers: Genomes, Variation and Evolution* (2019). doi:10.1007/978-3-319-46807-5.

61. Klemm, J., Herzschuh, U. & Pestryakova, L. A. Vegetation, climate and lake changes over the last 7000 years at the boreal treeline in north-central Siberia. *Quat. Sci. Rev.* **147**, 422–434 (2016).

62. Baumer, M. M. *et al.* Climatic and environmental changes in the Yana Highlands of north-eastern Siberia over the last c. 57 000 years, derived from a sediment core from Lake Emanda. *Boreas* **50**, 114–133 (2021).

63. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).

64. Gansauge, M.-T. *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, gkx033 (2017).

65. Mosca, E. *et al.* Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evol. Appl.* **5**, 762–775 (2012).

66. Semerikov, V. L., Semerikova, S. A. & Polezhaeva, M. A. Nucleotide diversity and linkage disequilibrium of adaptive significant genes in *Larix* (Pinaceae). *Russ. J. Genet.* **49**, 915–923 (2013).

67. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).

68. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

69. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *bioRxiv* 762302 (2019) doi:10.1101/762302.

5 Discussion and synthesis

The overall goal of this thesis was to contribute to the understanding of the Asian larch forest dominance and how they recolonized northern Asia after the LGM. The three research articles presented in this thesis all contributed differently to this goal. In the following I synthesize and discuss the findings of the conducted research.

5.1 Hybridization capture is a well-suited method to study ancient species dynamics

An important part of this thesis was to establish a method which would be suitable to study *Larix* species dynamics in the past. Therefore, we tested the potential of target enrichment by hybridization capture using the complete *Larix* chloroplast genome by sequencing both unenriched samples, so called "shotgun" sequenced samples and target enriched "capture" samples.

5.1.1 Advantages and limitations of shotgun sequencing

The shotgun sequenced samples allow to analyze plants as well as bacteria, archaea and fungi (Ahmed et al., 2018; Pedersen et al., 2016). However, due to database limitations and short reads (highly fragmented by age) only a small fraction of the data can be reliably assigned when compared to a taxonomic sequence database. In our case, only 0.3% of the quality-filtered reads (i.e. reads shorter than 30 bp are already excluded) could be assigned to the non-redundant nucleotide database (nt) of NCBI (manuscript II). This number can be increased by lowering bioinformatic stringency parameters or choosing more sensitive but less specific assignment tools (Harbert, 2018). This, however, reinforces the existing database bias which is especially problematic in the nt database, meaning that well-studied and deeply sequenced taxa are more likely to be assigned (Parducci et al., 2017). Even with these restrictions, an analysis of genetic variation is to some extent possible from (deeply sequenced) shotgun data if the species in question is sufficiently abundant (Meucci et al., 2021). Nonetheless, the fraction of reads of a specific target is very low. In our dataset, even with the lowest stringency parameters in classification, only 0.001% of the reads were classified as genus *Larix* using complete chloroplast genomes of Genbank as database. Thus, the amount of sequencing needed to study one taxon group in detail is not cost efficient with shotgun sequencing.

5.1.2 Successful hybridization capture enrichment using chloroplast DNA

Our study showed that target enrichment by hybridization capture using the complete chloroplast genome of *Larix* increased the number of assigned reads by two orders of magnitude. The number of assigned reads increased from 0.3% for shotgun sequencing to 28% for our dataset and the complete *Larix* chloroplast genome was retrieved (manuscript II). In comparison to PCR-based techniques, this allowed us to study many variable positions along the chloroplast genome. However, to access this depth of information a complex analysis of the data is needed. The first challenge is to identify molecules as belonging to the taxon of interest which was targeted by the baits. We first taxonomically classified the reads based on a well curated database with a lowest common ancestor approach (Wood et al., 2019) and only used *Larix*-assigned reads for mapping. The thus filtered result is an alignment of very uneven coverage with those regions best covered which can be unambiguously identified as belonging to the genus *Larix.* These are mostly non-coding regions or pseudogenes as ycf1 and ycf2 harboring most of the genetic variation on the *Larix* chloroplast (Guo et al., 2021). The ycf1-pseudogene is a variation hotspot in many plant

groups and was for this reason even proposed as a plant marker (Dong et al., 2015). It is thus a highly valuable region for the analysis of genetic variation. On the other hand, other regions are highly conserved between species and get filtered out by the lowest common ancestor approach. The most conserved region between plant species is coding for ribosomal RNA. Due to high similarities between chloroplast sequences of plants, Meucci et al. (2021) could also use reads of our *Larix*-capture dataset to analyze chloroplast variation in arctic shrubs. This is possible as the wet-lab hybridization capture protocol is less specific than bioinformatic analyses tools. However, in the experiments for our subsequent project (manuscript III), we left out the very conserved sites of ribosomal RNA, as this region had the highest coverage prior to taxonomic assignment and the lowest when using only reads assigned to the genus *Larix.*

5.1.3 Challenges in single-copy target enrichment

In the follow-up project of manuscript II, we upscaled the approach on samples from 8 lakes across Siberia. In addition to using the chloroplast as target in hybridization capture, we also included 65 candidate adaptive genes of the *Larix* nuclear genome in the bait set. However, this enrichment attempt was not successful. Instead of the nuclear genes, repetitive elements of the *Larix* nuclear genome got enriched, namely the most abundant satellite repeat of *Larix,* EulaSat1 (Heitkam et al., 2021; Hizume et al., 2002). In the especially large genomes of conifers such as *Larix,* repetitive elements make up for about 80% of the genome (Nystedt et al., 2013). Satellite repeats are one class of repetitive elements, characterized by a length of about 150 to 360 bp and their formation of long arrays of tandemly repeated non-coding DNA stretches (Plohl et al., 2012). Our analysis of modern genomes of the Siberian larches showed that sequences of the most abundant satellite, EulaSat1, made up for 0.62% (*L. sibirica)* to 2.52% (*L. gmelinii)* of the respective genomes (manuscript III). As the *Larix* genome spans around 12 to 13 giga bp (Kuzmin et al., 2019), EulaSat1 sequences add up to 60 to 300 mega bp. Off-target enrichment of repetitive elements is a known phenomenon in hybridization capture studies on modern individuals and is used to assemble multi-copy regions such as the chloroplast or mitochondrial genomes (Schmickl et al., 2016; Weitemier et al., 2014). Such an off-target "by-catch" of multi-copy sequences in individuals would not have been expected from *sed*aDNA samples. Here, the huge "background noise" is not the repetitive part of a single species, but the majority of reads are either non-assignable or of prokaryotic origin (manuscript II). It is therefore more likely that we specifically enriched the *Larix* satellite repeat in our capture approach. By the comparison of two datasets originating from identical samples, capturing only the chloroplast in one dataset and only the nuclear genes in the other dataset, we could show that the satellite repeat got specifically enriched by the nuclear gene bait set (manuscript III). This could either be a result of contamination of the bait set with sequences of the satellite repeat, or partial sequence similarities between the targeted genes and the satellite sequence. The specific enrichment of the *Larix* satellite repeat in our experiment is also corroborated by the repeat data showing the same abundance patterns as the other proxies used for comparison in our study (i.e. enriched chloroplast DNA, pollen and ancient DNA metabarcoding).

In the case of *Larix* satellite EulaSat1, we could not detect any trends in variation over time and space. This is probably related to the young evolutionary history of Asian *Larix* species, which are morphologically very similar, easily hybridize in nature and are still in an ongoing speciation process (Abaimov, 2010; Polezhaeva et al., 2010). In general, satellite repeats are among the fastest evolving repetitive elements which can be used both for abundance and sequence variation in phylogenetic analyses (Costa et al., 2021; Heitkam et al., 2021; Vitales et al., 2020). Sequences of ancient satellite repeats have hitherto not been reported and little is known about

their development over millennia. Although we were not successful in enriching nuclear genes, we could still show that off-target reads in *sed*aDNA capture samples can hold valuable information and that satellite repeats can be an interesting target for future studies.

5.1.4 Limitations and potentials to improve *sed*aDNA capture studies

Our studies (manuscript II and III) showed that chloroplast enrichment by hybridization capture is a powerful method to track *Larix* species dynamics in time and space. However, we were not able to delineate between single populations. A current limitation is the lack of reference sequences – only the sequence of a single chloroplast genome of *L. sibirica* is hitherto published (Bondar et al., 2019). The analysis could greatly benefit from multiple reference genomes per species from different geographic origins. A more important limitation is however, that chloroplast genomes of the two eastern Asian species *L. gmelinii* and *L. cajanderi* carry few sequence variations and even over long distances do not show spatial structure (Zimmermann et al., 2019). It is thus not possible to distinguish between these two species using chloroplast DNA. This is why we chose to summarize them in manuscript III in the broader sense of *L. gmelinii*, which comprises both *L. gmelinii* and *L. cajanderi* (Abaimov, 2010). Modern genetic analysis of Siberian larches showed that mitochondrial markers are better in delineating populations (Polezhaeva et al., 2010; Semerikov et al., 2013). Plant mitochondrial genomes are big and often complex and only few have been published so far (Putintseva et al., 2020). The mitogenome of *L. sibirica* has only recently been sequenced and assembled; it is highly complex (containing nine contigs) and with 11.7 mega bp the currently largest mitogenome reported (Putintseva et al., 2020). The *Larix* mitogenome, or variable parts of it, represent highly interesting targets for future *sed*aDNA capture studies and would likely allow to track past *Larix* dynamics on population level.

In both studies (manuscript II and III), our data shows high values of non-target reads. Only about 3.3% of the quality-filtered and deduplicated reads of the first study (manuscript II) were classified as *Larix* using a chloroplast database. With the changed bioinformatic pipeline of manuscript III with prior deduplication and a higher confidence stringency on classification, this number is reduced to 0.56%. Studies on capture efficiency recommend two consecutive rounds of hybridization capture for samples with low concentration of the target (Furtwängler et al., 2020; Hernandez-Rodriguez et al., 2018; White et al., 2019). As the *Larix* chloroplast poses such a small target in lake sediments (in the shotgun dataset (manuscript II), *Larix* reads constituted only 0.001% of the quality-filtered reads), we applied the double capture approach in the second capture project (manuscript III). However, the percentage of *Larix*-classified reads form deduplicated, quality-filtered reads resulted only in 0.14%. It seems thus that the second round of capture was not effective in increasing the proportion of target reads. However, this could be due to the difference between the samples, with the second study comprising many more samples with low expected proportion of *Larix* in the vegetation, e.g. samples dated to the LGM.

Both datasets contained high values of PCR duplicates. This is due to the high number of PCR amplification cycles which are necessary to obtain a sufficient concentration, since the capture approach only retains a small fraction from the original DNA. When applying two consecutive rounds of capture, three rounds of PCR amplification are necessary (before, in between and after the capture rounds). To avoid this in future studies, it can be beneficial to try improving on target proportion using other methods than double-capture. This could include optimizing capture specificity parameters like temperature, including the use of temperature ramps, or salt concentration (Paijmans et al., 2016).

5.2 Factors promoting Asian larch dominance

Although larch species are prevalent on all northern hemisphere continents, their distribution differs markedly (Herzschuh, 2020). Larches have their greatest distribution in northern Asia, where they form vast forests of mostly monospecific tracts and only in their western range they are replaced in dominance by evergreen species (Abaimov et al., 1999). North American larches have a wide distribution across the continent but pure old-growth stands are often confined to ecologically marginal sites and tree line habitats (Cheliak et al., 1988). Dominant boreal forest forming species in North America are often evergreen conifers including *Picea mariana, P. glauca, Abies balsamea* and *Pinus banksiana* (Warren et al., 2016). In Europe, native populations of *Larix* are absent from the boreal forests in Fennoscandia and *Larix* naturally occurs only in the central European mountain ranges as the Alps, Carpathians, Tatra Mountains and Sudetes and with a small population in the Polish lowlands (Da Ronch et al., 2016). This pronounced difference of distribution is still not well understood.

In manuscript I, we provide a synthesis and a detailed comparison of *Larix* distribution on the northern hemisphere and the factors that control its domination. Our comparative analysis of *Larix* in the northern hemisphere firstly states concrete numbers of the existing differences. Whereas European *L. decidua* and American *L. laricina* dominate only in 0.2% of their whole range, the Asian species dominate in 11% (*L. sibirica*) and 55% (*L. gmelinii* and *L. cajanderi*) of their respective ranges. This striking disparity is the starting point for our question of which factors are leading to the dominance of Asian larches.

The assertion that Asian *Larix* dominance is conveyed by its good adaptation to harsh climates (Kharuk et al., 2007) could only partly be confirmed. Our results show that only the most continental climates are covered exclusively by two Asian larches *L. gmelinii* and *L. cajanderi.* At the same time the two species dominate also vast areas outside of these extreme climates. Thus, climate cannot be the only factor. This is in line with the results of Herzschuh (2020) who showed that analogue climate spaces in North America and northern Asia are dominated by evergreen needle trees on the former and summergreen larches on the latter continent.

We also comparatively analyzed how the distribution of *Larix* dominance is allocated over different depth of active layer thickness (ALT, i.e. the layer thickness of the ground that is subject to annual thawing and freezing in areas underlain by permafrost (Obu et al., 2021a)), as well as over different extents of permafrost (i.e. the yearly fraction of permafrost-underlain and permafrost-free area within a pixel of approximately 1 km^2 (Obu et al., 2021b)). The ability to grow on extremely low permafrost active layer depths has been well documented for the Asian species *L. gmelinii* and *L. cajanderi* (Abaimov, 2010; Abaimov et al., 1999; Dylis, 1981). American *L. laricina* grows outside of the zone of continuous permafrost with only a small distribution in zones of discontinuous permafrost. In our analysis we show that the small fraction of the *L. laricina* populations on permafrost indeed grows on comparatively low ALT of around 1 meter, lower than the average distribution on ALT of *L. gmelinii.* This suggests that the ability to grow on low ALT is not sufficient to convey wide spread *Larix* dominance.

The connection of Asian larches and the extent of permafrost had long been proposed (e.g. Dylis, 1981). Our results corroborate these findings and show that dominance of *L. cajanderi* is almost exclusively in areas of continuous permafrost, and that dominance of *L. gmelinii* is clearly shifted towards high extent of permafrost. Beside the coupled system of larch vegetation and permafrost, the frequency of fire-return intervals has been discussed as an important factor promoting *Larix* dominance (Schulze et al., 2012; Uemura et al., 1997). Asian larches *L. gmelinii*

and *L. cajanderi* are fire-resistant with thick barks, the shedding of low branches, and serotinous cones which are stored for several years in the tree canopy (Osawa et al., 2010b). With seeds from the surviving trees available, they regenerate well on burned sites whereas evergreen succession is hindered if fire-return-frequencies are below 300 years (Schulze et al., 2012; Uemura et al., 1997). Interestingly, although the permafrost sensitive *L. sibirica* has similar fire-resisting traits, wild fires typically are followed by stand replacement through *Betula sp.* and larch regeneration takes many decades (Abaimov & Sofronov, 1996). In contrast, the American *L. laricina* trees and seeds get easily killed by fire as it possesses a thin bark and non-serotinous cones (Brown & Zobel, 1988; Busque & Arseneault, 2005; Johnston, 1990). Therefore, it seems that it is the interaction of several factors which are needed to secure larch dominance. Two of these factors are permafrost and fire. Adaptations to permafrost give *L. gmelinii* and *L. cajanderi* a competitive advantage above other species in permafrost regions and adaptations to regular occurring disturbances as fire help to maintain this dominance. Both species are not only able to grow on continuous permafrost but also efficiently decouple permafrost from climate (Zhang et al., 2011). By intercepting solar radiation and evapotranspiration of soil moisture it reduces heat flux to reach deeper soils and thus prevents permafrost thaw (Fisher et al., 2016; Zhang et al., 2011). On the other hand, permafrost with low seasonal active layer depth impedes the invasion of other tree species as *Picea, Abies* or *Pinus* which are not as well adapted to grow on permafrost (Kharuk et al., 2007). Thus, by a positive feedback mechanism between soil and vegetation, with the additional factor of wild fire frequencies, *L. gmelinii* and *L. cajanderi* can dominate in climatic areas which would otherwise be dominated by evergreen species. The importance of both fire-disturbance and permafrost adaptations has recently been confirmed by modelling (Abis & Brovkin, 2019).

