Methods in Ecology and Evolution

BRITISH ECOLOGICAL SOCIETY

Methods in Ecology and Evolution 2017, **8,** 1374–1388 doi: 10.1111/2041-210X.12785

HyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA

Sarah Schmid[1,2], Rémi Genevest[1], Erika Gobet[2], Tomasz Suchan[1,3], Christoph Sperisen†[,4], Willy Tinner†[,2] and Nadir Alvarez*[,†,1]

[1]*Department of Ecology and Evolution, UNIL-Sorge, University of Lausanne, Biophore Building, 1015 Lausanne, Switzerland;* [2]*Institute of Plant Sciences and Oeschger Centre for Climate Change Research, University of Bern, Altenbergrain 21, CH-3013 Bern, Switzerland;* [3]*W. Szafer Institute of Botany, Polish Academy of Sciences, Lubicz 46, 31-512 Kraków, Poland; and* [4]*Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland*

Summary

1. Over the last decade, protocols aimed at reproducibly sequencing reduced-genome subsets in non-model organisms have been widely developed. Their use is, however, limited to DNA of relatively high molecular weight. During the last year, several methods exploiting hybridization capture using probes based on RAD-sequencing loci have circumvented this limitation and opened avenues to the study of samples characterized by degraded DNA, such as historical specimens.

2. Here, we present a major update to those methods, namely *hybridization capture from RAD-derived probes obtained from a reduced eXome template* (hyRAD-X), a technique applying RAD sequencing to messenger RNA from one or few fresh specimens to elaborate bench-top produced probes, that is, a reduced representation of the exome, further used to capture homologous DNA from a samples set. In contrast to previous hybridization capture methods, the reference catalogue on which reads are aligned does not rely on *de novo* assembly of anonymous RAD-sequencing loci, but on an assembled transcriptome obtained from RNAseq data, thus increasing the accuracy of loci definition and single-nucleotide polymorphisms (SNP) call, and targeting, specifically, expressed genes. Finally, the capture step of hyRAD-X relies on RNA probes, increasing stringency of hybridization, making it well suited for low-content DNA samples. As a proof of concept, we applied hyRAD-X to subfossil needles from the coniferous tree *Abies alba*, collected in lake sediments (Origlio, Switzerland) and dating back from 7200 to 5800 years before present (BP). More specifically, we investigated genetic variation before, during and after an anthropogenic perturbation that caused an abrupt decrease in *A. alba* population size, 6500–6200 years BP.

3. HyRAD-X produced a matrix encompassing 524 exome-derived SNPs. Despite a lower observed heterozygosity was found during the 6500–6200 years BP time slice, genetic composition was nearly identical before and after the perturbation, indicating that re-expansion of the population after the decline was most likely driven by local specimens.

4. To the best of our knowledge, this is the first time a population genomic study incorporating ancient DNA samples of tree subfossils is conducted at a moderate cost using reproducible exome-reduced complexity.

Key-words: *Abies alba*, ancient DNA, anthropogenic perturbations, exome reduction, genetic diversity, hybridization capture, population genomics, RAD sequencing, subfossils, time series

Introduction

The advent of next-generation sequencing technologies has made possible conducting genomic-scale studies in non-model organisms (Ellegren 2014). As sequencing full genomes is not practical in many cases, approaches of reduced representation genome sequencing emerged, which include genotyping by sequencing (GBS; Elshire *et al.* 2011) and restriction-site associated DNA sequencing (RADseq; Baird *et al.* 2008; Peterson *et al.* 2012). These methods rely on digesting DNA with specific restriction enzymes and performing size selection on the resulting fragments, yielding homologous and reproducible markers in a set of samples. However, the use of such methods in ecological, phylogenetic and phylogeographic studies is constrained by the level of preservation of the starting genomic DNA, therefore limiting its application in historical or ancient specimens showing typical DNA degradation patterns (Suchan *et al.* 2016). Since 2010, elegant alternatives to circumscribe this issue by using hybridization capture of non-anonymous DNA fragments (e.g. exome, mitochondrial

*Correspondence author. E-mail: nadir.alvarez@unil.ch
†These authors supervised this work together and are considered as joint senior authors.

