

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

Luise Schulte<sup>1,2</sup>  | Nadine Bernhardt<sup>1</sup>  | Kathleen Stoof-Leichsenring<sup>1</sup>  |  
Heike H. Zimmermann<sup>1</sup>  | Ludmila A. Pestryakova<sup>3</sup>  | Laura S. Epp<sup>1</sup>  |  
Ulrike Herzschuh<sup>1,2,4</sup> 

<sup>1</sup>Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Forschungsstelle Potsdam, Potsdam, Germany

<sup>2</sup>Institut für Biochemie and Biologie, Universität Potsdam, Potsdam, Germany

<sup>3</sup>Institute of Natural Sciences, North-Eastern Federal University of Yakutsk, Yakutsk, Russia

<sup>4</sup>Institut für Geowissenschaften, Universität Potsdam, Potsdam, Germany

## Correspondence

Ulrike Herzschuh, Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar und Meeresforschung, Forschungsstelle Potsdam, Potsdam, Germany.

## Present address

Nadine Bernhardt, Julius Kühn-Institut Bundesforschungsinstitut für Kulturpflanzen, Quedlinburg, Germany  
Laura S. Epp, Limnologisches Institut, Universität Konstanz, Konstanz, Germany

## Funding information

European Research Council, Grant/Award Number: 772852; German Research Foundation, Grant/Award Number: EP98/2-1

## 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.

## KEYWORDS

chloroplast genome, hybridization capture, *Larix*, sediment core, sedimentary ancient DNA, target enrichment

L. S. Epp: Second senior author.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

## 1 | 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 (*sedaDNA*) 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 *sedaDNA* 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 (Sønstebo et al., 2010; Taberlet et al., 2007), let alone show sub-specific 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. Kirstler 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 (Szmids 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 *sedaDNA* samples from a small lake in the Taymyr region of north-eastern 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.

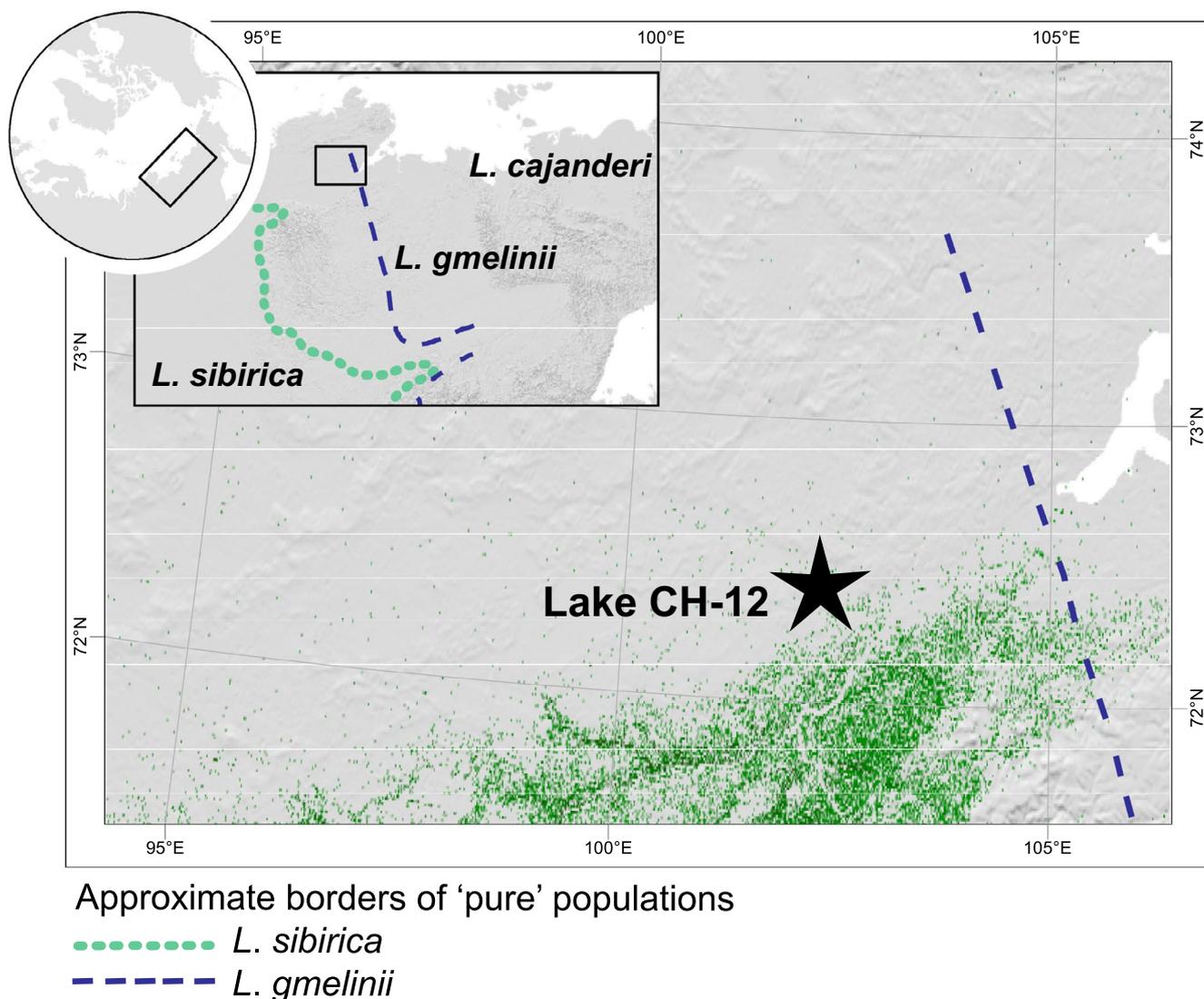
As a proof of concept, four samples were both shotgun-sequenced and enriched by hybridization capture for the chloroplast genome of *L. gmelinii*, to evaluate how well we can retrieve and assemble genome-scale data—here the complete chloroplast genome of *Larix*—from sedimentary ancient DNA. We demonstrate 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.