In addition to these two factors, the glacial history has been proposed as a third factor, i.e. Asian larch dominance as a legacy of the severe Last Glacial (Herzschuh, 2020; Herzschuh et al., 2016). To investigate this factor, we compared glacial refugia and post-glacial migration patterns in our analysis (manuscript I). In Europe and most of North America, except for ice free Alaska, *Larix* populations were pushed to the south as big ice shields were covering the continents (Warren et al., 2016). In Asia, apart from mountain glaciers, the continent was mostly ice free and *Larix* survival in northern refugia was documented for a number of sites across the continent (Binney et al., 2009; Brubaker et al., 2005; Müller et al., 2010). On the example of the Alaskan population we show that glacial survival is not sufficient to ensure larch domination and that the genetic fitness of glacial refugial populations probably is an important factor. The unique situation of the Alaskan larches, being separated from the rest of the North American populations by the geographical barrier of the Rocky Mountains resulted in a genetic impoverishment as revealed by modern genetic analysis (Napier et al., 2020; Warren et al., 2016). The patterns of *Larix* proxies in northern Asia during and after the LGM show a wide spread occurrence across Siberia and a fast dispersal after the LGM (manuscript I). In contrast to the other continents, for Asia there is no clear direction of recolonization visible from the paleo proxy data. This hints towards propagation out of high latitude refugial populations as opposed to recolonization from a southern front (Brubaker et al., 2005). However, this is connected to a high uncertainty as the sample density in northern Asia is very low (Li et al., 2022). Only from pollen, macrofossil and metabarcoding data it is therefore difficult to infer the origin of postglacial colonization.

Our results of chloroplast data captured from *sed*aDNA samples of lakes across Siberia give for the first time important insights on recolonization patterns (manuscript III). We show that glacial survival in high latitudes differed between the *L. sibirica* and *L. gmelinii* (*sensu lato*). For the analyzed sites we could show that high latitude glacial refugia were constituted by *L. gmelinii,*

also in the current distribution range of *L. sibirica.* Invariable patterns of *L. gmelinii* in eastern samples could hint to a continuous population without invasion events, for which also modern genetic analysis give hints (Polezhaeva et al., 2010). Samples from western sites, in the current range of *L. sibirica,* show a turnover in the genetic variants of the chloroplast DNA. Pre-Holocene samples show the presence of *L. gmelinii,* and only in the Holocene samples we see patterns of *L. sibirica*. This is a strong indication for *L. sibirica* recolonizing its current range not from northern refugia but from populations situated more to the south. This was also suggested from modern genetic analysis of *L. sibirica* populations. Semerikov et al. (2013) stated that *L. sibirica* migrated in the West Siberian Plain from populations in the foothills of southern mountain ranges and that northern glacial refugia contributed only to a limited degree. This minor contribution of northern refugia could be originating from refugial populations of *L. gmelinii* which got invaded by *L. sibirica* during the Holocene. With the onset of climate warming after the LGM *Picea* also started to migrate out of southern mountain ranges into the West Siberian Plain (Tollefsrud et al., 2015). The situation of *L. sibirica* would thus have been similar to the situation of *L. laricina* in North America, i.e. a northward migration of larch accompanied or shortly followed by evergreen species (Warren et al., 2016). And both species, *L. sibirica* and *L. laricina* are today confined to less favorable habitats as poorly drained wetlands (*L. laricina*, Cheliak et al., 1988) or areas of harsh climatic conditions as the latitudinal and altitudinal tree line (*L. sibirica,* Kharuk et al., 2007). These differences between the *L. sibirica* and *L. gmelinii* in high latitude glacial survival and recolonization mode could have influenced their role in forming the dominant tree species of today. That is, a substantial advance of larches over evergreen conifers in the postglacial recolonization may have enhanced, or even be a prerequisite for their present dominance.

5.3 Drivers of *Larix* species distribution

Our results of chloroplast capture data across Siberia reveal two strong patterns (manuscript III): 1) In the past, *L. sibirica* had a wider distribution towards the east and 2) glacial refugial populations were constituted by *L. gmelinii,* even in regions which are today dominated by *L. sibirica.* Prior to the LGM, during the warmer interstadial Marine Isotope Stage 3 (MIS 3), genetic patterns suggest *L. sibirica* occurrences even in the Verkhoyansk Mountains (lake Billyakh) which are thousands of kilometers east of its current distribution. These results were new and surprising, although findings of fossilized cones and modern genetic analyses corroborate a wider distribution of *L. sibirica* in the past (Dylis, 1981; Semerikov et al., 2013). Another indicator is that *L. olgensis,* which occurs on the Korean Peninsula and adjacent areas in Russia and China, is genetically close to *L. sibirica,* although the two species are spatially separated by *L. gmelinii* (Polezhaeva et al., 2010; Semerikov & Lascoux, 2003). A possible explanation for this pattern is that a formerly connected distribution was disrupted by the invasion of a better cold-adapted *L. gmelinii* (Dylis, 1961; Semerikov et al., 2003). Fossilized cones of *L. sibirica* from the early Pleistocene were found in several river basins of the Russian Far East and even at the Sea of Okhotsk, thousands of kilometers east of its current distribution (Dylis, 1981). This suggests a rapid long-distance dispersal along rivers (Neilson et al., 2005). River shores have also deeper permafrost thaw depth, facilitating the successful establishment of seedlings (Wieczorek et al., 2017). In addition, a spread of *L. sibirica* during warm phases fits to the ecological traits of the species. *L. sibirica* does not grow on continuous permafrost with low active layer depth and is in general more heat demanding than *L. gmelinii* (Abaimov, 2010; Abaimov et al., 1999). On the other hand, *L. sibirica* is less light-demanding than *L. gmelinii,* grows faster and produces more seeds (Abaimov et al., 1999; Epp et al., 2018; Osawa et al., 2010b). Thus, *L. sibirica* is able to outcompete *L. gmelinii* if climatic conditions are favorable.

Another finding was that almost exclusively patterns of *L. gmelinii* are found in the LGM samples, even in those sites which are located in the current range of *L. sibirica*. This fits well with the ecological traits of *L. gmelinii*, being able to grow on continuous permafrost with low seasonal thaw depth (Abaimov, 2010). In adverse climates at the latitudinal or altitudinal treeline, *Larix gmelinii* can grow as dwarf form (Abaimov et al., 1999) and reproduce asexually by growth of adventitious roots on lateral branches touching the ground (Cooper, 1911). Today, refugial populations of *L. gmelinii* dwarf forms can be found north of the treeline as relicts of the Holocene Thermal Optimum, when the treeline extended further north (Kruse et al., 2020; MacDonald et al., 2008). In addition, modern genetic studies, in the ranges of *L. sibirica* and *L. gmelinii* found patterns suggesting a rapid recolonization from the south for *L. sibirica* (Khatab et al., 2008; Semerikov et al., 2013) and a local origin of populations of *L. gmelinii* (Polezhaeva et al., 2010). Thus, it seems highly plausible that *L. gmelinii* soley formed the refugial populations where pollen and macrofossil evidence was found in many instances (Binney et al., 2009; Brubaker et al., 2005; Lozhkin et al., 2018; Tarasov et al., 2009).

The current split of *Larix* species ranges between east and west could suggest a strong influence of biogeography in the distribution patterns, i.e. that the current distribution is determined by the past distribution and dispersal limitations of the species. Our findings show that environment is a strong determinant of *Larix* species distribution, that is species were not bound to their current ranges but were highly mobile in response to climate change. In warm phases, *L. sibirica* was able to grow far east, whereas cold phases were dominated by *L. gmelinii.* As the species substantially differ in their ecological traits, this has important implications for vegetation-climate feedbacks.

5.3.1 Implications for larch forests under climate warming

The now warming climate might favor *L. sibirica* over *L. gmelinii,* as the species is more heat-tolerant (Abaimov, 2010, manuscript I). So far, the permafrost limits the eastward extent of *L. sibirica* (Tchebakova et al., 2006) and its range is even invaded from the east by *L. gmelinii* (Epp et al., 2018; Semerikov et al., 2013, manuscript II). Climate warming causing increased permafrost thawing can lead to a reversal of the invasion pattern, i.e. *L. sibirica* invading *L. gmelinii.* Modelling studies predict an evergreen invasion of Siberian larch forest in a warming climate (Shuman et al., 2011), a phenomenon which has already been observed in field studies (Kharuk et al., 2007). However, different species traits are rarely included in modelling studies (Kruse et al., 2021; Tchebakova et al., 2006). Modelling results show that in a future warming climate, *L. gmelinii* will move towards the east and be replaced by *L. sibirica* on the western edge (Tchebakova et al., 2006). Shuman et al. (2011) showed by modelling that the theoretical introduction of heat-tolerant European *L. decidua* in central and north-western Siberia can delay or even prevent the collapse of larch forests and their replacement by evergreen species. Although *L. sibirica* occupies a much colder bioclimatic niche than *L. decidua* (manuscript I), *L. sibirica* invasion could possibly also delay an evergreen invasion into the zone of *L. gmelinii* to some extent. However, a permanent prevention of the invasion of evergreens by *L. sibirica* is less likely. As shown in manuscript I, *L. sibirica* does not dominate in its current range to the extent as *L. gmelinii* does. In great parts of its natural distribution range, evergreen conifers such as *Pinus sylvestris* and *Picea obovata* are dominant (Abaimov et al., 1999; Wirth et al., 2002). A similar positive feedback mechanism, with permafrost and fire promoting the dominance of *L. gmelinii,* does not exist for *L. sibirica.* Thus, in a warming climate, *L. sibirica* could outcompete *L. gmelinii,* but might eventually itself be pushed to ecologically marginal sites by evergreen species. Our findings

emphasize the need of including species traits in climate-vegetation models to accurately forecast the future of the Siberian boreal forests.

5.4 Conclusion

The presented study contributes to the knowledge on Asian larch dominance and their glacial history. By synthesizing available data and knowledge on major larch species of the northern hemisphere, I was able to pinpoint the main factors promoting dominance of larches in Asia. The comparison shows that it is not a single but the interaction of multiple factors leading to the vast Asian forests dominated by two species, *L. gmelinii* and *L. cajanderi.* Namely, these factors are their good adaptation to permafrost and their adaptations to regenerate well after frequent wild fires. In addition, it is very likely that northern glacial refugia did promote their early establishment and dominance. The question of the possible role of northern glacial refugia in the postglacial recolonization was further pursued in this thesis. For this, a method enabling to distinguish between *Larix* species from ancient lake sediment DNA was established, using chloroplast enrichment by hybridization capture. The results of a large-scale capture study across Siberia showed that the recolonization patterns differ between *Larix* species, more specifically between the widely dominant *L. gmelinii* and the less dominant *L. sibirica*. For the studied sites I could show that recolonization of *L. sibirica* was not enhanced by high latitude refugia, but that *L. sibirica* invaded high latitude populations of *L. gmelinii* in the course of the Holocene. On the other hand, more eastern sites in the range of *L. gmelinii* showed no turn-over in genetic variants, hinting towards a continuous population. The results support the hypothesis that northern glacial refugia contributed to the dominance of *L. gmelinii*. I further could show that *L. sibirica* had a much wider distribution towards the east during warmer MIS3. I infer that climate is a strong driver of *Larix* species distribution, which has important implications for projections of future *Larix* distributions in a warming climate.

5.5 Outlook

The applied method of target enrichment by hybridization capture using the paternally inherited chloroplast genome gave already first important insights into pre- and post-glacial migration patterns of Siberian larches. Specifically, I was able to track occurrences of *L. sibirica* and *L. gmelinii* back in time. However, the *Larix* chloroplast genome does not carry enough genetic variation to provide insights into patterns on the population level. Therefore, including the mitochondrial genome as a target for enrichment might give valuable insights in ancient *Larix* populations. As the *Larix* mitochondrial genome is large and complex, it is advisable to start with selected regions carrying known variants. As of yet, only the reference genome of the *L. sibirica* is available. The target capture analysis would gain validity by accompanying it with the assembly of additional modern mitochondrial reference genomes of *L. gmelinii* and *L. cajanderi* using publicly available short reads.

Our study included lake sediment samples of only coarse time resolution. A next step is to study one, or preferably two or more lake records, with high resolution in time to get a better picture of population processes. A high sampling depth is particularly interesting in times of climatic changes as before and after the LGM and at the beginning of the Holocene. Of special interest is also a selection of lakes along a north-south transect to gain information on the direction of postglacial recolonization.

The analysis of *sed*aDNA samples provides valuable information with a precise chronological reference. Additional information can be gained by the study of modern larch populations, also in regard to the data analysis and interpretation of the ancient samples. Few genetic studies with only limited numbers of genetic markers are reported on modern Siberian larch populations. Reduced representation sequencing methods are able to give information of millions of positions across the whole genome at comparably low costs. This can provide a good overview of variation in extant populations as well as potentially high spatial resolution of possible migration pathways of the past (depending on the number of individuals included). Combining modern information with information from *sed*aDNA can provide a detailed picture of the migration patterns of larches before, during and after the last glacial period.

6 References

Abaimov, A. P. (2010). Geographical Distribution and Genetics of Siberian Larch Species. In A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, & R. W. Wein (Eds.), *Permafrost ecosystems Siberian larch forests* (Vol. 209, pp. 41–58). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9693-8_3

Abaimov, A. P., Lesinski, J. A., Martinsson, O., & Milyutin, L. (1999). *Variability and Ecology of Siberian Larch Species* (Vol. 43). Swedish University of Agricultural Sciences, Department of Silviculture.

Abaimov, A. P., & Sofronov, M. A. (1996). The Main Trends of Post-Fire Succession in Near-Tundra Forests of Central Siberia. In J. G. Goldammer & V. Furyaev (Eds.), *Fire in Ecosystems of Boreal Eurasia* (1st ed., pp. 372–386). Springer, Dordrecht. https://doi.org/10.1007/978-94-015-8737-2_33

Abis, B., & Brovkin, V. (2019). Alternative tree-cover states of the boreal ecosystem: A conceptual model. *Global Ecology and Biogeography*, *28*(5), 612–627. https://doi.org/10.1111/geb.12880

Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J. E., Han, L., Muschitiello, F., Pedersen, M. W. M. W., Smittenberg, R. H., Yamoah, K. A. K. A., Slotte, T., & Wohlfarth, B. (2018). Archaeal community changes in Lateglacial lake sediments: Evidence from ancient DNA. *Quaternary Science Reviews*, *181*, 19–29. https://doi.org/10.1016/j.quascirev.2017.11.037

Alsos, I. G., Lammers, Y., Yoccoz, N. G., Jørgensen, T., Sjögren, P., Gielly, L., & Edwards, M. E. (2018). Plant DNA metabarcoding of lake sediments: How does it represent the contemporary vegetation. *PLoS ONE*, *13*(4), 1–23. https://doi.org/10.1371/journal.pone.0195403

Andreev, A. A., Tarasov, P. E., Klimanov, V. A., Melles, M., Lisitsyna, O. M., & Hubberten, H.-W. (2004). Vegetation and climate changes around the Lama Lake, Taymyr Peninsula, Russia during the Late Pleistocene and Holocene. *Quaternary International*, *122*(1), 69–84. https://doi.org/10.1016/j.quaint.2004.01.032

Araki, N. H. T., Khatab, I. A., Hemamali, K. K. G. U., Inomata, N., Wang, X. R., & Szmidt, A. E. (2008). Phylogeography of Larix sukaczewii Dyl. and Larix sibirica L. inferred from nucleotide variation of nuclear genes. *Tree Genetics and Genomes*, *4*(4), 611–623. https://doi.org/10.1007/s11295-008-0137-1

Betts, H. (1939). Larix occidentalis. *American Woods*, 1–7.

Bigelow, N. H., Brubaker, L. B., Edwards, M. E., Harrison, S. P., Prentice, I. C., Anderson, P. M., Andreev, A. A., Bartlein, P. J., Christensen, T. R., Cramer, W., Kaplan, J. O., Lozhkin, A. V., Matveyeva, N. V., Murray, D. F., McGuire, A. D., Razzhivin, V. Y., Ritchie, J. C., Smith, B., Walker, D. A., … Volkova, V. S. (2003). Climate change and Arctic ecosystems: 1. Vegetation changes north of 55°N between the last glacial maximum, mid-Holocene, and present. *Journal of Geophysical Research: Atmospheres*, *108*(19). https://doi.org/10.1029/2002jd002558

Binney, H. A., Willis, K. J., Edwards, M. E., Bhagwat, S. A., Anderson, P. M., Andreev, A. A., Blaauw, M., Damblon, F., Haesaerts, P., Kienast, F., Kremenetski, K. V., Krivonogov, S. K., Lozhkin, A. V., MacDonald, G. M., Novenko, E. Y., Oksanen, P., Sapelko, T. V., Väliranta, M., & Vazhenina,

L. (2009). The distribution of late-Quaternary woody taxa in northern Eurasia: evidence from a new macrofossil database. *Quaternary Science Reviews*, *28*(23–24), 2445–2464. https://doi.org/10.1016/j.quascirev.2009.04.016

Birks, H. J. B. (1981). The use of pollen analysis in the reconstruction of past climates: a review. In T. M. L. Wigley, M. J. Ingram, & G. Farmer (Eds.), *Climate and history: studies in past climates and their impact on man* (pp. 111–138). Cambridge University Press.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., … Lantuit, H. (2019). Permafrost is warming at a global scale. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-018-08240-4

Bondar, E. I., Putintseva, Y. A., Oreshkova, N. V., & Krutovsky, K. V. (2019). Siberian larch (Larix sibirica Ledeb.) chloroplast genome and development of polymorphic chloroplast markers. *BMC Bioinformatics*, *20*(S1), 38. https://doi.org/10.1186/s12859-018-2571-x

Brown, K. R., & Zobel, D. B. (1988). Seed dispersal, seedling emergence, and early survival of Larix larcina (DuRoi) K. Koch in the Tanana Valley, Alaska. *Canadian Journal of Forest Research*, *18*, 306–314.