genome) have been published (reviewed in Horn 2012; Burrell, Disotell & Bergey 2015; Orlando, Gilbert & Willerslev 2015). However, it is only last year that capture methods targeting GBS or RADseq loci arose. These methods use either bench-top produced (i.e. hyRAD; Suchan *et al*. 2016) or commercially synthesized hybridization probes corresponding to GBS or RADseq loci (Ali *et al*. 2016; Boucher *et al*. 2016; Hoffberg *et al*. 2016; Sánchez Barreiro *et al*. 2016). Those methods have largely increased the mean depth and matrix fullness of single-nucleotide polymorphisms (SNPs), as well as individual and SNP coverage of the studied samples, compared to classical GBS or RADseq approaches (Suchan *et al*. 2016). In addition, Sánchez Barreiro *et al*. (2016) and Suchan *et al*. (2016) have demonstrated the applicability of such methods to museum or herbarium samples, some of which collected more than a century ago.

While opening new avenues in population and ecological genomics targeting non-model organisms, most of those RADseq- or GBS-based capture methods are yet costly, as they require commercially synthesized probes (with the exception of Suchan *et al*. 2016). In addition, those methods are limited in their potential to yield orthologous loci, given that a single genome-based probe could capture several paralogous fragments from the same or different individuals, since hybridization capture may retrieve fragments with up to 5–15% local nucleotide divergence (Yeates *et al*. 2016). Because DNA regions present in multiple copies across the genome are more frequent in neutral than in expressed genes (i.e. the genome of metazoan and land plants contains 10–80% non-coding repeated elements; Metcalfe & Casane 2013), such an issue could be circumscribed by using the exome as the raw material for building the RADseq-based probes. In addition, the relatively ease of transcriptome sequencing (RNAseq; Wang, Gerstein & Snyder 2009) makes possible obtaining a high-quality reference to which aligning reads, a solution that may reveal both more powerful than a *de novo* assembly from captured fragments and much affordable economically than producing whole genome sequencing. Here, we present a method called *hybridization capture from RAD-derived probes obtained from a reduced eXome template* (hereafter, hyRAD-X), a technique derived from hyRAD, but performing single-digest RADseq on cDNA obtained from messenger RNA extracted from fresh samples, which serves as a template. This allows gathering a reproducible reduced representation library of the exome in a set of samples from the same species or from closely related species at an unprecedented low cost.

In addition, hyRAD-X implements an RNA conversion of the probes before hybridizing them to shotgun libraries, adapting a technique employed for capturing whole-genome sequences in human ancient DNA research (Carpenter *et al*. 2013). Compared to DNA probes, RNA probes not only eliminate the risk of contamination and chimera formation with the captured libraries – an issue producing erroneous SNP composition (Khaleel 2012) – but also forms DNA–RNA hybrid duplexes thermodynamically more stable than DNA–DNA duplexes (Gyi *et al*. 1998), enabling a more efficient capture step.

While population genomics involving ancient specimens has been applied to humans and several other large mammals (see Orlando, Gilbert & Willerslev 2015 for a review), such studies in non-mammal organisms are still scarce due to the lack of genome reduction techniques similar to RAD-seq, but applicable to low-content and highly degraded DNA. This may indicate why no studies so far have made use of the large through-time series of tree subfossil remains (e.g. leaves and needles), which have been caught in lake sediments across several millennia (e.g. Birks & Birks 2000; Birks 2003), to perform demo-genetic analyses using a large number of SNPs. In this study, we apply hyRAD-X to a samples set of subfossil needles of the silver fir tree *Abies alba*, collected from lake sediments in Origlio (Ticino, Switzerland) and dating back from 7200 to 5800 years before present (BP). More specifically, we illustrate the application of hyRAD-X to ancient DNA (aDNA) samples by comparing genetic diversity before, during and after an environmental anthropogenic perturbation recorded in the southern Alps, which took place between 6500 and 6200 years BP. This perturbation affected forest community composition due to higher rates of fire and grazing of cattle in the understorey following the establishment of the first agrarian societies in the area (Tinner *et al*. 1999; Wick & Möhl 2006). By applying hyRAD-X to a set of 48 samples spanning the period 7200–5800 years BP, we test whether such an anthropogenic perturbation affected within-species genetic diversity of the silver fir, *A. alba*, a co-dominant tree species in the Origlio area during this time frame. For evaluation purposes, we also applied classical hyRAD (Suchan *et al*. 2016) to a subset of silver fir samples from the same time interval in order to compare the resulting depth, coverage and SNP matrix fullness.