## 2 | METHODS

### 2.1 | 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



**FIGURE 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 (2013) [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

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.

## 2.2 | Laboratory work

### 2.2.1 | 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<sup>9</sup> to 10<sup>2</sup> copies/µl. qPCR was carried out in 1x 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). Cyclor 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.

### 2.2.2 | 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: 1x 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).

### 2.2.3 | 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  $\mu$ l. The sheared pool was blunt-ended using the Fast DNA End Repair Kit (Thermo Scientific) with 1.3  $\mu$ g DNA, 4  $\mu$ l End Repair Enzyme Mix and 1 $\times$  End Repair Reaction Mix in a volume of 100  $\mu$ l 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  $\mu$ l 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  $\mu$ l blunt-ended DNA, 1.25  $\mu$ M hybridized adapters Bio-T/B, 1 $\times$  rapid ligation buffer and 0.5 U T4 DNA ligase in a volume of 40  $\mu$ l 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  $\mu$ l elution buffer. The concentration was estimated with the ds-DNA BR assay and the Qubit<sup>®</sup> 2.0 Fluorometer (Invitrogen) using 1  $\mu$ l of the purified baits. Directly prior to the hybridization capture reaction, 500 ng of the purified baits was ligated to Dynal<sup>™</sup> Dynabeads<sup>™</sup> M-270 Streptavidin (Invitrogen) following the protocol described in Maricic et al. (2010).

### 2.2.4 | Hybridization capture

The enrichment was done following the protocol of Maricic et al. (2010). Another 24  $\mu$ l 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  $\mu$ g, 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: AATGATACGCGACCACCGATCTACAC-phosphate, BO2.P5.part2.F: ACACTCTTCCCTACACGACGCTCTCCGATCT-phosphate, BO3.P5.part1.R: GTGTAGATCTCGGTGGTCGCCGTATCATT-phosphate and BO4.P5.part2.R: AGATCGGAAGAGCGTCGTAG GGAAAGAGTGT-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  $\mu$ l 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.