Brubaker, L. B., Anderson, P. M., Edwards, M. E., & Lozhkin, A. V. (2005). Beringia as a glacial refugium for boreal trees and shrubs: new perspectives from mapped pollen data. *Journal of Biogeography*, *32*(5), 833–848. https://doi.org/10.1111/j.1365-2699.2004.01203.x

Busque, D., & Arseneault, D. (2005). Fire disturbance of larch woodlands in string fens in northern Québec. *Canadian Journal of Botany*, *83*(6), 599–609. https://doi.org/10.1139/b05-028

Carlson, C. E. (1998). Larix lyallii. *Enzyklopädie Der Holzgewächse: Handbuch Und Atlas Der Dendrologie*, 1–6. http://mfkp.org/INRMM/article/13745263

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., Rasmussen, M., Gravel, S., Guillén, S., Nekhrizov, G., Leshtakov, K., Dimitrova, D., Theodossiev, N., Pettener, D., Luiselli, D., Sandoval, K., Moreno-Estrada, A., Li, Y., Wang, J., … Bustamante, C. D. (2013). Pulling out the 1%: Whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics*, *93*(5), 852–864. http://www.sciencedirect.com/science/article/pii/S000292971300459X

Cheliak, W. M., Wang, J., & Pitel, J. A. (1988). Population structure and genic diversity in tamarack, *Larix laricina* (Du Roi) K. Koch. *Canadian Journal of Forest Research*, *18*(10), 1318–1324. https://doi.org/10.1139/x88-203

Clarke, C. L., Edwards, M. E., Gielly, L., Ehrich, D., Hughes, P. D. M., Morozova, L. M., Haflidason, H., Mangerud, J., Svendsen, J. I., & Alsos, I. G. (2019). Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. *Scientific Reports*, *9*(1), 19613. https://doi.org/10.1038/s41598-019-55989-9

Cooper, W. S. (1911). Reproduction by Layering Among Conifers. *Botanical Gazette*, *52*(5), 369–379. https://doi.org/10.1086/330666

Costa, L., Marques, A., Buddenhagen, C., Thomas, W. W., Huettel, B., Schubert, V., Dodsworth, S., Houben, A., Souza, G., & Pedrosa-Harand, A. (2021). Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA. *Annals of Botany*, 1–14. https://doi.org/10.1093/aob/mcab063

Da Ronch, F., Caudullo, G., Tinner, W., & de Rigo, D. (2016). Larix decidua and other larches in

Europe: distribution, habitat, usage and threats. In J. San-Miguel-Ayanz, D. de Rigo, G. Caudullo, T. Houston Durrant, & A. Mauri (Eds.), *European Atlas of Forest Tree Species.* (p. e01e492+). Publ. Off. EU, Luxembourg.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.-L., & Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, *110*(39), 15758–15763. https://doi.org/10.1073/pnas.1314445110

de Klerk, P., Teltewskoi, A., Theuerkauf, M., & Joosten, H. (2014). Vegetation patterns, pollen deposition and distribution of non-pollen palynomorphs in an ice-wedge polygon near Kytalyk (NE Siberia), with some remarks on Arctic pollen morphology. *Polar Biology*, *37*(10), 1393–1412. https://doi.org/10.1007/s00300-014-1529-3

Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., Cheng, T., Guo, J., & Zhou, S. (2015). ycf1, the most promising plastid DNA barcode of land plants. *Scientific Reports*, *5*, 8348. https://doi.org/10.1038/srep08348

Dylis, N. V. (1961). *Larch of Eastern Siberia and Far East*. Academy of Sciences of USSR.

Dylis, N. V. (1981). *Listvennitsa (Larch)*. Lesnaya Promyshlennost.

Enk, J., Devault, A., Widga, C., Saunders, J., Szpak, P., Southon, J., Rouillard, J.-M., Shapiro, B., Golding, G. B., Zazula, G., Froese, D., Fisher, D. C., MacPhee, R. D. E., & Poinar, H. (2016). Mammuthus population dynamics in late Pleistocene North America: Divergence, phylogeography, and introgression. *Frontiers in Ecology and Evolution*, *4*(April), 1–13. https://doi.org/10.3389/fevo.2016.00042

Epp, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A., & Herzschuh, U. (2018). Temporal and spatial patterns of mitochondrial haplotype and species distributions in Siberian larches inferred from ancient environmental DNA and modeling. *Scientific Reports*, *8*(1), 17436. https://doi.org/10.1038/s41598-018-35550-w

ESA. (2017). *Land Cover CCI Product User Guide Version 2. Tech. Rep.* maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Fisher, J. P., Estop-Aragonés, C., Thierry, A., Charman, D. J., Wolfe, S. A., Hartley, I. P., Murton, J. B., Williams, M., & Phoenix, G. K. (2016). The influence of vegetation and soil characteristics on active-layer thickness of permafrost soils in boreal forest. *Global Change Biology*, *22*(9), 3127–3140. https://doi.org/10.1111/gcb.13248

Furtwängler, A., Neukamm, J., Böhme, L., Reiter, E., Vollstedt, M., Arora, N., Singh, P., Cole, S. T., Knauf, S., Calvignac-Spencer, S., Krause-Kyora, B., Krause, J., Schuenemann, V. J., & Herbig, A. (2020). Comparison of target enrichment strategies for ancient pathogen DNA. *BioTechniques*, *69*(6), 455–460. https://doi.org/10.2144/btn-2020-0100

Gauthier, S., Bernier, P., Kuuluvainen, T., Shvidenko, A. Z., & Schepaschenko, D. G. (2015). Boreal forest health and global change. *Science*, *349*(6250), 819–822. https://doi.org/10.1126/science.aaa9092

Geburek, T. (2014). *Larix decidua*. In B. Stimm, A. Roloff, U. M. Lang, & H. Weisgerber (Eds.), *Enzyklopädie der Holzgewächse: Handbuch und Atlas der Dendrologie*. Wiley. https://doi.org/10.1002/9783527678518.ehg2002016

Ginolhac, A., Rasmussen, M., Gilbert, M. T. P., Willerslev, E., & Orlando, L. (2011). mapDamage: Testing for damage patterns in ancient DNA sequences. *Bioinformatics*, *27*(15), 2153–2155.

https://doi.org/10.1093/bioinformatics/btr347

Givnish, T. (2002). Adaptive significance of evergreen vs. deciduous leaves: solving the triple paradox. *Silva Fennica*, *36*(3), 703–743. https://doi.org/10.14214/sf.535

Gowan, E. J., Zhang, X., Khosravi, S., Rovere, A., Stocchi, P., Hughes, A. L. C., Gyllencreutz, R., Mangerud, J., Svendsen, J., & Lohmann, G. (2021). A new global ice sheet reconstruction for the past 80 000 years. *Nature Communications*, *12*(1), 1199. https://doi.org/10.1038/s41467-021-21469-w

Gower, S. T., & Richards, J. H. (1990). Larches: Deciduous Conifers in an Evergreen World. *BioScience*, *40*(11), 818–826. https://doi.org/10.2307/1311484

Graham, R. W., Belmecheri, S., Choy, K., Culleton, B. J., Davies, L. J., Froese, D., Heintzman, P. D., Hritz, C., Kapp, J. D., Newsom, L. A., Rawcliffe, R., Saulnier-Talbot, É., Shapiro, B., Wang, Y., Williams, J. W., & Wooller, M. J. (2016). Timing and causes of mid-Holocene mammoth extinction on St. Paul Island, Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(33), 9310–9314. https://doi.org/10.1073/pnas.1604903113

Guo, Q., Li, H., Qian, Z., Lu, J., & Zheng, W. (2021). Comparative study on the chloroplast genomes of five Larix species from the Qinghai-Tibet Plateau and the screening of candidate DNA markers. *Journal of Forestry Research*, *0123456789*. https://doi.org/10.1007/s11676-020-01279-4

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, *342*(6160), 850–853. https://doi.org/10.1126/science.1244693

Harbert, R. S. (2018). Algorithms and strategies in short-read shotgun metagenomic reconstruction of plant communities. *Applications in Plant Sciences*, *6*(3), e1034. https://doi.org/10.1002/aps3.1034

Heitkam, T., Schulte, L., Weber, B., Liedtke, S., Breitenbach, S., Kögler, A., Morgenstern, K., Brückner, M., Tröber, U., Wolf, H., Krabel, D., & Schmidt, T. (2021). Comparative Repeat Profiling of Two Closely Related Conifers (Larix decidua and Larix kaempferi) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs. *Frontiers in Genetics*, *12*(July 2021), 2021.03.21.436054. https://doi.org/10.3389/fgene.2021.683668

Hernandez-Rodriguez, J., Arandjelovic, M., Lester, J., de Filippo, C., Weihmann, A., Meyer, M., Angedakin, S., Casals, F., Navarro, A., Vigilant, L., Kühl, H. S., Langergraber, K., Boesch, C., Hughes, D., & Marques-Bonet, T. (2018). The impact of endogenous content, replicates and pooling on genome capture from faecal samples. *Molecular Ecology Resources*, *18*(2), 319–333. https://doi.org/10.1111/1755-0998.12728

Herzschuh, U. (2020). Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography*, *29*(2), 198–206. https://doi.org/10.1111/geb.13018

Herzschuh, U., Birks, H. J. B., Laepple, T., Andreev, A., Melles, M., & Brigham-Grette, J. (2016). Glacial legacies on interglacial vegetation at the Pliocene-Pleistocene transition in NE Asia. *Nature Communications*, *7*(1), 11967. https://doi.org/10.1038/ncomms11967

Hizume, M., Shibata, F., Matsumoto, A., Maruyama, Y., Hayashi, E., Kondo, T., Kondo, K., Zhang, S., & Hong, D. (2002). Tandem repeat DNA localizing on the proximal DAPI bands of chromosomes in \textit{Larix, Pinaceae}. *Genome*, *45*(4), 777–783.

Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M. N., Smith, S. W., Middle, C. M., Rodesch, M. J., Albert, T. J., Hannon, G. J., & McCombie, W. R. (2007). Genome-wide in situ exon capture for selective resequencing. *Nature Genetics*, *39*(12), 1522–1527. https://doi.org/10.1038/ng.2007.42

IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huan, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou (Eds.)). Cambridge University Press.

Johnston, W. F. (1990). *Larix laricina* (Du Roi) K. Koch tamarack. In R. M. Burns & B. H. Honkala (Eds.), *Silvics of North America* (pp. 141–151). Forest Service, United States Department of Agriculture.

Jørgensen, T., Haile, J., Möller, P., Andreev, A., Boessenkool, S., Rasmussen, M., Kienast, F., Coissac, E., Taberlet, P., Brochmann, C., Bigelow, N. H., Andersen, K., Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2012). A comparative study of ancient sedimentary DNA, pollen and macrofossils from permafrost sediments of northern Siberia reveals long-term vegetational stability. *Molecular Ecology*, *21*(8), 1989–2003. https://doi.org/10.1111/j.1365-294X.2011.05287.x

Kajimoto, T., Osawa, A., Usoltsev, V. A., & Abaimov, A. P. (2010). Biomass and Productivity of Siberian Larch Forest Ecosystems. In A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, & R. W. Wein (Eds.), *Permafrost ecosystems Siberian larch forests* (pp. 99–122). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9693-8_6

Kharuk, V. I., Ranson, K. J., & Dvinskaya, M. L. (2007). Evidence of Evergreen Conifers Invasion into Larch Dominated Forests During Recent Decades in Central Siberia. *Eurasian Journal of Forest Research*, *10*(2), 163–171. https://doi.org/10.1007/978-90-481-8641-9

Khatab, I. A., Ishiyama, H., Inomata, N., Wang, X.-R., & Szmidt, A. E. (2008). Phylogeography of Eurasian *Larix* species inferred from nucleotide variation in two nuclear genes. *Genes & Genetic Systems*, *83*(1), 55–66. https://doi.org/10.1266/ggs.83.55

Kistler, L., Montenegro, A., Smith, B. D., Gifford, J. A., Green, R. E., Newsom, L. A., & Shapiro, B. (2014). Transoceanic drift and the domestication of African bottle gourds in the Americas. *Proceedings of the National Academy of Sciences*, *111*(8), 2937–2941. https://doi.org/10.1073/pnas.1318678111

Kruse, S., Kolmogorov, A. I., Pestryakova, L. A., & Herzschuh, U. (2020). Long-lived larch clones may conserve adaptations that could restrict treeline migration in northern Siberia. *Ecology and Evolution*, *10*(18), 10017–10030. https://doi.org/10.1002/ece3.6660

Kruse, S., Stuenzi, S. M., Boike, J., Langer, M., Gloy, J., & Herzschuh, U. (2021). Novel coupled permafrost-forest model revealing the interplay between permafrost , vegetation , and climate across eastern Siberia. *Geoscientific Model Development Discussions*, *October*, 1–35.

Kullmann, L. (1998). Palaeoecological , Biogeographical and Palaeoclimatological Implications of Early Holocene Immigration of *Larix sibirica* Ledeb . into the Scandes Mountains, Sweden. *Global Ecology and Biogeography Letters*, *7*(3), 181–188. https://www.jstor.org/stable/2997373

Kuzmin, D. A., Feranchuk, S. I., Sharov, V. V., Cybin, A. N., Makolov, S. V., Putintseva, Y. A., Oreshkova, N. V., & Krutovsky, K. V. (2019). Stepwise large genome assembly approach: a

case of Siberian larch (Larix sibirica Ledeb). *BMC Bioinformatics*, *20*(S1), 37. https://doi.org/10.1186/s12859-018-2570-y

Lepage, B. A. (2003). The evolution, biogeography and palaeoecology of the Pinaceae based on fossil and extant representatives. *Acta Horticulturae*, *615*, 29–52.

LePage, B. A., & Basinger, J. F. (1995). The evolutionary history of the genus *Larix* (Pinaceae). *U.S. Dept. Agric., For Ser., Intermountain Res. Sta.*, *GTR-INT-31*(January 1995), 19–219.

Li, C., Postl, A. K., Böhmer, T., Cao, X., Dolman, A. M., & Herzschuh, U. (2022). Harmonized chronologies of a global late Quaternary pollen dataset (LegacyAge 1.0). *Earth System Science Data*, *14*(3), 1331–1343. https://doi.org/10.5194/essd-14-1331-2022

Liu, S., Stoof-Leichsenring, K. R., Kruse, S., Pestryakova, L. A., & Herzschuh, U. (2020). Holocene Vegetation and Plant Diversity Changes in the North-Eastern Siberian Treeline Region From Pollen and Sedimentary Ancient DNA. *Frontiers in Ecology and Evolution*, *8*(September), 1–17. https://doi.org/10.3389/fevo.2020.560243

Lozhkin, A., Anderson, P., Minyuk, P., Korzun, J., Brown, T., Pakhomov, A., Tsygankova, V., Burnatny, S., & Naumov, A. (2018). Implications for conifer glacial refugia and postglacial climatic variation in western Beringia from lake sediments of the Upper Indigirka basin. *Boreas*, *47*(3), 938–953. https://doi.org/10.1111/bor.12316

MacDonald, G. M., Kremenetski, K. V., & Beilman, D. W. (2008). Climate change and the northern Russian treeline zone. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1501), 2283–2299. https://doi.org/10.1098/rstb.2007.2200

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., & Turner, D. J. (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*(2), 111–118. https://doi.org/10.1038/nmeth.1419

Mamet, S. D., Brown, C. D., Trant, A. J., & Laroque, C. P. (2019). Shifting global *Larix* distributions: Northern expansion and southern retraction as species respond to changing climate. *Journal of Biogeography*, *46*(1), 30–44. https://doi.org/10.1111/jbi.13465

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS ONE*, *5*(11), e14004. https://doi.org/10.1371/journal.pone.0014004

Meucci, S., Schulte, L., Zimmermann, H. H., Stoof-Leichsenring, K. R., Epp, L., Bronken Eidesen, P., & Herzschuh, U. (2021). Holocene chloroplast genetic variation of shrubs (Alnus alnobetula, Betula nana, Salix sp.) at the siberian tundra-taiga ecotone inferred from modern chloroplast genome assembly and sedimentary ancient DNA analyses. *Ecology and Evolution*, *May 2020*, 1–21. https://doi.org/10.1002/ece3.7183

Meyer, H., Opel, T., Laepple, T., Dereviagin, A. Y., Hoffmann, K., & Werner, M. (2015). Long-term winter warming trend in the Siberian Arctic during the mid- to late Holocene. *Nature Geoscience*, *8*(2), 122–125. https://doi.org/10.1038/ngeo2349

Moore, C. R., Brooks, M. J., Goodyear, A. C., Ferguson, T. A., Perrotti, A. G., Mitra, S., Listecki, A. M., King, B. C., Mallinson, D. J., Lane, C. S., Kapp, J. D., West, A., Carlson, D. L., Wolbach, W. S., Them, T. R., Harris, M. S., & Pyne-O'Donnell, S. (2019). Sediment Cores from White Pond, South Carolina, contain a Platinum Anomaly, Pyrogenic Carbon Peak, and Coprophilous Spore Decline at 12.8 ka. *Scientific Reports*, *9*(1), 15121. https://doi.org/10.1038/s41598-019-51552-8

Müller, S., Tarasov, P. E., Andreev, A. A., Tütken, T., Gartz, S., & Diekmann, B. (2010). Late

Quaternary vegetation and environments in the Verkhoyansk Mountains region (NE Asia) reconstructed from a 50-kyr fossil pollen record from Lake Billyakh. *Quaternary Science Reviews*, *29*(17–18), 2071–2086. https://doi.org/10.1016/j.quascirev.2010.04.024

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., MacPhee, R., Froese, D., & Poinar, H. (2020). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quaternary Research*, *99*(September), 305–328. https://doi.org/10.1017/qua.2020.59

Napier, J. D., Fernandez, M. C., Lafontaine, G., & Hu, F. S. (2020). Ice-age persistence and genetic isolation of the disjunct distribution of larch in Alaska. *Ecology and Evolution*, *10*(3), 1692–1702. https://doi.org/10.1002/ece3.6031

Neale, D. B., & Wheeler, N. C. (2019). The Conifers: Genomes, Variation and Evolution. In *The Conifers: Genomes, Variation and Evolution*. https://doi.org/10.1007/978-3-319-46807-5

Neilson, R. P., Pitelka, L. F., Solomon, A. M., Nathan, R., Midgley, G. F., Fragoso, J. M. V., Lischke, H., & Thompson, K. (2005). Forecasting regional to global plant migration in response to climate change. *BioScience*, *55*(9), 749–759. https://doi.org/10.1641/0006-3568(2005)055[0749:FRTGPM]2.0.CO;2

Niemeyer, B., Klemm, J., Pestryakova, L. A., & Herzschuh, U. (2015). Relative pollen productivity estimates for common taxa of the northern Siberian Arctic. *Review of Palaeobotany and Palynology*, *221*, 71–82. https://doi.org/10.1016/j.revpalbo.2015.06.008

Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D. G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., … Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, *497*(7451), 579–584.