Materials and methods

SAMPLING SITE, CORING AND RADIOCARBON DATING

Lago di Origlio is a small lake (8 ha) situated at 416 m a.s.l. near Lugano in southern Switzerland, in the southern Prealps. In June 2014, two new parallel cores were taken with a UWITEC piston corer at the location of a previous coring site (Tinner *et al*. 1999) in the deepest part of the lake. The core diameter was 8 cm. After coring, the sediment cores were stored at 4 °C at the Institute of Plant Sciences, University of Bern. Given that the sediments consisted of discernible layers of silt and gyttja (brown organic-rich sediments), the old (coring 1993, Tinner *et al*. 1999) and new cores (2014) could be parallelized with a precision of c. 1 cm, corresponding to about 12 years for the period of interest (7200–5800 years ago). The published chronology of Origlio relies on accelerator mass spectrometry radiocarbon dating of terrestrial macrofossils that were converted to calibrated years before present (cal. BP) with the program Calib Version 3.03 (see Tinner *et al*. 1999). We checked the original chronology by recalibrating the radiocarbon ages (6 dates between 832 and 994 cm for the period 5200–7200 cal. BP, Fig. 1) with the new Calib Version 7.1. On average the new lowess-smoothed calibrated ages differ from the original ones that were used to build a depth-age model (Tinner *et al*. 1999) by 44 years, with the most extreme deviations being +111 and −79 years. Although this difference corresponds to the error that is inherent to radiocarbon dating for Holocene sequences (c. ±100 years; Tinner *et al*. 2003), we decided