## 2.3 | Data analysis

### 2.3.1 | 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).

### 2.3.2 | 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).

### 2.3.3 | 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.

### 2.3.4 | 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 of 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 GG-PLOT2 (R Core Team, 2013).

### 2.3.5 | 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).

### 2.3.6 | 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 variation 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 | RESULTS

### 3.1 | Overview of the shotgun and hybridization capture data sets

#### 3.1.1 | 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).

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.

### 3.1.2 | 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).

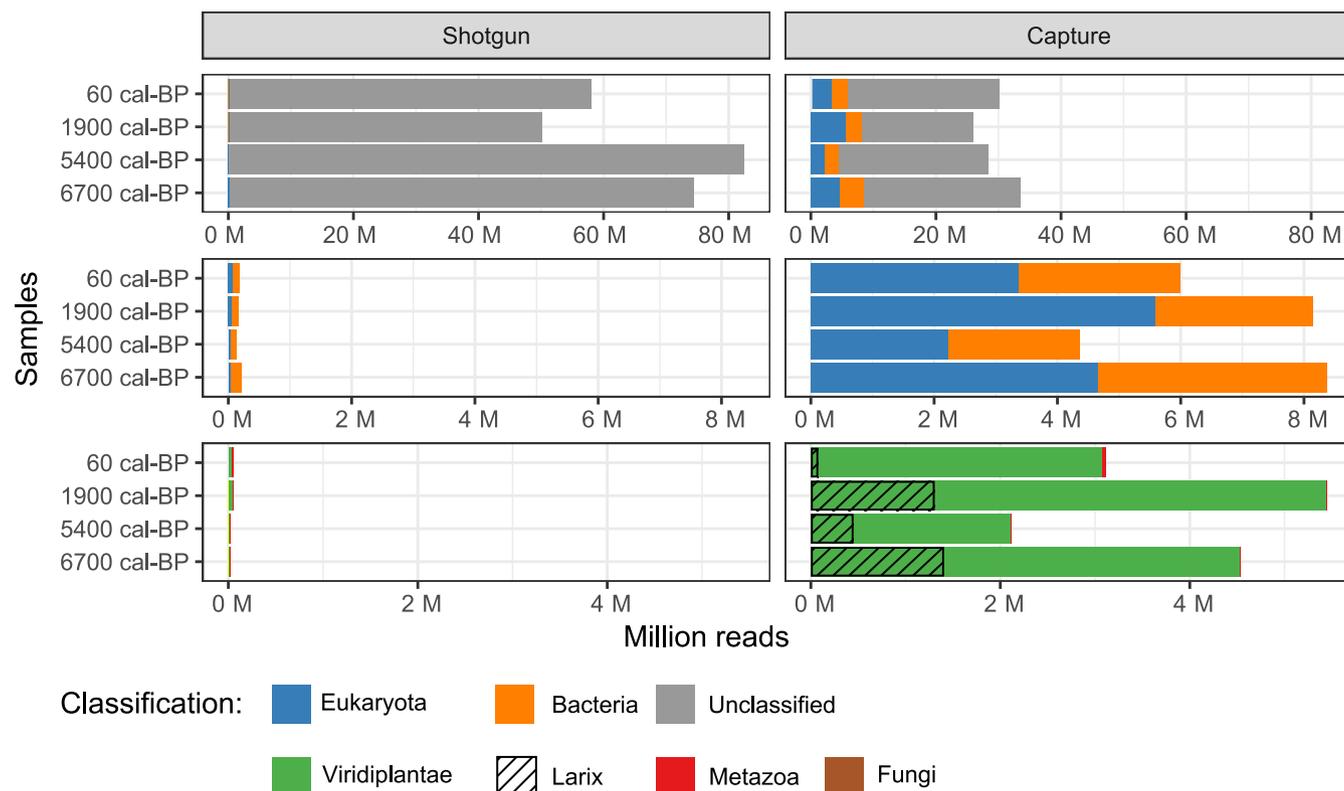
Classification with *KRAKEN2* using the nt database with a confidence threshold of 0.8 could classify 28% of the capture sample reads. Of the classified sample reads, the majority was classified as

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, de-duplicated reads of both *shotgun* and *capture* data sets, the enrichment ranged from 6.4- to 16.2-fold (Table S4).