Obu, J., Westermann, S., Barboux, C., Bartsch, A., Delaloye, R., Grosse, G., Heim, B., Hugelius, G., Irrgang, A., Kääb, A. M., Kroisleitner, C., Matthes, H., Nitze, I., Pellet, C., Seifert, F. M., Strozzi, T., Wegmüller, U., Wieczorek, M., & Wiesmann, A. (2021a). ESA Permafrost Climate Change Initiative (Permafrost_cci): Permafrost active layer thickness for the Northern Hemisphere, v3.0. *NERC EDS Centre for Environmental Data Analysis*. https://doi.org/http://dx.doi.org/10.5285/29c4af5986ba4b9c8a3cfc33ca8d7c85

Obu, J., Westermann, S., Barboux, C., Bartsch, A., Delaloye, R., Grosse, G., Heim, B., Hugelius, G., Irrgang, A., Kääb, A. M., Kroisleitner, C., Matthes, H., Nitze, I., Pellet, C., Seifert, F. M., Strozzi, T., Wegmüller, U., Wieczorek, M., & Wiesmann, A. (2021b). ESA Permafrost Climate Change Initiative (Permafrost_cci): Permafrost extent for the Northern Hemisphere, v3.0. *NERC EDS Centre for Environmental Data Analysis*. https://doi.org/10.5285/6e2091cb0c8b4106921b63cd5357c97c

Osawa, A., & Zyryanova, O. A. (2010). Introduction. In *Permafrost ecosystems Siberian larch forests* (pp. 3–13).

Osawa, A., Zyryanova, O. A., Matsuura, Y., Kajimoto, T., & Wein, R. W. (Eds.). (2010a). *Permafrost Ecosystems* (Vol. 209). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9693-8

Osawa, A., Zyryanova, O. A., Matsuura, Y., Kajimoto, T., & Wein, R. W. (2010b). Permafrost ecosystems Siberian larch forests. In A. Osawa, O. A. Zyryanova, Y. Matsuura, T. Kajimoto, & R. W. Wein (Eds.), *Ecological Studies* (Vol. 209). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9693-8

Oswald, W. W., Anderson, P. M., Brown, T. A., Brubaker, L. B., Feng, S. H., Lozhkin, A. V., Tinner, W., & Kaltenrieder, P. (2005). Effects of sample mass and macrofossil type on radiocarbon dating of arctic and boreal lake sediments. *Holocene*, *15*(5), 758–767. https://doi.org/10.1191/0959683605hl849rr

Paijmans, J. L. A., Fickel, J., Courtiol, A., Hofreiter, M., & Förster, D. W. (2016). Impact of enrichment conditions on cross-species capture of fresh and degraded DNA. *Molecular Ecology Resources*, *16*(1), 42–55. https://doi.org/10.1111/1755-0998.12420

Pansu, J., Giguet-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., Arnaud, F., Poulenard, J., Taberlet, P., & Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, *24*(7), 1485–1498. https://doi.org/10.1111/mec.13136

Pâques, L. E., Foffová, E., Heinze, B., Lelu-Walter, M.-A., Liesebach, M., & Philippe, G. (2013). Larches (Larix sp.). In L. E. Pâques (Ed.), *Forest Tree Breeding in Europe* (Vol. 25, Issue 1, pp. 13–122). Springer, Dortrecht. https://doi.org/10.1007/978-94-007-6146-9_2

Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M., & Eckert, A. J. (2018). RADseq approaches and applications for forest tree genetics. *Tree Genetics & Genomes*, *14*(3), 39. https://doi.org/10.1007/s11295-018-1251-3

Parducci, L., Bennett, K. D., Ficetola, G. F., Alsos, I. G., Suyama, Y., Wood, J. R., & Pedersen, M. W. (2017). Ancient plant DNA in lake sediments. *New Phytologist*, *214*(3), 924–942. https://doi.org/10.1111/nph.14470

Patschke, W. (1913). Über die extratropischen ostasiatischen Coniferen und ihre Bedeutung für die pflanzengeographische Gliederung Ostasiens. *Botanische Jahrbücher Für Systematik, Pflanzengeschichte Und Pflanzengeographie*, *48*, 622–776.

Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., Mendoza, M. L. Z., Beaudoin, A. B., Zutter, C., Larsen, N. K., Potter, B. A., Nielsen, R., Rainville, R. A., Orlando, L., Meltzer, D. J., Kjær, K. H., & Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, *537*(7618), 45–49. https://doi.org/10.1038/nature

Petit, R. J., Aguinagalde, I., De Beaulieu, J. L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martín, J. P., Rendell, S., & Vendramin, G. G. (2003). Glacial refugia: Hotspots but not melting pots of genetic diversity. *Science*, *300*(5625), 1563–1565. https://doi.org/10.1126/science.1083264

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils*, *45*(3), 219–235. https://doi.org/10.1007/s00374-008-0345-8

Plohl, M., Meštrović, N., Mravinac, B., Meštrovic, N., & Mravinac, B. (2012). Satellite DNA evolution. *Repetitive DNA*, *7*, 126–152. https://doi.org/10.1159/000337122

Polezhaeva, M. A., Lascoux, M., & Semerikov, V. L. (2010). Cytoplasmic DNA variation and biogeography of *Larix* Mill, in Northeast Asia. *Molecular Ecology*, *19*(6), 1239–1252. https://doi.org/10.1111/j.1365-294X.2010.04552.x

Putintseva, Y. A., Bondar, E. I., Simonov, E. P., Sharov, V. V., Oreshkova, N. V., Kuzmin, D. A., Konstantinov, Y. M., Shmakov, V. N., Belkov, V. I., Sadovsky, M. G., Keech, O., & Krutovsky, K. V. (2020). Siberian larch (Larix sibirica Ledeb.) mitochondrial genome assembled using

both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics*, *21*(1), 654. https://doi.org/10.1186/s12864-020-07061-4

Scheffer, M., Hirota, M., Holmgren, M., Van Nes, E. H., & Chapin, F. S. (2012). Thresholds for boreal biome transitions. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(52), 21384–21389. https://doi.org/10.1073/pnas.1219844110

Schmickl, R., Liston, A., Zeisek, V., Oberlander, K., Weitemier, K., Straub, S. C. K., Cronn, R. C., Dreyer, L. L., & Suda, J. (2016). Phylogenetic marker development for target enrichment from transcriptome and genome skim data: the pipeline and its application in southern African Oxalis (Oxalidaceae). *Molecular Ecology Resources*, *16*(5), 1124–1135. https://doi.org/10.1111/1755-0998.12487

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017). HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods in Ecology and Evolution*, *8*(10), 1374–1388. https://doi.org/10.1111/2041-210X.12785

Schulze, E. D., Wirth, C., Mollicone, D., Von Lüpke, N., Ziegler, W., Achard, F., Mund, M., Prokushkin, A., & Scherbina, S. (2012). Factors promoting larch dominance in central Siberia: Fire versus growth performance and implications for carbon dynamics at the boundary of evergreen and deciduous conifers. *Biogeosciences*, *9*(4), 1405–1421. https://doi.org/10.5194/bg-9-1405-2012

Seddon, A. W. R., Macias-Fauria, M., Long, P. R., Benz, D., & Willis, K. J. (2016). Sensitivity of global terrestrial ecosystems to climate variability. *Nature*, *531*(7593), 229–232. https://doi.org/10.1038/nature16986

Semerikov, V. L., & Lascoux, M. (1999). Genetic relationship among Eurasian and American *Larix* species based on allozymes. *Heredity*, *83*(1), 62–70. https://doi.org/10.1038/sj.hdy.6885310

Semerikov, V. L., & Lascoux, M. (2003). Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae) species. *American Journal of Botany*, *90*(8), 1113–1123. https://doi.org/10.3732/ajb.90.8.1113

Semerikov, V. L., Semerikova, S. A., Polezhaeva, M. A., Kosintsev, P. A., & Lascoux, M. (2013). Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. *Molecular Ecology*, *22*(19), 4958–4971. https://doi.org/10.1111/mec.12433

Semerikov, V. L., Zhang, H., Sun, M., & Lascoux, M. (2003). Conflicting phylogenies of *Larix* (Pinaceae) based on cytoplasmic and nuclear DNA. *Molecular Phylogenetics and Evolution*, *27*(2), 173–184.

Shirazi, S., Meyer, R., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Authorea Preprints*, 1–22. https://doi.org/10.22541/au.159309876.62184178/v2

Shuman, J. K., Shugart, H. H., & O'Halloran, T. L. (2011). Sensitivity of Siberian larch forests to climate change. *Global Change Biology*, *17*(7), 2370–2384. https://doi.org/10.1111/j.1365-2486.2011.02417.x

Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gušić, I., … Meyer, M. (2017). Neandertal and

Denisovan DNA from Pleistocene sediments. *Science*, *356*(6338), 605–608. https://doi.org/10.1126/science.aam9695

Sonstebo, J. H., Gielly, L., Brysting, A. K., Elven, R., Edwards, M., Haile, J., Willerslev, E., Coissac, E., Rioux, D., Sannier, J., Taberlet, P., & Brochmann, C. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, *10*(6), 1009–1018. https://doi.org/10.1111/j.1755-0998.2010.02855.x

Sturm, M., & Lotter, A. F. (1995). Lake sediments as environmental archives. *EAWAG News*, *38 E*(January 1995), 6–9.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T., Corthier, G., Brochmann, C., & Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3), e14–e14. https://doi.org/10.1093/nar/gkl938

Taggart, R. E., & Cross, A. T. (2009). Global greenhouse to icehouse and back again: The origin and future of the Boreal Forest biome. *Global and Planetary Change*, *65*(3–4), 115–121. https://doi.org/10.1016/j.gloplacha.2008.10.014

Tarasov, P., Müller, S., Andreev, A., Werner, K., & Diekmann, B. (2009). Younger Dryas Larix in eastern Siberia: A migrant or survivor? *PAGES News*, *17*(3), 122–123. https://doi.org/10.22498/pages.17.3.122

Tchebakova, N. M., Rehfeldt, G. E., & Parfenova, E. I. (2006). Impacts of Climate Change on the Distribution of Larix Spp. and Pinus Sylvestris and Their Climatypes in Siberia. *Mitigation and Adaptation Strategies for Global Change*, *11*(4), 861–882. https://doi.org/10.1007/s11027-005-9019-0

Tollefsrud, M. M., Latałowa, M., van der Knaap, W. O., Brochmann, C., & Sperisen, C. (2015). Late Quaternary history of North Eurasian Norway spruce (Picea abies) and Siberian spruce (Picea obovata) inferred from macrofossils, pollen and cytoplasmic DNA variation. *Journal of Biogeography*, *42*(8), 1431–1442. https://doi.org/10.1111/jbi.12484

Tsetsos, F., Drineas, P., & Paschou, P. (2018). Genetics and population analysis. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*, 363–378. https://doi.org/10.1016/B978-0-12-809633-8.20114-3

Uemura, S., Kanda, F., Isaev, A., & Tsuji, T. (1997). Forest structure and succession in southeastern Siberia. *Vegetation Science*, *14*(2), 119–127. https://doi.org/10.15031/vegsci.14.119

Van Geel, B. (2002). Non-Pollen Palynomorphs. In J. P. Smol, H. J. B. Birks, W. M. Last, R. S. Bradley, & Alverson K. (Eds.), *Tracking Environmental Change Using Lake Sediments. Developments in Paleoenvironmental Research* (Vol. 3, pp. 99–119). Springer, Dortrecht. https://doi.org/10.1007/0-306-47668-1_6

Vitales, D., Garcia, S., & Dodsworth, S. (2020). Reconstructing phylogenetic relationships based on repeat sequence similarities. *Molecular Phylogenetics and Evolution*, *147*(February), 106766. https://doi.org/10.1016/j.ympev.2020.106766

Wagner, S., Litt, T., Sánchez-Goñi, M.-F. F., & Petit, R. J. (2015). History of *Larix decidua* Mill. (European larch) since 130 ka. *Quaternary Science Reviews*, *124*, 224–247. https://doi.org/10.1016/j.quascirev.2015.07.002

Wang, Y., Heintzman, P. D., Newsom, L., Bigelow, N. H., Wooller, M. J., Shapiro, B., & Williams, J. W. (2017). The southern coastal Beringian land bridge: cryptic refugium or pseudorefugium for woody plants during the Last Glacial Maximum? *Journal of Biogeography*, *44*(7), 1559–

1571. https://doi.org/10.1111/jbi.13010

Warren, E., de Lafontaine, G., Gérardi, S., Senneville, S., Beaulieu, J., Perron, M., Jaramillo-Correa, J. P., & Bousquet, J. (2016). Joint inferences from cytoplasmic DNA and fossil data provide evidence for glacial vicariance and contrasted post-glacial dynamics in tamarack, a transcontinental conifer. *Journal of Biogeography*, *43*(6), 1227–1241. https://doi.org/10.1111/jbi.12675

Wei, X.-X., & Wang, X.-Q. (2003). Phylogenetic split of Larix: Evidence from paternally inherited cpDNA trnT-trnF region. *Plant Systematics and Evolution*, *239*(1–2), 67–77.