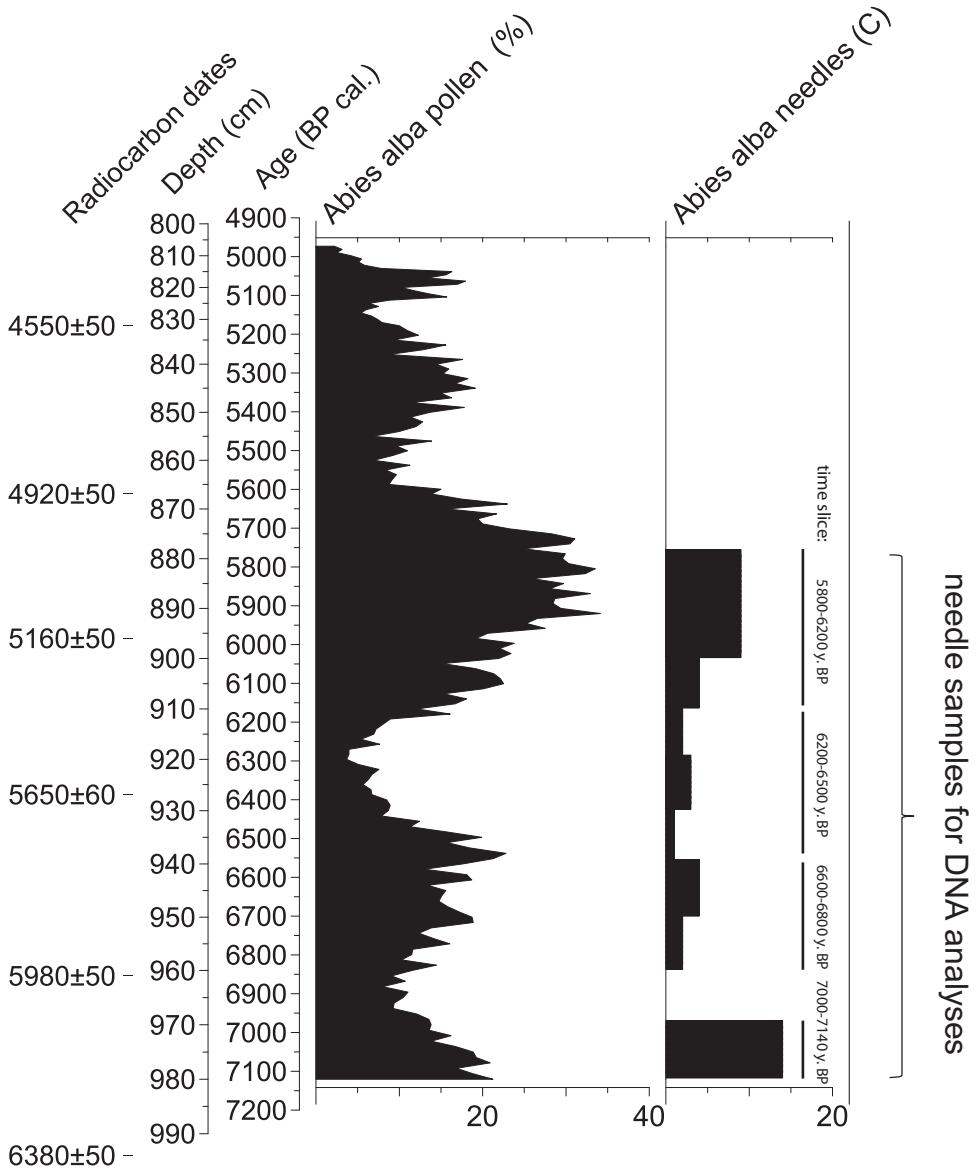

Fig. 1. *Abies alba* pollen percentage and needle concentration diagram of the Origlio lake. Percentages are given for a pollen sum that includes all terrestrial pollen types (for details see Tinner *et al.* 1999). Needle concentrations are in numbers per 100 cm^3, only intact and undamaged needles are shown. Vertical bars on the right of the graph show the delineations of the four time slices considered.

to present the paleobotanical *A. alba* data (pollen, needles) with the new chronology.

CORE SUBSAMPLING

The subsampling of the core was carried at the Institute of Plant Sciences, University of Bern, in a laboratory, which has never hosted any DNA extractions or subsequent DNA analyses. We cut the 1-m sediment core in 10-cm long and c. 4-cm wide slices (c. 100 cm^3) and removed c. 1 cm around the slice to avoid any kind of contemporary and past contamination (e.g. macrofossil down core transport during coring). We sieved each slice with a 200-μm mesh to retrieve subfossil needles and assessed species identification to *A. alba* under a binocular. A total of 48 undamaged subfossil needles from the time interval ranging between 7200 and 5800 BP were found in the core (Fig. 1), and all were stored in separate Eppendorf tubes at −18 °C until use. Due to the small coring diameter compared to the entire lake surface, we considered the probability of subsampling plant macrofossils from the same individual as negligible. All material employed during the samples preparation was sterilized by autoclaving (10 min, 30 psi and 132 °C) before use. Eppendorf tubes were further treated under UV light for 30 min. To control for potential modern aerosol contaminations, an empty Eppendorf tube was left open all along the procedure.

DNA EXTRACTION

All ancient DNA extractions and library preparations were performed in a dedicated clean air laboratory at the Swiss Federal Institute for Forest, Snow and Landscape Research WSL (Birmensdorf, Switzerland), which is conformed to the highest standards in the field.

Before DNA extraction, we wiped the needles using a foamed sterile tip humidified with molecular grade water under a binocular. Then, the needles were rinsed in molecular grade water, dipped in ethanol and

placed on a sterile Petri dish to dry. Following drying, the needles were exposed to UV light for 1 min on each side and finally placed in new tubes prefilled with molecular grade 0·1 mm zirconium beads (Benchmark Scientific Inc., Edison, NJ, USA). Subsequently, we immersed each tube in liquid nitrogen for 1 min, grinded the frozen needles with a disposable pestle (cleaned in 10% bleach for 10 min, rinsed in molecular grade water, followed by UV irradiation for 30 min) and repeated these steps until the needles were turned into powder. The resulting powder was supplemented with 1·2 ml of *N*-phenacylthiazolium bromide (PTB) extraction buffer – a buffer successfully used in various ancient plant tissues extractions (Jaenicke-Despres *et al*. 2003; Erickson *et al*. 2005) – and the mix was incubated for 18 h at 37 °C (Kistler 2012) on a tube rotator. Then, we centrifuged the mixture at 9000×*g* for 5 min, transferred the supernatant into a new tube, added 0·325 volumes of DNeasy Plant Mini Kit P3 buffer (Qiagen, Hilden, Germany) and incubated it on ice for 5 min. We completed the extraction by following the DNeasy Plant Mini Kit protocol provided by the manufacturer from step 10 (Qiagen), except that the DNA was purified on MinElute columns (Qiagen) and eluted in 20 µl of 10 mM Tris-HCl pH 8.5. For each series of six samples extracted in a row, we prepared one negative extraction control, which followed the same procedure as the above-mentioned extraction, minus the subfossil sample.