**FIGURE 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 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 3.2 | 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  $92 \pm 36.7$  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).

### 3.3 | 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  $21.0x \pm 14.2x$  (6700 cal-BP),  $5.0x \pm 4.5x$  (5400 cal-BP),  $6.6x \pm 5.9x$  (1900 cal-BP) and  $0.3x \pm 0.7x$  (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.

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 variations, and 4% of the reads carry

*L. sibirica* variants (Figure 4). Almost no reads (0.4%) carry neither of the two species-specific variations ('other').

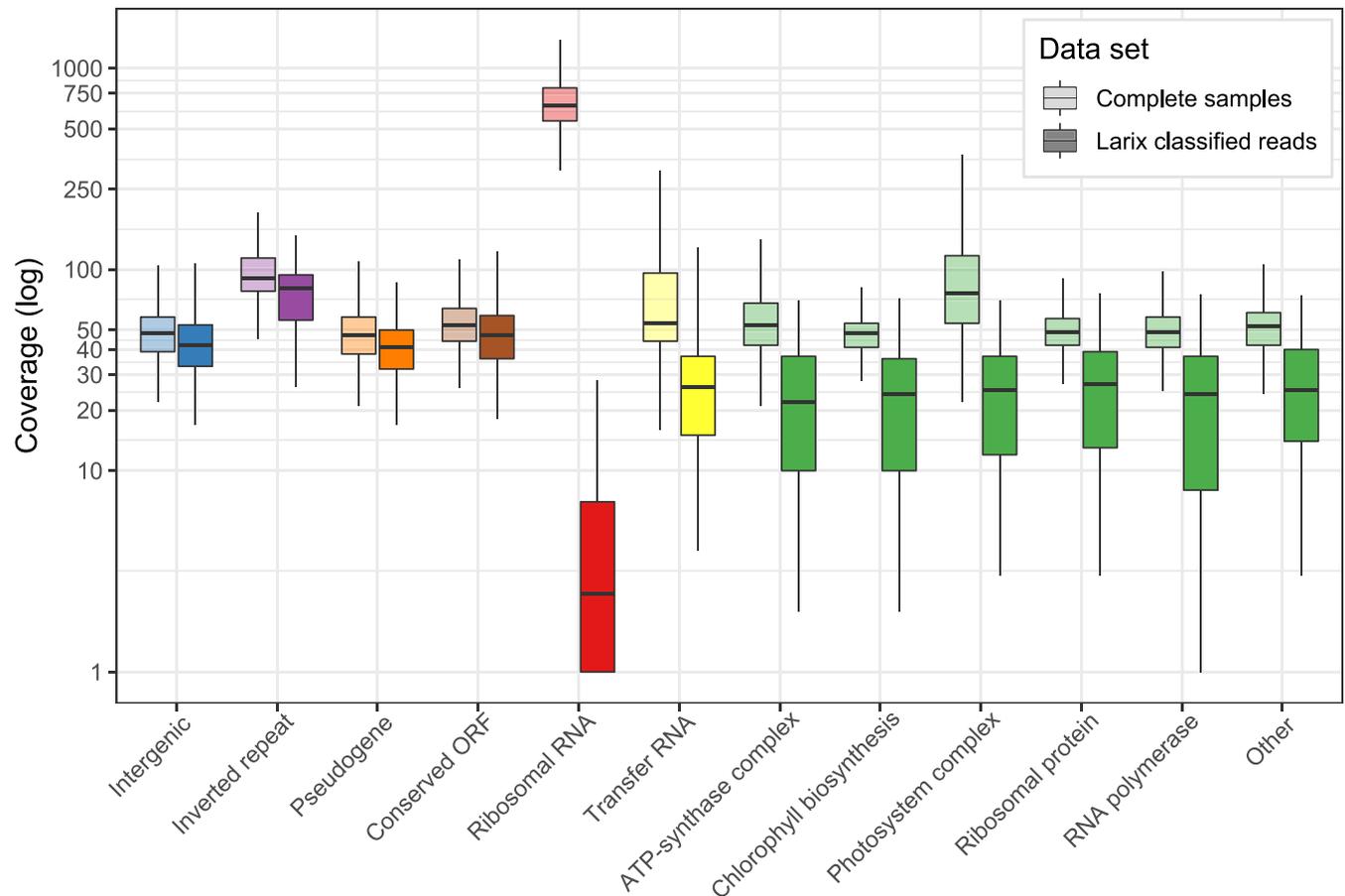
Between 0.3% and 51% of the analysed positions contained at least one read which was classified as *L. sibirica*, with the highest percentage detected at 6700 cal-BP and the lowest percentage detected at 60 cal-BP (Table S6). The ratio of *L. sibirica* variants over all positions and reads varied from 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 variations assigned to *L. gmelinii*, while 19.1% of the reads are classified as *L. sibirica* and 1.3% as 'other' (Table S6).