Weitemier, K., Straub, S. C. K., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., & Liston, A. (2014). Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant Phylogenomics. *Applications in Plant Sciences*, *2*(9), 1400042. https://doi.org/10.3732/apps.1400042

White, L. C., Fontsere, C., Lizano, E., Hughes, D. A., Angedakin, S., Arandjelovic, M., Granjon, A. C., Hans, J. B., Lester, J. D., Rabanus-Wallace, M. T., Rowney, C., Städele, V., Marques-Bonet, T., Langergraber, K. E., & Vigilant, L. (2019). A roadmap for high-throughput sequencing studies of wild animal populations using noninvasive samples and hybridization capture. *Molecular Ecology Resources*, *19*(3), 609–622. https://doi.org/10.1111/1755-0998.12993

Wieczorek, M., Kruse, S., Epp, L. S., Kolmogorov, A., Nikolaev, A. N., Heinrich, I., Jeltsch, F., Pestryakova, L. A., Zibulski, R., & Herzschuh, U. (2017). Dissimilar responses of larch stands in northern Siberia to increasing temperatures-a field and simulation based study. *Ecology*, *98*(9), 2343–2355. https://doi.org/10.1002/ecy.1887

Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., Lorenzen, E. D., Vestergård, M., Gussarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L. S., Pearman, P. B., Cheddadi, R., Murray, D., Bråthen, K. A., Yoccoz, N., … Taberlet, P. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, *506*(7486), 47–51. https://doi.org/10.1038/nature12921

Wirth, C., Schulze, E. D., Lühker, B., Grigoriev, S., Siry, M., Hardes, G., Ziegler, W., Backor, M., Bauer, G., & Vygodskaya, N. N. (2002). Fire and site type effects on the long-term carbon and nitrogen balance in pristine Siberian Scots pine forests. *Plant and Soil*, *242*(1), 41–63. https://doi.org/10.1023/A:1020813505203

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, *20*(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Zhang, N., Yasunari, T., & Ohta, T. (2011). Dynamics of the larch taiga–permafrost coupled system in Siberia under climate change. *Environmental Research Letters*, *6*(2), 024003. https://doi.org/10.1088/1748-9326/6/2/024003

Zimmermann, H. H., Harms, L., Epp, L. S., Mewes, N., Bernhardt, N., Kruse, S., Stoof-Leichsenring, K. R., Pestryakova, L. A., Wieczorek, M., Trense, D., & Herzschuh, U. (2019). Chloroplast and mitochondrial genetic variation of larches at the Siberian tundra-taiga ecotone revealed by de novo assembly. *PLOS ONE*, *14*(7), e0216966. https://doi.org/10.1371/journal.pone.0216966

Zimmermann, H. H., Raschke, E., Epp, L., Stoof-Leichsenring, K. R., Schirrmeister, L., Schwamborn, G., & Herzschuh, U. (2017). The History of Tree and Shrub Taxa on Bol'shoy Lyakhovsky Island (New Siberian Archipelago) since the Last Interglacial Uncovered by Sedimentary Ancient DNA and Pollen Data. *Genes*, *8*(10), 273. https://doi.org/10.3390/genes8100273

7 Appendix

7.1 Appendix to manuscript I

Supplementary information for

Schulte, L., Li, C., Lisovski, S., Herzschuh, U., Forest-permafrost feedback loops and glacial refugia help explain the unequal distribution of larch across continents, submitted to *Journal of Biogeography*

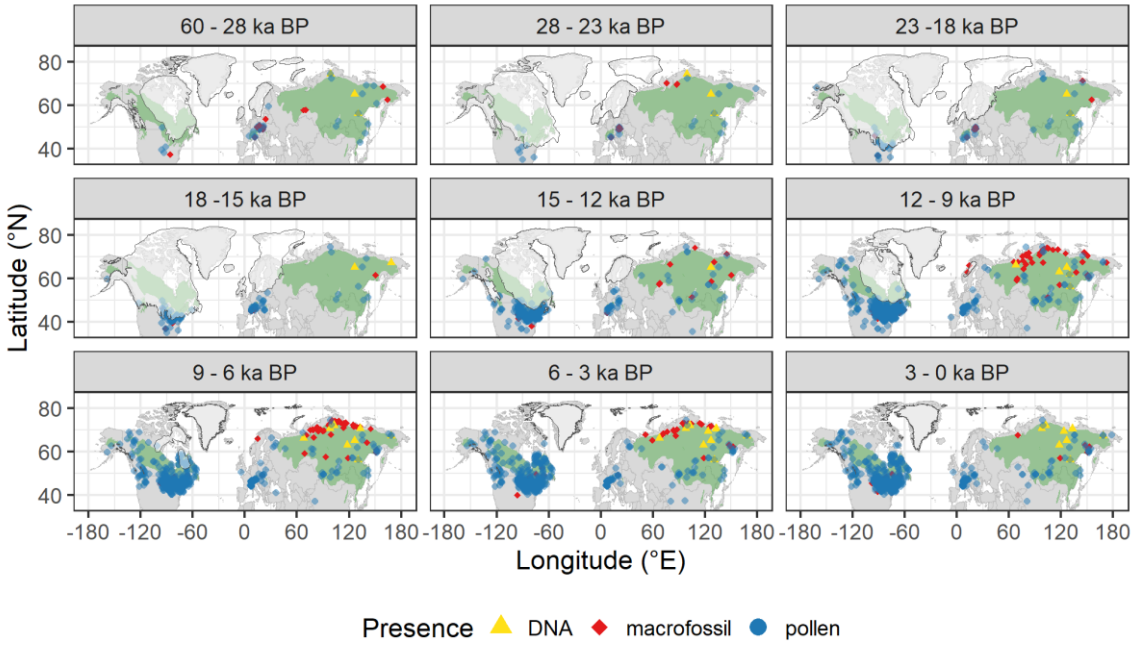

Fig. S1 Past occurences of *Larix*. In green, modern distribution of Larix (Mamet et al., 2019), in white, reconstructed ice margins (Gowan et al., 2021). Colors and shapes indicate the present of Larix in a dataset of the respective proxy in at least one sample of the time period indicated.

Gowan, E. J., Zhang, X., Khosravi, S., Rovere, A., Stocchi, P., Hughes, A. L. C., Gyllencreutz, R., Mangerud, J., Svendsen, J., & Lohmann, G. (2021). A new global ice sheet reconstruction for the past 80 000 years. *Nature Communications*, *12*(1), 1199. https://doi.org/10.1038/s41467-021-21469-w

Mamet, S. D., Brown, C. D., Trant, A. J., & Laroque, C. P. (2019). Shifting global Larix distributions: Northern expansion and southern retraction as species respond to changing climate. *Journal of Biogeography*, *46*(1), 30–44. https://doi.org/10.1111/jbi.13465

7.2 Appendix to manuscript II

Supplementary information for:

Schulte, L., Bernhardt, N., Stoof-Leichsenring, K., Zimmermann, H.H., Pestryakova, L.A., Epp, L.S. and Herzschuh, U. (2021), Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. Mol Ecol Resour, 21: 801-815. https://doi.org/10.1111/1755-0998.13311

Table S1 Sequencing numbers for shotgun sequencing and hybridization capture sequencing.

Shotgun	**6700 cal-BP**	**5400 cal-BP**	**1900 cal-BP**	**60 cal-BP**	**Ext. blank**	**Lib. blank**
Raw reads	145,972,964	112,220,286	78,650,028	87,627,075	10,370,556	14,264,796
QC reads	74,395,179	82,578,790	50,280,976	58,194,020	23,014	15,692
Merged reads	59,402,747	70,277,127	40,827,070	48,646,742	19,117	10,662
Capture	**6700 cal-BP**	**5400 cal-BP**	**1900 cal-BP**	**60 cal-BP**	**Ext. blank**	**Lib. blank**
Raw reads	50,743,540	47,562,965	42,857,744	50,564,822	4,309,730	5,796,937
QC reads	36,512,567	30,066,323	27,832,438	32,166,404	20,696	7,322
Merged reads	33,115,641	27,802,780	25,304,175	29,301,513	19,898	6,788

Cal-BP = calibrated years before present (present = 1950); Ext. = extraction; Lib. = library; QC = quality control passed

Table S2 Classification against nt database with kraken2 (--confidence 0.8). Counts from merged and unmerged paired-end reads

Shotgun	**6700 cal-BP**	**5400 cal-BP**	**1900 cal-BP**	**60 cal-BP**	**Ext. blank**	**Lib. blank**
Unclassified	74,143,837	82,423,133	50,093,108	57,980,345	20,172	12,753
Root	251,342	155,657	187,868	213,675	2,842	2,939
Archaea	4,511	2,057	2,087	1,099	-	-
Bacteria	182,235	101,074	110,087	112,852	375	322
Viruses	98	938	60	107	3	2
Eukaryota	29,879	31,885	55,021	72,380	1,838	1,901
Fungi	494	1,273	898	961	9	10
Metazoa	2,011	3,530	3,004	17,940	1,797	1,874
Viridiplantae	19,115	17,270	46,359	39,477	30	16
Streptophyta	18,185	16,435	45,756	38,819	30	16
Larix	1,196	495	1,150	45	-	-
Capture	**6700 cal-BP**	**5400 cal-BP**	**1900 cal-BP**	**60 cal-BP**	**Ext. blank**	**Lib. blank**
Unclassified	25,071,453	23,956,488	17,858,027	24,178,726	8,885	3,156

Root	11,441,113	6,109,835	9,974,411	7,987,678	11,013	3,632
Archaea	41,891	16,271	21,801	11,507	-	-
Bacteria	3,718,173	2,142,634	2,558,563	2,616,197	1,641	772
Viruses	730	2,821	266	463	-	-
Eukaryota	4,654,826	2,217,357	5,587,144	3,373,420	5,112	896
Fungi	1,270	4,101	3,213	9,220	38	24
Metazoa	3,993	6,600	4,561	34,006	795	868
Viridiplantae	4,527,757	2,111,186	5,440,761	3,070,656	4,274	4
Streptophyta	4,406,508	2,045,301	5,365,130	2,986,741	4,274	4
Larix	1,385,860	429,979	1,286,646	57,707	4	1

Cal-BP = calibrated years before present (present = 1950); Ext. = extraction; Lib. = library

Table S3 Classification against chloroplast database with kraken2 (--confidence default). Counts from merged and unmerged paired-end reads

Shotgun	**6700 cal-BP**	**5400 cal-BP**	**1900 cal-BP**	**60 cal-BP**	**Ext. blank**	**Lib. blank**
Unclassified	74,271,855	82,495,622	50,165,405	58,082,107	22,966	15,673
Viridiplantae	123,338	83,202	115,614	111,946	48	20
Streptophyta	77,381	55,529	86,353	77,549	36	12
Larix	1,581	541	1,472	78	-	-
Capture	**6700 cal-BP**	**5400 cal-BP**	**1900 cal-BP**	**60 cal-BP**	**Ext. blank**	**Lib. blank**
Unclassified	21,892,038	22,137,637	14,956,260	21,355,789	8,144	4,579
Viridiplantae	14,620,529	7,928,686	12,876,178	10,810,640	12,552	2,743
Streptophyta	10,166,027	5,016,616	9,908,283	7,214,903	10,220	1,643
Larix	1,905,593	562,216	1,697,052	73,243	7	1

Cal-BP = calibrated years before present (present = 1950); Ext. = extraction; Lib. = library

Table S4 Number of *Larix*-classified reads mapped to *L. gmelinii* chloroplast genome. Mapped reads with PCR duplicates removed

	Capture	**Shotgun**	**Enrichment**
Merged reads			
6700 cal-BP	13906	1393	9.98x
5400 cal-BP	4125	478	8.63x
1900 cal-BP	3918	1149	3.41x
60 cal-BP	156	40	3.90x
Unmerged reads			
6700 cal-BP	10503	111	94.62x
5400 cal-BP	2175	29	75.00x
1900 cal-BP	4357	138	31.57x
60 cal-BP	201	7	28.71x
Sum of merged & unmerged			
6700 cal-BP	24409	1504	16.23x
5400 cal-BP	6300	507	12.43x
1900 cal-BP	8275	1287	6.43x
60 cal-BP	357	47	7.60x

Table S5 Breadth of coverage at 1-fold and 10-fold coverage of the *Larix gmelinii* choroplast genome. Alignment of the capture dataset as subset of only reads classified as *Larix* and the complete dataset, PCR duplicates removed.

	Capture *Larix* reads	**Capture full sample**	**Capture *Larix* reads**	**Capture full sample**
Fold coverage	1x	1x	10x	10x
Merged reads				
6700 cal-BP	89.3897	99.9943	67.2888	99.0978
5400 cal-BP	79.1905	99.1949	4.94583	19.0842
1900 cal-BP	81.3171	99.11	5.72079	18.7033
60 cal-BP	10.7254	29.7395	0	5.93614
Unmerged reads				
6700 cal-BP	90.0529	99.9356	26.2995	59.4406
5400 cal-BP	62.0363	87.108	0.56694	10.0026
1900 cal-BP	79.8308	97.1139	2.77923	18.0858
60 cal-BP	9.13466	37.5184	0	10.6984
Merged & unmerged				
6700 cal-BP	91.416	99.9943	76.4186	99.9054
5400 cal-BP	80.2795	99.4339	14.388	40.6435
1900 cal-BP	85.1013	99.4828	25.6877	55.8864
60 cal-BP	14.171	42.7823	0	12.511

Table S6 Number of reads classified as *L. gemlinii* or *L. sibirica* in the alignment of *Larix*-classified reads against the *L. gmelinii* and *L. sibirica* chloroplast genome respectively. Number from the 294 variable sites, considering only bases with a base quality above 30.

		L. gmelinii as reference			*L. sibirica* as reference		
	Sample	*L. gmelinii*	*L. sibirica*	Other	*L. gmelinii*	*L. sibirica*	other
No. of sites with at least one read coverage for each group	60 cal-BP	19	1	1	62	28	9
	1900 cal-BP	283	39	9	276	87	18
	5400 cal-BP	220	17	3	201	50	8
	6800 cal -BP	289	150	23	280	194	29
Percent of sites with at least one read coverage for each group	60 cal-BP	6.5	0.3	0.3	21.1	9.5	3.1
	1900 cal-BP	96.3	13.3	3.1	93.9	29.6	6.1
	5400 cal-BP	74.8	5.8	1.0	68.4	17.0	2.7
	6800 cal -BP	98.3	51.0	7.8	95.2	66.0	9.9
Total number of reads over all positions for each group	60 cal-BP	35	1	1	164	113	17
	1900 cal-BP	2671	72	15	2275	462	46
	5400 cal-BP	1621	27	4	1034	189	10
	6800 cal -BP	7686	412	30	6786	1703	91
Percent of total reads classified to each group	60 cal-BP	94.6	2.7	2.7	55.8	38.4	5.8
	1900 cal-BP	96.9	2.6	0.5	81.7	16.6	1.7
	5400 cal-BP	98.1	1.6	0.2	83.9	15.3	0.8
	6800 cal -BP	94.6	5.1	0.4	79.1	19.8	1.1
Median of total number of reads for each group	60 cal-BP	0	0	0	0	0	0
	1900 cal-BP	8	0	0	7	0	0
	5400 cal-BP	5	0	0	3	0	0
	6800 cal -BP	25	1	0	23	1	0
Mean of total number of reads for each group	60 cal-BP	0.1	0.0	0.0	0.6	0.4	0.1
	1900 cal-BP	9.1	0.2	0.1	7.7	1.6	0.2
	5400 cal-BP	5.5	0.1	0.0	3.5	0.6	0.0
	6800 cal -BP	26.1	1.4	0.1	23.1	5.8	0.3
Standard deviation of total number of reads for each group	60 cal-BP	0.50	0.06	0.06	1.74	1.77	0.36
	1900 cal-BP	6.73	0.73	0.32	4.76	3.64	1.04
	5400 cal-BP	5.25	0.43	0.14	3.40	1.74	0.23
	6800 cal -BP	12.44	2.31	0.45	11.16	12.01	2.03

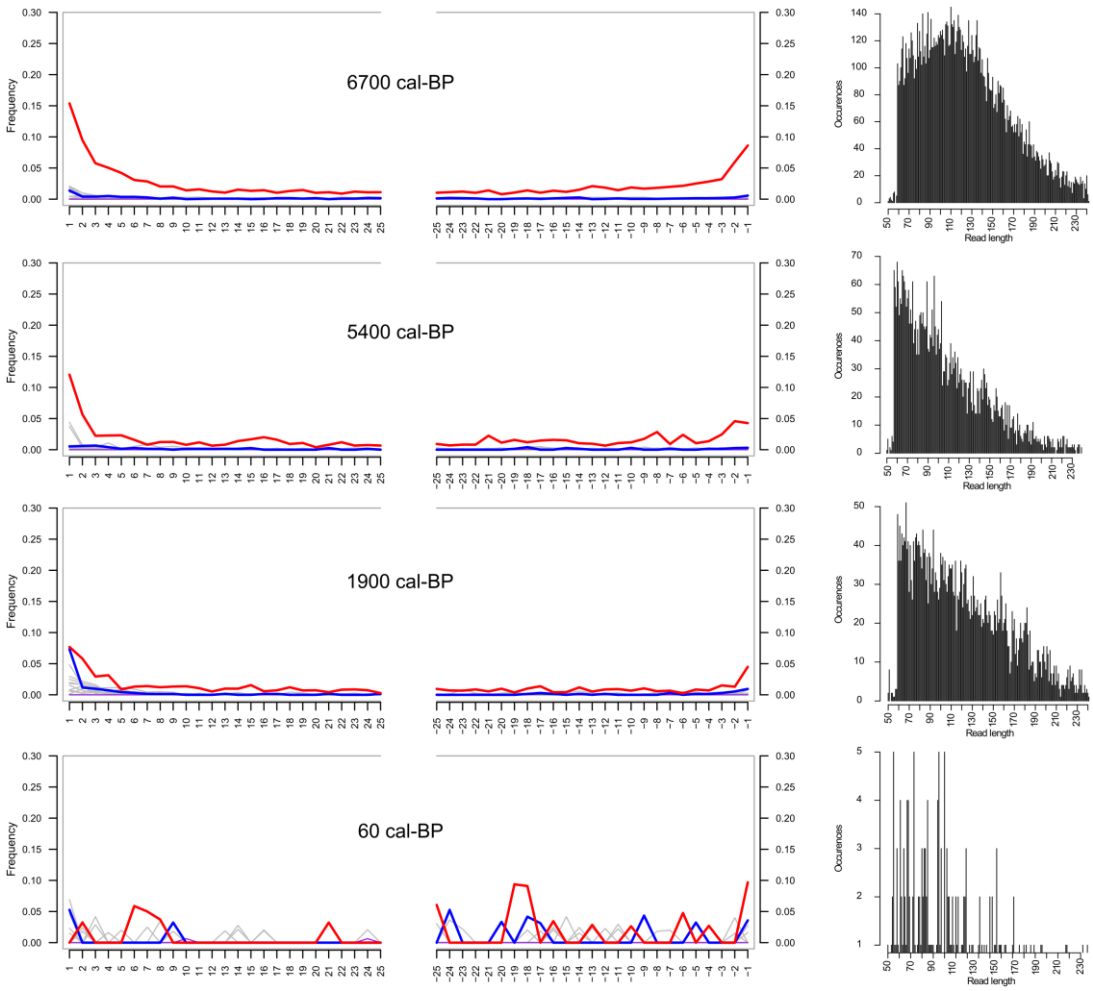

Figure S1 MapDamage plots for overlapping merged reads of the hybridization capture dataset aligned against the *Larix* chloroplast genome: Left: Misincorporation plots, red: C to T substitutions, blue: G to A substitutions, grey: all other substitutions. Right: Read length distributions.

7.3 Appendix to manuscript III

Supplemental information for

Schulte, L., Meucci, S., Stoof-Leichsenring, K, Heitkam, T., Schmidt N., von Hippel, B., Andreev, A. A., Diekmann, B., Biskaborn, B. B., Wagner, B., Melles, M., Pestryakova, L. A., Alsos, I. G., Clarke, C., Krutovsky, K. V., & Herzschuh, U., Dynamics of larch species ranges in Siberia since the Last Glacial captured from sedimentary ancient DNA, *communications biology*. (under review)

7.3.1 Material and Methods

Hybridization capture targeting the chloroplast genome and nuclear genes of Larix

Sample material and chronostratigraphy

Sample material was obtained from 8 lake sediment cores from across Siberia, for location and published age model see Supplementary Table 4.

Radiocarbon dating and age modelling of sediment record PG2361, Lake Satagay

We used Bacon in R and the IntCal20 calibration curve to model the age-depth relationship based on 11 radiocarbon dates from the MICADAS laboratory (Bremerhaven, Supplementary Table 5). Two bulk samples 5281.1.1 and 5193.1.1 from 0.5 and 341.75 cm core depth, respectively, were identified as outliers because they contained too high amounts of old carbon compared to a linear relationship with adjacent samples. We used linear regression of C^{14} ages (n=10, excluding the bottom sample 5188.1.1) to find the intersection with the age axis at 0 cm depth and used this value as an old carbon reservoir effect in the lake system (delta.R=1955.2 yrs). The 0.5 cm bulk sample supports the assumption of an old carbon effect in the lake, although it is ca. 570 yrs too old compared to the inferred reservoir value. We used the maximum dating error from the C^{14} dataset as the standard deviation for the reservoir estimate (delta.STD=52 yrs) applied to 11 samples. A dated water plant from a surface sample revealed a roughly modern age (48 +-18 C^{14} yrs at 0-1 cm core depth). Outliers were not used for age-depth modelling but are indicated by red crosses (Fig. S1).

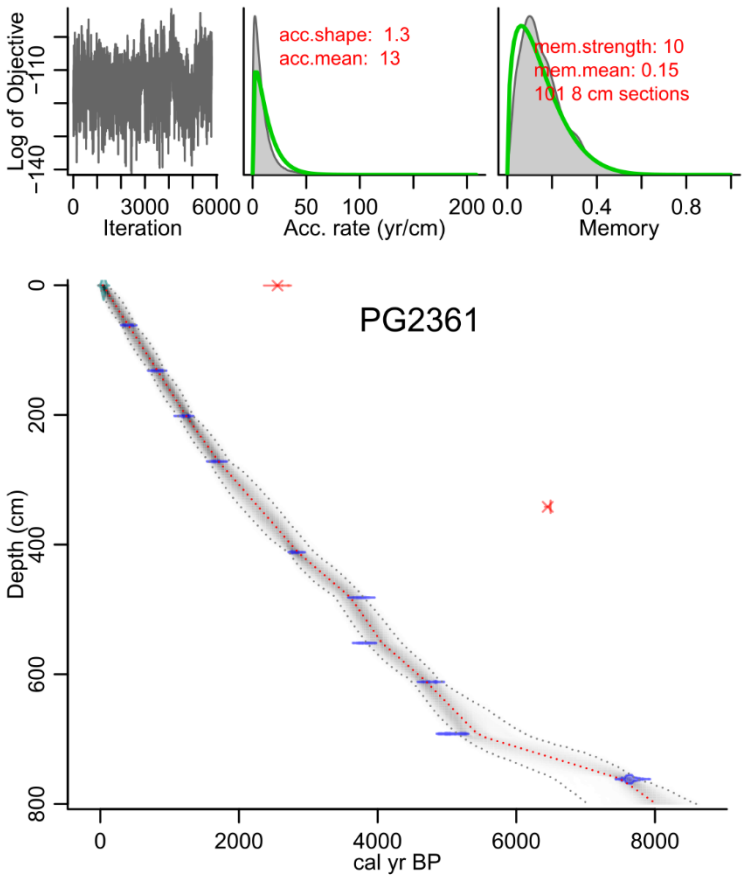

Fig. S1: Age-depth model of PG2361 (Lake Satagay). Performed using R package 'Bacon'[1]

DNA extraction and library preparation

Subsampling of all cores, with the exception of the core from Lake Bolshoye Shchuchye, was performed as described in Epp et al.[2] in the climate chamber of the Helmholtz Centre Potsdam – German Research Centre for Geosciences (GFZ) with the following adjustments. For each sample, the top of the sediment core half was removed twice with a sterile scalpel blade. A one-centimeter thick slice of the sediment was cut out, put on a cleaned sampling plate, and sides previously touching the coring tube were cut with cleaned knives. Sampling equipment was cleaned with 5% sodium hydroxide (VWR, Germany) and DNA-ExitusPlus™ (PanReac AppliChem, Germany). DNA extraction was done in a dedicated ancient DNA laboratory at Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, as described in Schulte et al.[3]. Subsampling from Lake Bolshoye Shchuchye was performed in a laminar flow hood in a clean laboratory at the Centre for Geobiology and Microbiology, University of Bergen, and DNA extraction was performed in the dedicated ancient DNA facility at Tromsø University Museum as described in Clarke et al.[4].

For the preparation of sequencing libraries, a single stranded DNA library protocol developed for ancient DNA was followed as published by Gansauge et al.[5,6] with minor adjustments as described in Schulte et al.[3]. With each batch of 7 libraries, 1 blank (control) library with only diethylpyrocarbonate (DEPC) purified water as input instead of DNA was included. Additional

libraries were constructed for the extraction blanks. As samples from different extraction batches were included in the study, often several extraction blanks were pooled together prior to building the library (see Supplementary Table 6). Libraries were quantified with qPCR as described in Gansauge and Meyer[5,6] and Schulte et al.[3].

Bait construction

Baits targeting the chloroplast genome of *Larix*

Baits targeting the chloroplast genome of *Larix* were produced using long-range PCR on DNA extracted from a *Larix gmelinii* individual (collected in the Botanical Gardens of the University of Potsdam, Accession CC-0POTSD-3867) as described in Schulte et al.[3]. In contrast to Schulte et al.[3], two of the 18 primer pairs covering the complete chloroplast genome[7] were left out as they cover the ribosomal RNA genes and we found that these very conserved regions cannot be unequivocally assigned to *Larix.* The omitted primer pairs are primer pair 8 (*Larix*_49890_F/*Larix*_57954_R) and primer pair 9 (*Larix*_57611_F/Lg_03_R).

Baits targeting a set of nuclear genes of *Larix*

Primer pairs for candidate adaptive genes found in the nuclear genome of *Larix* were selected from the literature. Specific primers for 59 candidate adaptive genes were originally designed for *Pinus taeda* by Eckert et al.[8], and tested successfully on *Larix decidua* by Mosca et al.[9]. A further 6 primer pairs were designed and tested on *L. sibirica* by Semerikov et al.[10], of which 5 single primers were modified, and 2 newly designed in order to improve the primer binding affinity (see Supplementary Table 7).

DNA for the bait production was extracted from 80 mg of needles of a *Larix sibirica* individual sampled in 2018 from the Tazovsky District, Yamalo-Nenets Autonomous Okrug of Tyumen Oblast, Russia. The needles were transferred into impact-resistant 2 ml tubes together with two DNA-free steel beads of 5 mm diameter and ground to powder with FastPrep-24 (MP Biomedicals) for 50 s at 4 m s^{-1}. The DNeasy Plant Mini Kit (Qiagen) was used to isolate total genomic DNA, according to the manufacturer's protocol. The PCR products' lengths span between 169 and 683 bp (see Table S5). The baits/amplicons were produced by PCR using 18 ng input DNA, 2.5 U Platinum™ Taq High Fidelity DNA Polymerase (Invitrogen, USA), 1x PCR buffer (Invitrogen), 0.25 mM dNTPs (Invitrogen), 0.8 mg ml^{-1} BSA (VWR, Germany), 2 mM $MgSO_4$ and 0.2 µM forward and reverse primers. PCR was carried out with the following cycling conditions: 5 minutes (min) initial denaturation at 94 °C, 45 cycles with 30 seconds (sec) at 94 °C, 30 sec at the specific annealing temperature (see Supplementary Table S8) and 1 min at 68 °C, followed by a final extension at 10 min at 72 °C. To check for potential contamination a no template control (NTC) was carried along with each PCR extraction and treated identically to the samples. To check successful PCR amplification, the amplicon size was checked using 1-2 % agarose (Roth, Germany) gels.

According to the hybridization protocol from Maricic et al.[11], bait lengths should be below 1000 bp. In order to have a uniform bait pool, five PCR products with a length higher than 1000 bp were pooled in equimolar ratios in a volume of 130 µl and sonicated using a Covaris M220 Focused-ultrasonicator (Covaris, USA) to a target peak of 450 bp (which is the mean length of the

other 60 PCR products) with settings of peak incident power 50 W, duty factor 20%, cycles per burst 200 and treatment time 70 s. The fragment size and distribution were visualized with Agilent TapeStation (D1000 ScreenTape, Agilent Technologies). Resulting fragment sizes ranged from 100 to 1000 bp with an average size of 450 bp.

The sheared and non-sheared amplicons were pooled together in equimolar ratios and purified using the MinElute PCR Purification Kit (Qiagen), following the manufacturer's recommendations and eluted in 30 µl.

Blunt-ending, Adapters (Bio-T and B) ligation and Dynal™ Dynabeads™ M-270 Streptavidin (Invitrogen) ligation were conducted as described in Maricic et al.[11], with modification as described in Schulte et al.[3].

Hybridization capture

DNA libraries of the ancient lake sediment samples were pooled in equimolar amounts according to the lake of origin. Negative controls of DNA extraction and library preparation were also pooled with a fixed volume of 1 µl. According to Maricic et al.[11], ~2000 ng DNA of pooled libraries (details in Table S6) and 500 ng of baits were used for the hybridization capture experiment. In particular, the library pools of seven lakes (see Tables S3 and S6) were hybridized with 250 ng of baits targeting the chloroplast of *Larix* and 250 ng of baits targeting the 65 *Larix* candidate adaptive genes, while the Lake CH12 pool was hybridized only with 500 ng of baits covering the 65 candidate adaptive genes.

The hybridization capture experiment targeting both nuclear genes and the chloroplast genome was performed following the protocol of Maricic et al.[11] with the following modifications: 1) Two rounds of hybridization capture were performed for each pool. The input concentration, the number of amplification cycles and the concentration of captured material after each hybridization are described in Table S5**.** 2) To prevent binding of library molecules to the adapter sequences, which would result in off-target capture, the adapter sequences were blocked prior to the capture experiment by blocking oligonucleotides. The blocking oligos were implemented according to Schulte et al.[3]. 3) In order to reduce the tube surface and prevent the binding of magnetic beads to the tube wall during the rotation, 0.5 ml DNA LoBind® tubes (Eppendorf, Germany) were used for the hybridization capture incubation. The hybridization capture experiment of Lake CH12 pool targeting only the 65 nuclear genes was performed as described in Schulte et al.[3], i.e. with only one round of capture in 1.5 ml DNA LoBind® tubes (Eppendorf).

Sequencing

The final enriched library pools targeting both nuclear genes and the chloroplast genome, were pooled equimolarly to one sequencing pool of 10 nM and sequenced by Fasteris SA Sequencing Service (Geneva, Switzerland) on one SP flow-cell on an Illumina NovaSeq6000 instrument with V1.5 chemistry, 500 cycles kit and 2 x 250 bp paired-end sequencing. From the capture enriched Lake CH12 pool targeting the nuclear genes, 10 nM were sequenced by Fasteris SA Sequencing Service on one lane of an Illumina MiSeq instrument with V2 chemistry and 2 x 150 bp paired-end sequencing.

Data analysis

Demultipelxed and adapter-trimmed fastqc files as obtained from the sequencing company were quality checked using FASTQC[12] before and after deduplication with CLUMPIFY[13] (v.38.87) and trimming with FASTP[14] (v.0.20.1) with the parameters --merge, --length_required=30, --overlap_len_require=5, --correction, --low_complexity_filter, --cut_front, --cut_tail, --cut_window_size=4, and --cut_mean_quality=10.

Analysis of chloroplast enrichment

Reads were classified using KRAKEN2[15] (version 2.1.1) with a confidence threshold of 0.8 against a plastid database (RefSeq release of NCBI[16]) downloaded in July 2021. Reads classified to *Larix* at genus or species level were extracted using KRAKENTOOLS extract_kraken_reads.py[17] (v. 0.1). *Larix*-classified reads were aligned against an *L. gmelinii* chloroplast reference genome (NCBI GenBank accession number MK468637) using BWA ALN[18] (v. 0.7.17) with the parameters -l 1024 -o 2 -n 0.01 as recommended for ancient DNA read mapping by Oliva et al.[19]. Further processing of the alignment files such as conversion, sorting and indexing was done using SAMTOOLS[20] (v. 1.11). Duplicates were removed a second time, as deduplication is more efficient in aligned reads. Deduplication was done using PICARD MARKDUPLICATES[21] (v.2.24.1) for merged and unpaired reads and SAMTOOLS MARKDUP[20] for unmerged reads with default parameters. Ancient damage patterns were assessed and quality scores of likely damaged positions rescaled using MAPDAMAGE2[22] (v. 2.2.1). Alignments produced with overlapping merged reads, unmerged reads and unpaired reads were merged to one bam file using SAMTOOLS MERGE[20].

Variants were called for all samples conjointly using FREEBAYES[23] (v. 1.3.2) with the options --pooled-continuous and --min-base-quality 10. The produced vcf file was processed using PLINK[24] (v.1.90b4) to filter samples with over 90% missing sites and VCFTOOLS[25] (v. 0.1.16) to filter sites for maximal 70% of missing data and a minor allele frequency greater than 1%.

For a comparison of the variants in the samples with known references, all available chloroplast genomes from the Siberian boreal larch species western range *L. gmelinii* (11 species), eastern range *L. gmelinii* (or *L. cajanderi* (7 spec.)) and *L. sibirica* (1 spec.) were downloaded (NCBI GenBank accession numbers: MK468630.1-MK468636.1 and MK468638.1-MK468648.1, NC_036811.1; number of available genomes given in parentheses). All genomes were aligned against the same *L. gmelinii* reference genome as used for the sample MK468637 using BWA MEM[18]. Variants were called conjointly using FREEBAYES[23] with parameter settings --min-alternate-count, 1 --min-alternate-fraction 0, --haplotype-length 0, and --pooled-continuous.

Variants of samples and references were compared and plotted using R[26] and the packages tidyverse[27], readxl[28], cowplot[29] and ggh4x[30] with colors based on Okabe & Ito[31]. Maps were plotted in R using the packages ggmap[32], rgdal[33], sp[34], broom[35] and scatterpie[36].

Analysis of nuclear bait set enrichment

Analysis was conducted as described above. Differences in the analysis are: as database for classification, the plant database of RefSeq was used and 4 genomes of sequenced Pinaceae added using KRAKEN2 (*Picea abies, P. glauca, Pinus taeda,* NCBI GenBank accession numbers GCA_900067695.1, GCA_000411955.6 and GCA_000404065.3). Reads classified to genus *Larix* using KRAKEN2 with a confidence threshold 0.8 were mapped against the set of nuclear genes (see Supplementary Table 2) with the same tools and parameters as described for the analysis of chloroplast reads. As very few reads only mapped to the nuclear genes target probe set, unmapped *Larix*-classified reads were closely inspected using *de-novo* assembly and BLAST in

Geneious[37], which revealed a high content of repetitive DNA, in particular of the most abundant satellite repeat of *Larix.* This satellite repeat was first described by Hizume et al.[38] and more thoroughly described by Heitkam et al.[39] in European larch (*L. decidua)* and Japanese larch (*L. kaempferi).* To check whether or not the consensus sequence (most abundant version of the satellite repeat sequence in the genome) differs between the *Larix* species *L. gmelinii,* and *L. sibirica,* a comparative repeat analysis was undertaken (see description in next section). As the consensus sequence in the three species differed only at one position (*L. sibirica* carries at position 96 a cytosine instead of an adenosine), the alignment of *Larix*-classified reads was repeated with the published consensus sequence of EulaSat1[39]. To get reads mapping at the end of one repeat sequence (in the genome, satellite sequences are arranged in large arrays of repeated monomers), a triplet of the monomer was used as a reference for the alignment.