## 4 | 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.

### 4.1 | 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).



**FIGURE 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 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

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

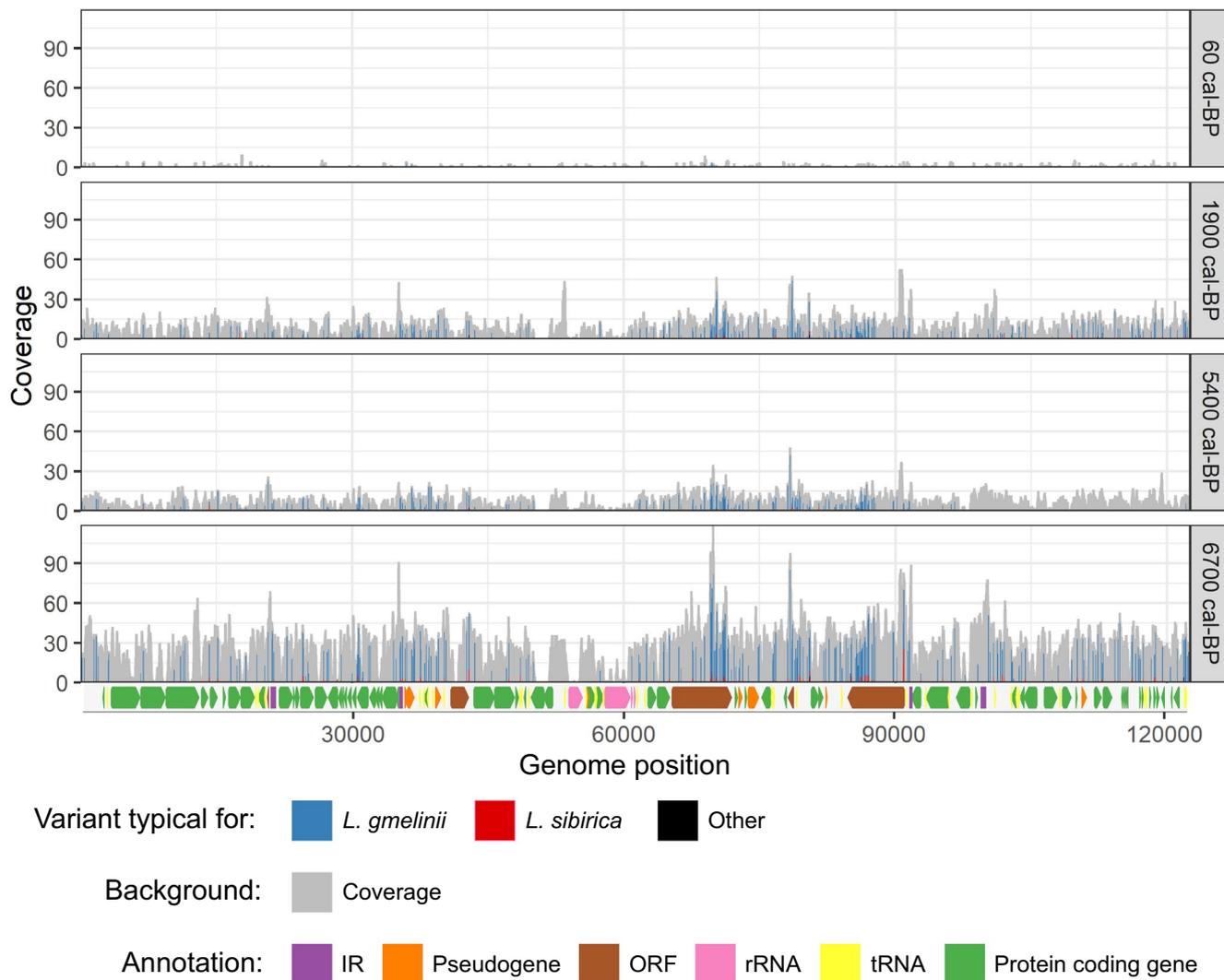
Hybridization capture resulted in an increase in taxonomically classified reads by orders of magnitude, especially with respect to the ratio of classified reads (0.3% to 28%), and also in absolute numbers of assigned reads (800 K reads to 35 M reads). These results show, for the first time, that DNA 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; Pajmans 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 (Pajmans 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.