Enrichment of the satellite repeat

To test whether the repeat sequence became enriched by the capture approach and by which set of baits, we compared three datasets produced from Lake CH12 samples: 1) the un-enriched shotgun dataset (ENA project number PRJEB35838, sample accession numbers SAMEA6430888-91), 2) the target-enriched capture dataset using the complete chloroplast genome as a hybridization probe set (ENA project number PRJEB35838, sample accession numbers SAMEA6430894-97) (datasets 1 and 2 published by Schulte et al.[3]), and 3) the samples of Lake CH12 enriched for the set of nuclear genes produced in this study. Reads were quality trimmed and merged as described above and aligned to the triplet of the EulaSat1 consensus sequence as described above. Percentages of mapped reads to the repeat of the total quality trimmed reads were calculated.

Comparative repeat analysis of *L. gmelinii* and *L. sibirica*

To estimate the amount of repetitive DNA and to quantify EulaSat1 in the three *Larix* genomes, read clustering was performed using the RepeatExplorer Pipeline[40] with standard parameters using paired-end reads of publicly available *Larix* accessions: SRR8555411 (*L. sibirica*), SRR9610223 (*L. gmelinii,* eastern range), and SRR9610240 (*L. gmelinii,* western range). Proportions of sequence families that constitute more than 0.01% of the genome, including long terminal repeat (LTR) retrotransposons, DNA transposons, ribosomal DNA, Long and Short Interspersed Nuclear Elements (LINES/SINES) and satellite repeats, are displayed in Fig. S2. Initial quality filtering of the reads was done using TRIMMOMATIC[41], to provide 1 Mio reads per species with a consistent length of 93 nucleotides (LEADING:5 TRAILING:30 CROP:93 MINLEN:93) for the analyses. Read quality was checked using FastQC[12]. Bowtie2[42] was applied to identify and subsequently remove reads representing plastid DNA from the sequence data by mapping them to a database containing chloroplast DNA of the three individuals (NC_036811.1, MK468640.1, MK468648.1).

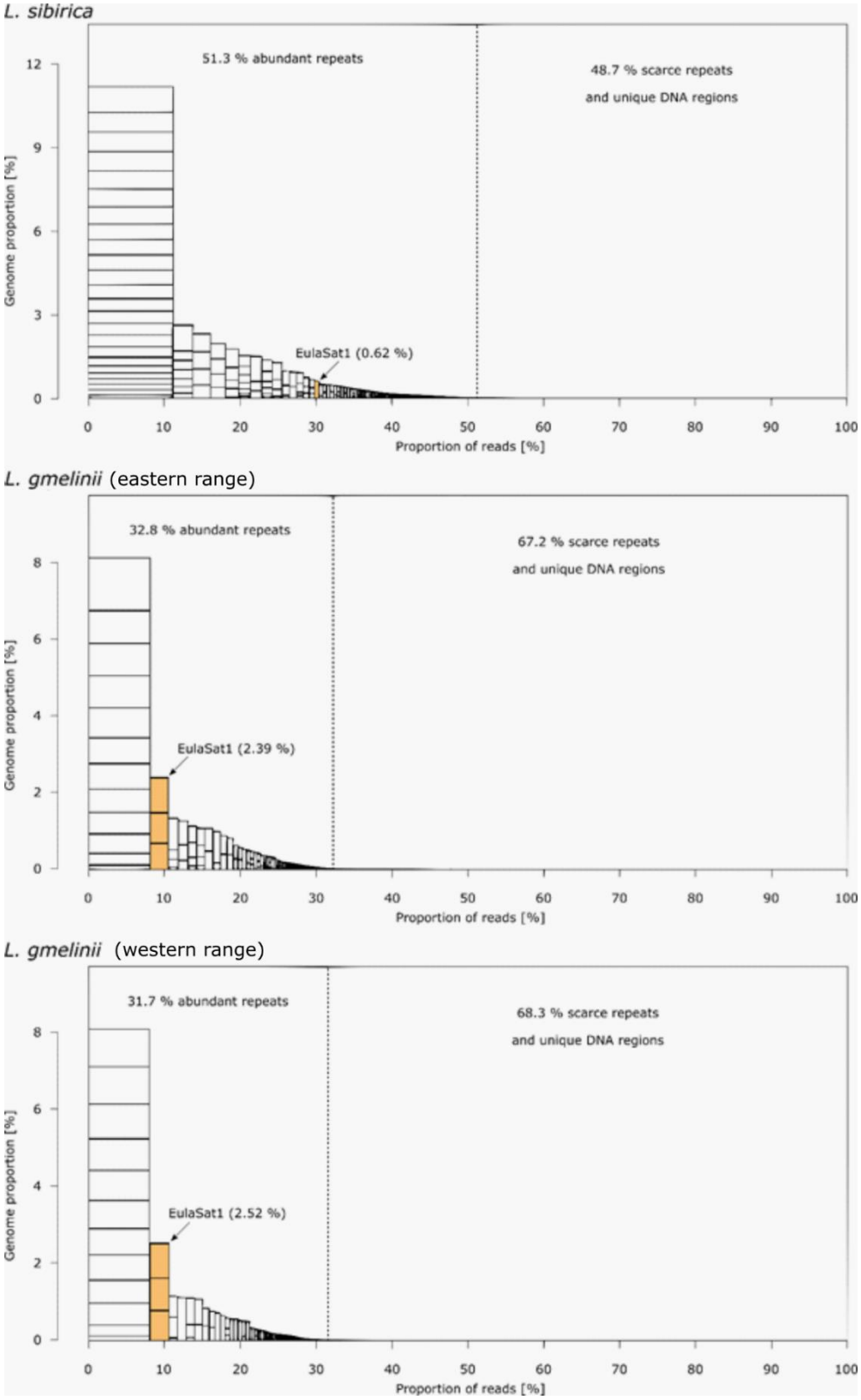

Fig. S2: Quantification of the most abundant repetitive sequences from *Larix sibirica* and *L. gmelinii* (western and eastern range) by genome-wide read clustering. Stacked bars represent repetitive DNA of the same sequence family, with the X- and Y-axes being a measure of abundance by read proportion. Repeats were considered as either abundant or scarce if their genome proportion exceeds or falls below 0.01% genome proportion, respectively. The read cluster corresponding to EulaSat1 is highlighted in all graphs.

Metabarcoding approach

Sampling, DNA extraction, PCR

To compare results from the hybridization capture approach with the metabarcoding approach, both published and newly produced data were used. Published datasets include Lake CH12[2] and Lake Bolshoye Shchuchye[4]. For lakes Billyakh, Kyutyunda, Lama, Malaya Chabyda and Satagay, new samples were processed (for sample list see Table S8). Core sampling and DNA extraction were done as described for the hybridization capture approach. PCR reactions were performed in three independent replicates using the trnL-g and trnL-h primers[43] modified to carry unique 8 bp tags on the 5' prime end preceded by NNN to improve cluster generation[44,45]. PCR set-up, conditions and sequencing were performed as described in Zimmermann et al.[46].

Data analysis of metabarcoding

Published and new metabarcoding datasets were analysed using OBITOOLS[47] (version 3.0.0b38). First, overlapping paired reads were merged using the command *alignpairedend*, then unmerged reads were removed with *grep -a mode:alignment.* The command *ngsfilter* was used to assign each sequence to its corresponding sample according to the tag combination. Consequently, reads were de-replicated with the command *unique* and cleaned from PCR and sequencing errors using *clean -r 0.05* (-r defines the maximum ratio allowed between the counts of sequence variants). Reads were assigned to a taxonomic level using *ecotag.* For taxonomic assignment, two databases were used: 1) a database based on the curated arctic and boreal vascular plant and bryophyte reference database published by Sønstebø et al.[48], Willerslev et al.[49] and Soininen et al.[50] and 2) a database based on the EMBL Nucleotide Database standard sequence release 143[51] (ftp://ftp.ebi.ac.uk/pub/databases/ena/sequence/release/). Databases were produced in OBITOOLS using the *ecopcr* command. In R[26], assignments of both databases were joined, preferentially using the scientific name assigned by the arctic and boreal vascular plant and bryophyte reference database, where the identity was higher or equal to the identity of the EMBL database. Subsequently, reads were filtered for a best identity higher than 0.97, counts for the PCR replicates were summed up and the percentage of reads assigned to *Larix* in the samples calculated.

Pollen

In addition to the Kyutyunda pollen record of core PG2022[52], nine samples of the parallel core PG2023 were selected for pollen analysis (Supplementary Table 12). Standard HF techniques were used for sample preparation[53]. *Lycopodium* marker spores were added to each sample to calculate total spore and pollen concentrations[54]. Water-free glycerol was used for sample storage and preparation of the microscopic slides. Pollen were analyzed using a 400x magnification and identified with the help of published pollen atlases[55–59]. Non-pollen-palynomorphs were identified when possible according to van Geel[60]. At least 250-300 pollen grains were counted in each sample. The relative frequencies of pollen taxa were calculated from the sum of the terrestrial pollen taxa. The percentages of fungal spores are based on the sum of the pollen and fungal spores, and the percentages of algae are based on the sum of pollen and algae. TGView software[61] (v. 1.7.16) was used for the calculation of percentages.

7.3.2 Additional Results & Discussions

Sequencing results

Three datasets were used in the study: 1) the hybridization capture dataset targeting both the chloroplast and a set of nuclear genes of *Larix* in 64 samples and 19 negative controls from seven

lake sediment cores, 2) the published hybridization capture data targeting the *Larix* chloroplast in four core samples of Lake CH12 (ENA project number PRJEB35838, sample accession SAMEA6430894-99), and 3) a dataset on the same CH12 core samples targeting the same set of nuclear genes as above. The three datasets comprised in total 1.15 billion paired reads: 946, 201, and 2 million (M) reads for the three datasets, respectively, of which 323, 54, and 1.5 M reads (380 M reads in total) remained after quality filtering and deduplication.

Enrichment of satellite repeat by nuclear gene bait set

To test whether the repeat sequence became enriched by the capture approach and by which set of baits, we compared three datasets produced from the same samples of Lake CH12: 1) the shotgun and 2) target enriched capture datasets produced in the study of Schulte et al.[3] using the complete chloroplast genome as a bait set and 3) a hybridization capture approach on the same sample libraries using only the nuclear gene bait set. The comparative analysis of the percentage of mapped reads against the sequence of EulaSat1 showed slight enrichment by the chloroplast bait set and a pronounced enrichment by the nuclear gene bait set (Fig. S3). This implies that the enrichment of off-target reads is linked either to the production or sequence composition of hybridization probes, and not a general by-product.

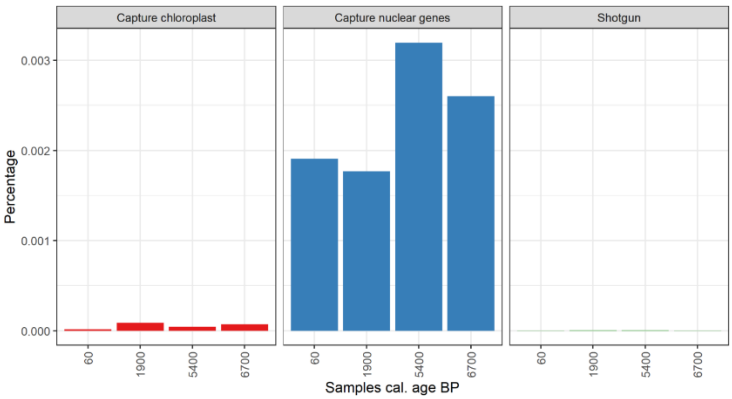

Fig. S3 Percentage of unclassified reads mapped against the EulaSat1 repeat sequence Percentages are displayed for three datasets: 1) capture enrichment using the *Larix* chloroplast genome, 2) capture enrichment using the nuclear bait gene set, and 3) shotgun sequencing. For all sequencing datasets the same libraries were used.

Nucleotide variants in EulaSat1 alignment

The nucleotide variants in the alignment of reads against the consensus sequence of the EulaSat1 repeat shows strong homogeneity (Fig. S4).

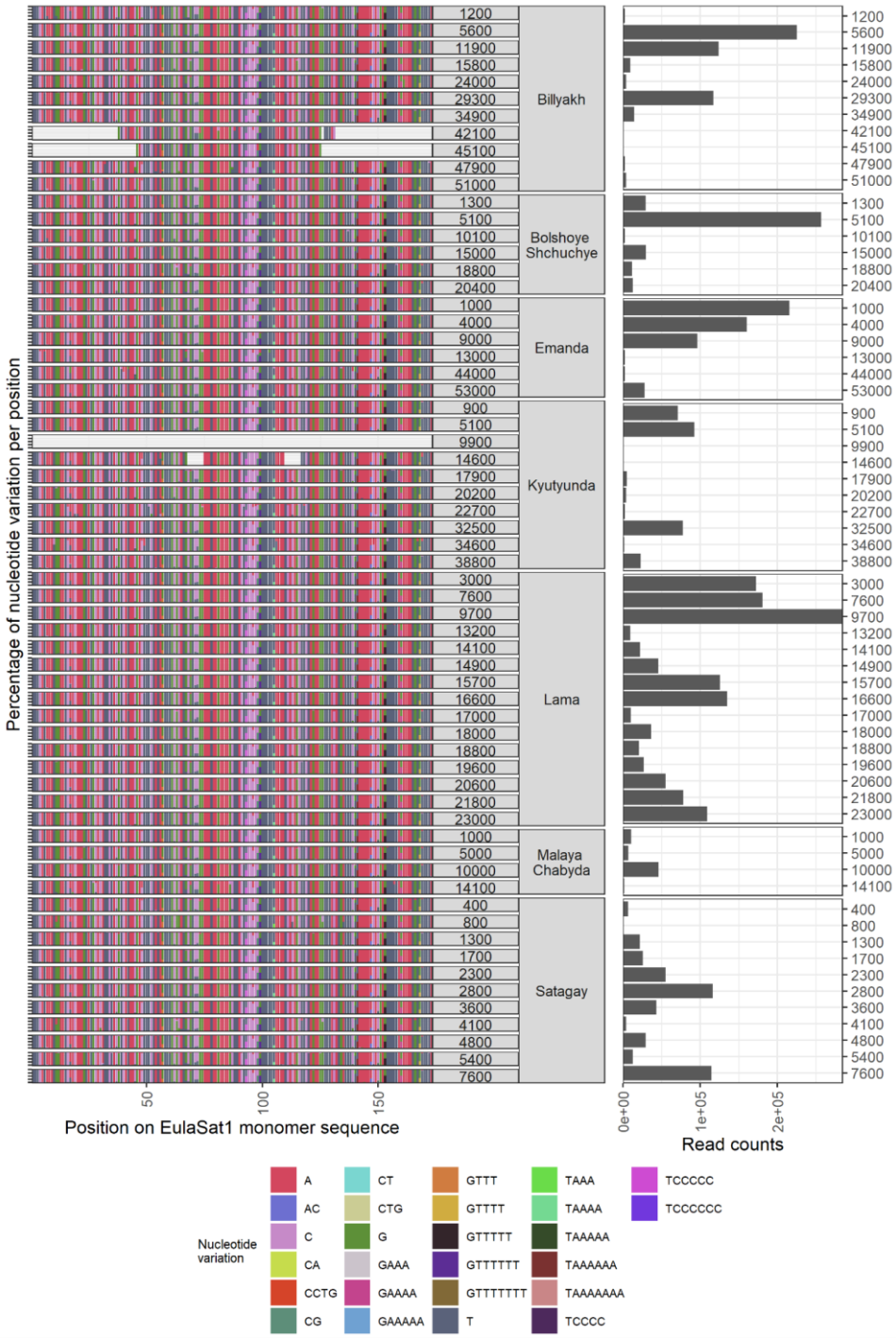

Fig. S4 Alignment of reads against the EulaSat1 repeat sequence. Left: Percentage of variants on the 173 positions of the repeat monomer arranged by calibrated age before present per lake, light gray background indicates positions with no coverage. Right: total read counts per sample.

Ancient damage patterns

Ancient damage patterns in the alignment of *Larix*-classified reads against the *Larix* chloroplast genome are visible for all samples with sufficient coverage (Supplementary Table 5). As typical for single-stranded library preparation methods, Cytosine (C) to Thymine (T) mutations are visible both at the 5’ and 3’ ends of the molecules[5]. Patterns are best visible in the biggest read fraction of overlapping merged reads. For high coverage samples, unmerged and unpaired reads also

show similar patterns. When looking at samples within one lake, the C to T mutation rate often increases with age, but this trend is not consistent in all samples and especially not in low coverage samples.