**FIGURE 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 variation found in each of the two species or if a read contained a variation 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 [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

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.

### 4.3 | 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).

#### 4.4 | *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 (Szmíd 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.

#### 4.5 | 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 (Giguët-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; Pisarić et al., 2001) and is in correspondence with the reconstructed global cooling trend of the middle to late Holocene (Marcott et al., 2013).

## 5 | CONCLUSIONS

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 *sedDNA* 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.

#### ACKNOWLEDGEMENTS

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.

#### 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.

#### DATA AVAILABILITY STATEMENT

The Illumina sequence data are submitted to the European Nucleotide Archive under Project Number PRJEB35838, Accession Numbers ERS4197088–ERS4197099 (Schulte et al., 2020).

#### ORCID

Luise Schulte  <https://orcid.org/0000-0002-0428-8107>

Nadine Bernhardt  <https://orcid.org/0000-0003-0716-6284>

Kathleen Stoof-Leichsenring  <https://orcid.org/0000-0002-6609-3217>

[org/0000-0002-6609-3217](https://orcid.org/0000-0002-6609-3217)

Heike H. Zimmermann  <https://orcid.org/0000-0002-0225-5176>

Luidmila A. Pstryakova  <https://orcid.org/0000-0001-5347-4478>

Laura S. Epp  <https://orcid.org/0000-0002-2230-9477>

Ulrike Herzschuh  <https://orcid.org/0000-0003-0999-1261>

#### 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. <https://doi.org/10.1007/978-1-4020-9693-8>
- Ahmed, E., Parducci, L., Unneberg, P., Ågren, R., Schenk, F., Rattray, J. E., Han, L. U., Muschitiello, F., Pedersen, M. W., Smittenberg, R. H., Yamoah, 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, 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., Fordyce, S. L., Montiel, R., Vielle-Calzada, J.-P., Willerslev, E., & 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>
- Ávila-Arcos, M. C., Sandoval-Velasco, M., Schroeder, H., Carpenter, M. L., Malaspina, A.-S., Wales, N., Peñaloza, F., Bustamante, C. D., & Gilbert, M. T. P. (2015). Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. *Methods Ecol Evol*, 6, 725–734. <https://doi.org/10.1111/2041-210X.12353>
- 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. <http://broadinstitute.github.io/picard/>
- Carpenter, M. L., Buenostro, 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.
- CBOL Plant Working Group, Hollingsworth, P. M., Forrest, L. L., Spouge, J. L., Hajibabaei, M., Ratnasingham, S., van der Bank, M., Chase, M. W., Cowan, R. S., Erickson, D. L., Fazekas, A. J., Graham, S. W., James, K. E., Kim, K.-J., Kress, W. J., Schneider, H., van AlphenStahl, J., Barrett, S. C., van den Berg, C., ... Little, D. P. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31), 12794–12797. <https://doi.org/10.1073/pnas.0905845106>
- Curat, 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., 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 of the United States of America*, 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., 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., Gussarova, G., Boessenkool, S., Olsen, J., Haile, J., Schröder-Nielsen, A., Ludikova, A., Hassel, K., Stenøien, H. K., Funder, S., Willerslev, E., Kjær, K., & 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, L. S., Kruse, S., Kath, N. J., Stoof-Leichsenring, K. R., Tiedemann, R., Pestryakova, L. A., & Herzsuh, 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., Riehl, L. M., Schmidt, A., & Meyer, M. (2017). Single-stranded DNA library preparation from highly degraded DNA using T4 DNA ligase. *Nucleic Acids Research*, 45, e79. <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>
- Giguët-Covex, C., Ficetola, G. F., Walsh, K., Poulénard, J., Bajard, M., Fouinat, L., Sabatier, P., Gielly, L., Messenger, E., Develle, A. L., David, F., Taberlet, P., Brisset, E., Guiter, F., Sinet, R., & 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., Malaspina, A.-S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maric, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kučan, Ž., ... 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>
- 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>
- Herzsuh, U. (2019). Legacy of the Last Glacial on the present-day distribution of deciduous versus evergreen boreal forests. *Global Ecology and Biogeography*, 29, 198–206. <https://doi.org/10.1111/geb.13018>
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., dePamphilis, C. W., Leebens-Mack, J., Muller, K. F., Guisinger-Bellian, M., Haberle, R. C., Hansen, A. K., Chumley, T. W., Lee, S.-B., Peery, R., McNeal, J. R., Kuehl, J. V., & 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., 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>
- 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 of the United States of America*, 111(8), 2937–2941. <https://doi.org/10.1073/pnas.1318678111>
- Klemm, J., Herzsuh, 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*. <https://github.com/genom-e/bam-readcount>
- 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. <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., Marth, G., Abecasis, G., & 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., & Herzsuh, 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>
- Mohandesan, E., Speller, C. F., Peters, J., Uerpmann, H.-P., Uerpmann, M., De Cupere, B., Hofreiter, M., & Burger, P. A. (2017). Combined hybridization capture and shotgun sequencing for ancient DNA analysis of extinct wild and domestic dromedary camel. *Molecular ecology resources*, 17(2), 300–313. <https://doi.org/10.1111/1755-0998.12551>
- Moore, C. R., Brooks, M. J., Goodyear, A. C., Ferguson, T. A., Perrotti, A. G., Mitra, S., Listeck, 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>
- Murchie, T., Kuch, M., Duggan, A., Ledger, M. L., Roche, K., Klunk, J., Klunk, J., Karpinski, E., Hackenberger, D., Sadoway, T., Macphee, R., Froese, D. G., & 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](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., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & 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>
- Pajmans, 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., Gelly, L., Boyer, F., Zinger, L., Arnaud, F., Poulencard, 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>
- Parducci, L., Alsos Greve, I., Unneberg, P., Pedersen, M. W., Han, L., Lammers, Y., Salonen, J. S., Väiliranta, M. M., Slotte, T., & 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., 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/nature19085>
- Peñalba, J. V., Smith, L. L., Tonione, M. A., Sass, C., Hykin, S. M., Skipwith, P. L., McGuire, J. A., Bowie, R. C. K., & 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](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.* *Northeast Asia. Molecular Ecology*, 19(6), 1239–1252.
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <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., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B. O., 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>
- Sønsteby, J. H., Gelly, 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>
- 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., Vermet, 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. <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 pseudoregium 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., Bunce, M., Wiuf, C., Gilichinsky, D. A., & 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., Stoof-Leichsenring, K. R., Pestryakova, L. A., Wieczorek, M., Trense, D., & Herzsich, 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>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Schulte L, Bernhardt N, Stoof-Leichsenring K, et al. Hybridization capture of larch (*Larix* Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. *Mol Ecol Resour.* 2021;21:801–815. <https://doi.org/10.1111/1755-0998.13311>