Evaluation of negative controls

For each library batch of 7 samples, one library was built only with HPLC grade water (Sigma Aldrich, USA) instead of sample DNA. The library blanks have very few reads that map to the chloroplast genome (1-5 reads) but most of them (8 from in total 12) have no reads mapping.

For each DNA extraction of 9 samples, one negative control without sediment was included. As samples from many different extraction batches were combined in this study, in some cases several extraction blanks were combined for one library preparation and therefore cannot be distinguished. In total 7 libraries with pools of extraction blanks were built. Two of the extraction blank libraries have no reads mapping to *Larix* (blanks for Lake Satagay, and one of the two blanks for Lake Billyakh). However, four of the extraction blank libraries have read counts mapping to the *Larix* chloroplast genome ranging from 49 to 94 and one of the libraries has 237 reads mapping to the reference (extraction blank for Lake Bolshoye Shchuchye). Considering that some of the low count samples have counts in the same range, a more detailed evaluation (see next section) is needed to determine whether any cross contamination occurred. Special attention is given to the signal of variants attributed to *L. sibirica,* which is one of the main results of the study.

Extraction batches showing potential cross contamination

For Lake Kyutyunda, samples were extracted in two extraction batches (LS130E and LS131E). In one batch (LS131E), the extraction blank library (JK102L-7) is extremely clean (the batch containing all the younger samples). The batch with the contaminated extraction blank library (JK106L-7-WDH) contains the older samples (samples of 38, 34 and 32 ka BP, extraction batch LS130E), which were extracted together with samples from Lake Malaya Chabyda. Samples of the two lakes show different signals, with Lake Kyutyunda exhibiting high proportions of *L. sibirica* variants and Lake Malaya Chabyda samples not showing any strong *L. sibirica* signal. Therefore, a cross-contamination between lakes is highly unlikely and the interpreted signal of *L. sibirica* in Lake Kyutyunda samples could not have come from a different source.

There are two libraries with pooled extraction blanks for Lake Billyakh. One of the blanks is extremely clean (JK102L-7, batch LS083E-LS131E-LS133E). This blank is an extraction blank for samples of ages 1, 5, 11 and 24 ka BP. The other blank is an extraction blank for samples of ages 34, 45, 47 and 51 ka BP (JK101L7-WDH, batch LS004E-LS005E-LS011E-LS022E). This means that the old samples are independent from the young samples. The samples from the blank in question were extracted in separate batches, each batch containing samples in a chronological manner. So, if there has been cross contamination, it is roughly from the same age/core depth and thus does not influence the interpretation of the observed patterns.

Samples of Lake Emanda were all extracted in one batch with one extraction blank. Cross contamination between blanks could have happened between the samples. However, none of the samples, nor the blank shows a strong signal of *L. sibirica* variants, so any potential cross contamination did not affect the interpretation of the results.

The blank for samples from Lake Satagay was extremely clean, cross-contamination is therefore highly unlikely.

The library of extraction blanks for Lake Bolshoye Shchuchye shows the highest read counts of the sequenced blanks. The blank also combines the highest number of extraction blanks pooled for one library. The whole core was extracted in a chronological manner, as a result all samples stem from different extraction batches. A possible cross-contamination in the course of the DNA extraction could have happened only between samples of similar age, and not between samples included in our study.

Exclusion of reads present in negative controls

To check whether the reads present in the extraction blank libraries had an impact on the results, we subtracted those reads from the sample reads. In detail, for each sample the extraction blank library containing the negative control of the respective extraction batch was selected. Each position covered in the extraction blank library was completely excluded from the belonging samples. The result is displayed in Fig. S5 and shows no difference in the interpreted results from the unfiltered samples.

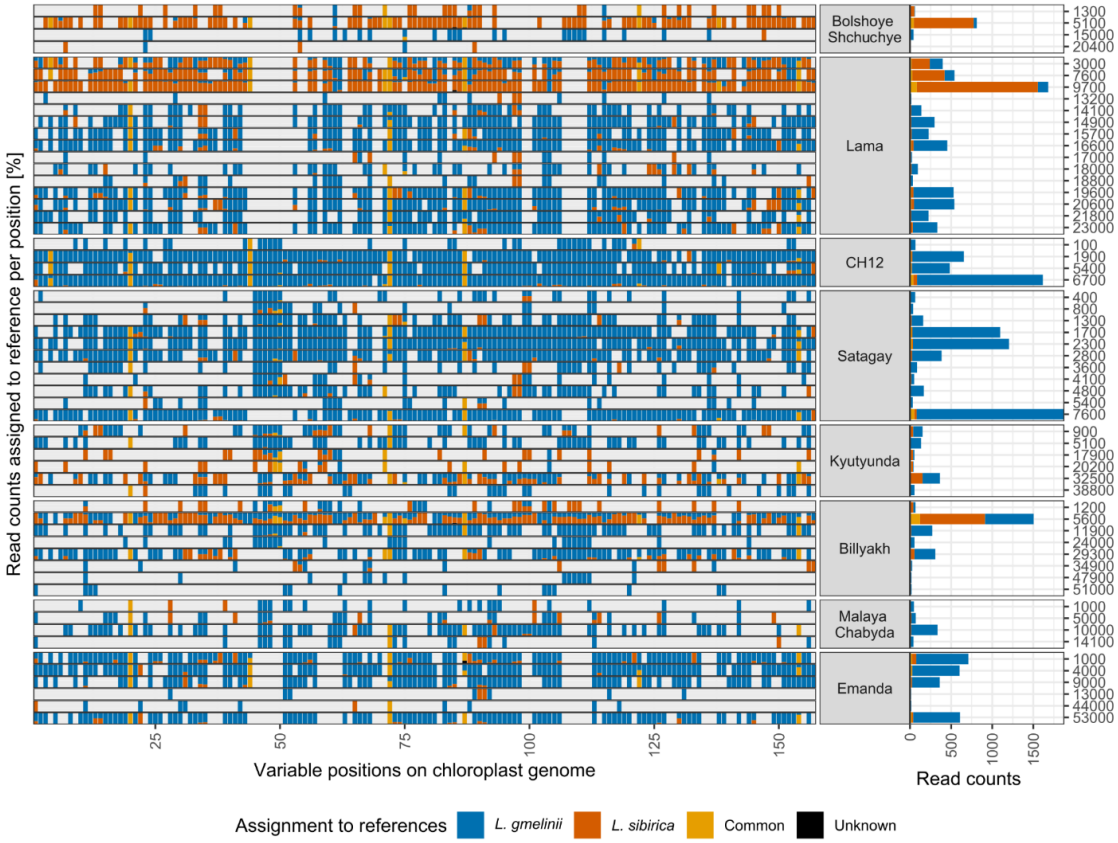

Fig. S5: Percentage and counts of variable positions on the chloroplast genome assigned to reference species with exclusion of blanks. Left: Alignment of *Larix*-classified reads against the chloroplast genome. Positions that are covered in the blank libraries are excluded from the corresponding samples. Only 157 variable positions between the reference species were considered. Colors indicate the percentage of variants at the positions assigned to the different *Larix* species. Each row represents one sample, arranged by calibrated age before present per lake. Gray background indicates positions with no coverage. Right: Number of read counts at the variable positions assigned to references for each sample.

7.3.3 References

1. Blaauw, M. & Christen, J. A. Bacon. 1–15 (2011).

2. Epp, L. S. *et al.* Temporal and spatial patterns of mitochondrial haplotype and species

distributions in Siberian larches inferred from ancient environmental DNA and modeling. *Sci. Rep.* **8**, 17436 (2018).

3. Schulte, L. *et al.* Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Mol. Ecol. Resour.* **21**, 801–815 (2021).

4. Clarke, C. L. *et al.* Persistence of arctic-alpine flora during 24,000 years of environmental change in the Polar Urals. *Sci. Rep.* **9**, 19613 (2019).

5. Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).

6. Gansauge, M.-T. *et al.* Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Res.* **45**, gkx033 (2017).

7. Zimmermann, H. H. *et al.* Chloroplast and mitochondrial genetic variation of larches at the Siberian tundra-taiga ecotone revealed by de novo assembly. *PLoS One* **14**, e0216966 (2019).

8. Eckert, A. J. *et al.* Patterns of population structure and environmental associations to aridity across the range of loblolly pine (Pinus taeda L., Pinaceae). *Genetics* **185**, 969–982 (2010).

9. Mosca, E. *et al.* Contrasting patterns of nucleotide diversity for four conifers of Alpine European forests. *Evol. Appl.* **5**, 762–775 (2012).

10. Semerikov, V. L. & Lascoux, M. Nuclear and cytoplasmic variation within and between Eurasian *Larix* (Pinaceae) species. *Am. J. Bot.* **90**, 1113–23 (2003).

11. Maricic, T., Whitten, M. & Pääbo, S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* **5**, e14004 (2010).

12. Andrews, S. FastQC: A quality control tool for high throughput sequence data. (2015).

13. Bushnell, B. BBMap. (2019).

14. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

15. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *bioRxiv* 762302 (2019) doi:10.1101/762302.

16. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

17. Lu, J. KrakenTools. (2020).

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

19. Oliva, A., Tobler, R., Cooper, A., Llamas, B. & Souilmi, Y. Systematic benchmark of ancient DNA read mapping. *Brief. Bioinform.* **00**, 1–12 (2021).

20. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

21. Broad Institute. Picard toolkit. *GitHub-Repository* (2019).

22. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).

23. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. 1–9 (2012).

24. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16 (2015).

25. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

26. R Core Team. R: A language and environment for statistical computing. (2013).

27. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

28. Wickham, H. & Bryan, J. readxl: Read Excel Files. *R package version 1.3.1* https://cran.r-project.org/package=readxl (2019).

29. Wilke, C. O. cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2." *R package version 1.1.1* https://cran.r-project.org/package=cowplot (2020).

30. van den Brand, T. ggh4x: Hacks for "ggplot2." *R package version 0.1.2.1* https://cran.r-project.org/package=ggh4x (2021).

31. Okabe, M. & Ito, K. Color Universal Design (CUD) - How to make figures and presentations that are friendly to Colorblind people -. https://jfly.uni-koeln.de/color/ (2002).

32. Kahle, D. & Wickham, H. ggmap: Spatial Visualization with ggplot2. *R J.* **5**, 144–161 (2013).

33. Bivand, R., Keitt, T. & Rowlingson, B. rgdal: Bindings for the "Geospatial" Data Abstraction Library. *R package version 1.5-23* https://cran.r-project.org/package=rgdal (2021).

34. Bivand, R., Pebesma, E. & Gomez-Rubio, V. *Applied spatial data analysis with R*. (Springer, 2013).

35. Robinson, D., Hayes, A. & Couch, S. broom: Convert Statistical Objects into Tidy Tibbles. *R package version 0.7.6* https://cran.r-project.org/package=broom (2021).

36. Yu, G. scatterpie: Scatter Pie Plot. *R package version 0.1.6* https://cran.r-project.org/package=scatterpie (2021).

37. Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

38. Hizume, M. *et al.* Tandem repeat DNA localizing on the proximal DAPI bands of chromosomes in \textit{*Larix*, Pinaceae}. *Genome* **45**, 777–783 (2002).

39. Heitkam, T. *et al.* Comparative Repeat Profiling of Two Closely Related Conifers (*Larix* decidua and *Larix* kaempferi) Reveals High Genome Similarity With Only Few Fast-Evolving Satellite DNAs. *Front. Genet.* **12**, (2021).

40. Novák, P., Neumann, P., Pech, J., Steinhaisl, J. & MacAs, J. \textit{RepeatExplorer}: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792–793 (2013).

41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–360 (2012).

43. Taberlet, P. *et al.* Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14–e14 (2007).

44. De Barba, M. *et al.* DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Mol. Ecol. Resour.* **14**, 306–323 (2014).

45. Binladen, J. *et al.* The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* **2**, 1–9 (2007).

46. Zimmermann, H. H. *et al.* Sedimentary ancient DNA and pollen reveal the composition of plant organic matter in Late Quaternary permafrost sediments of the Buor Khaya Peninsula (north-eastern Siberia). *Biogeosciences* **14**, 575–596 (2017).

47. Boyer, F. *et al.* OBITOOLS : a UNIX-inspired software package for DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 176–182 (2016).

48. Sonstebo, J. H. *et al.* Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Mol. Ecol. Resour.* **10**, 1009–1018 (2010).

49. Willerslev, E. *et al.* Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* **506**, 47–51 (2014).

50. Soininen, E. M. *et al.* Highly Overlapping Winter Diet in Two Sympatric Lemming Species Revealed by DNA Metabarcoding. *PLoS One* **10**, e0115335 (2015).

51. Kanz, C. *et al.* The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* **33**, D29-33 (2005).

52. Biskaborn, B. K. *et al.* Late Quaternary vegetation and lake system dynamics in north-eastern Siberia: Implications for seasonal climate variability. *Quat. Sci. Rev.* **147**, 406–421 (2016).

53. Berglund, B. E. & Ralska-Jasiewiczowa, M. Pollen analysis and pollen diagrams. in *Handbook of Holocene palaeoecology and palaeohydrology.* (ed. Berglund, B. E.) 455–484 (Wiley, Chichester, 1987).

54. J., S. Tablets with spores used in absolute pollen analysis. *Pollen et Spores* **13**, 614–621 (1971).

55. Kupriyanova, L. A. & Alyoshina, L. A. *Pollen and spores of plants from the flora of European part of USSR*. (Academy of Sciences USSR, Komarov Botanical Institute, 1978).

56. Bobrov, A. E., Kupriyanova, L. A. & Litvintseva, M. V. *Spores and pollen of gymnosperms from the flora of the European part of the USSR*. (Nauka, 1983).

57. Reille, M. *Pollen et spores d'Europe et d'Afrique du nord. Supplement 2*. (Laboratoire de Botanique Historique et Palynologie, 1998).

58. Reille, M. *Pollen et spores d'Europe et d'Afrique du nord Supplement 1*. (Laboratoire de Botanique Historique et Palynologie, 1995).

59. Reille, M. *Pollen et spores d'Europe et d'Afrique du nord*. (Laboratoire de Botanique Historique et Palynologie, 1992).

60. Van Geel, B. Non-Pollen Palynomorphs. in *Tracking Environmental Change Using Lake*

Sediments. Developments in Paleoenvironmental Research (eds. Smol, J. P., Birks, H. J. B., Last, W. M., Bradley, R. S. & Alverson K.) 99–119 (Springer, Dortrecht, 2002). doi:10.1007/0-306-47668-1_6.

61. Grimm, E. C. TGView. (2004).

Acknowledgements

I have received a great deal of support throughout the time of pursuing my PhD. I would like to start by thanking my supervisor Ulrike Herzschuh for sharing her expertise and knowledge with me and for her her ongoing support and guidance throughout the years.

Many thanks I owe to Kathleen Stoof-Leichsenring for her support in all laboratory issues and organizations. Special thanks go to Tony Heitkam and Laura Epp who were great mentors and always provided good advice.

I am deeply grateful to my superb fellow PhD students and colleagues for always open ears, good discussions in coffee and lunch breaks and help in various bigger and smaller problems that I encountered during the last years. I especially want to thank Barbara von Hippel, Simone Stünzi, Heike Zimmermann, Stefano Meucci, Sichao Huang, Sisi Liu, Raphael Hébert and Thomas Böhmer. I also thank Jeremy Courtin and Amedea Perfumo for our good times at our little coffee club and Raphael Köhler for a great DokTeam time.

For help with DNA extraction, library builds and support in the lab I would like to thank Janine Klimke, Sarah Olischläger, Svetlana Karachurina as well as all student helpers who helped me with the core sampling for the multi-core project.

Simeon Lisovski, Boris Biskaborn, Inger Alsos, Andrej Andreev, Chenzhi Li and all other co-authors I would like to thank for their valuable feedback on the manuscripts.

I would like to express particular gratitude to thank Claudia Hanfland and Claudia Sprengel from POLMAR graduate school for being such great supporters to all AWI PhD students including me. I would furthermore like to thank Sigrun Gräning for her never ending patience in supporting me in administrative issues.

I am forever grateful to my parents, my sisters and my brothers who always supported and encouraged me. Particularly, I want to thank my sisters Wilhelmine and Helene for their help in proof-reading this thesis.

Finally, I want to thank my partner, Nico, for his overall and never ending support and love.

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel „Dynamics of *Larix* (Mill.) species in Siberia during the last 50,000 years inferred from sedimentary ancient DNA" selbstständig und unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Wörtlich oder sinngemäß übernommenes Gedankengut habe ich als solches kenntlich gemacht. Diese Dissertation wird erstmalig an der Universität Potsdam eingereicht, dem Verfahren zu Grunde liegende Promotionsordnung ist mir bekannt.

Berlin, 22.12.2021

Luise Schulte