SHOTGUN LIBRARY PREPARATION

We prepared shotgun libraries for the subfossil DNA extractions following Suchan *et al*. (2016), whose protocol is optimized for degraded DNA samples. This method consists in the ligation of barcoded adapters to the double-stranded DNA, followed by a 25 cycles indexing PCR using the Q5 High-Fidelity DNA Polymerase (NEB, New England Biolabs, Ipswich, MA, USA). Here, we applied a dual indexing method (i.e. barcode + index) in a way that both barcode and index were specifically tagging a given specimen (not a combination of barcode × index) in order to detect possible chimeras during library preparation. The resulting products were purified with AMPure XP beads (ratio 1 : 1 with the sample; Beckman Coulter, Nyon, Switzerland) and eluted in 11 µl of 10 mM Tris-HCl pH 8.5 to increase DNA concentration, although larger volumes are recommended to reach an optimal DNA recovery (Beckman Coulter, 2013) – an alternative would be to elute in larger volume and further re-concentrate with SpeedVac. Finally, the success of the library preparation was assessed by Fragment Analyzer (Advanced Analytical, Heidelberg, Germany) and quantified on a Qubit Fluorometer (Thermo Fisher Scientific, Ecublens, Switzerland).

GENERATION OF RNA PROBES

RNA extraction

A 3-years-old *A. alba* seedling grown in a nursery bed at the WSL (Birmensdorf, Switzerland) was brought into the laboratory and 10 buds were directly cut into liquid nitrogen and subsequently grinded. Total RNA was extracted with the Agilent Plant RNA Isolation Mini Kit (Agilent Technologies, Palo Alto, CA, USA) according to the manufacturer's protocol with the following modifications: washes of the silica columns were done in 450 µl with the final wash centrifuged for 2 min to dry the silica membrane. The RNA was eluted in 30 µl of DNase-free water, quantified using Nanodrop (Thermo Fisher Scientific) and its integrity checked on a Bioanalyzer 2100 (Agilent Technologies) using the Agilent RNA 6000 Pico Kit (Agilent Technologies). The detailed protocol for RNA extraction is given in Appendix S1, Supporting Information.

Double-stranded cDNA synthesis

The extracted total RNA was used as a template for the double-stranded cDNA synthesis using the Maxima H Minus Double-Stranded cDNA Synthesis Kit (Thermo Fisher Scientific) following the manufacturer's guidelines, applying oligo$(dT)_{18}$ primer for the first strand cDNA synthesis, and therefore reverse-transcribing only messenger RNA. Purification of cDNA was performed using the GeneJET PCR Purification Kit (Thermo Fisher Scientific) and the final cDNA was eluted in 25 µl of 10 mM Tris-HCl, pH 8·5 with gradual centrifugation.

Adaptor ligation and amplification

To ligate adaptors to the generated cDNA, we first applied digestion using the following reaction mix: 6 µl of cDNA (c. 20 ng), 0·9 µl CutSmart buffer (10×; NEB, New England Biolabs), 0·1 µl MseI (10 U μl^{-1}; NEB) and 2 µl H_2O. The mix was incubated at 37 °C for 3 h followed by heat inactivation of MseI at 65 °C during 20 min. Nine microlitres of the resulting digested DNA were mixed with 0·26 µl CutSmart buffer (10×; NEB), 0·12 µl ATP (100 mM), 0·06 µl H_2O, 2 µl T7 adapters mix (25 µM) and 0·17 µl T4 DNA ligase (400 U μl^{-1}; NEB). The reaction was incubated for 3 h at 16 °C, purified with AMPure XP beads (ratio 2 : 1 with the sample; Beckman Coulter) and eluted in 31 µl of 10 mM Tris-HCl pH 8·5. The generated library was PCR amplified in 10 separate reactions to minimise stochastic PCR-induced over-representation of some fragments, using the following reagents: 2·15 µl H_2O, 2 µl Q5 buffer (5×; NEB), 0·67 µl PCR primer (corresponds to T7 oligo 1, 10 µM stock), 0·08 dNTPs (25 mM each), 0·1 µl Q5 Hot Start High-Fidelity DNA Polymerase (2U μl^{-1}; NEB), 2 µl Q5 High GC Enhancer (5×; NEB) and 3 µl of the purified ligation product. The mix was subjected to the following cycle: 98 °C for 30 s; 30 cycles of 98 °C for 20 s, 60 °C for 30 s, 72 °C for 40 s; 72 °C for 10 min. Subsequently, 1 µl of the following mix was added to each PCR reaction to proceed to the final PCR cycle: 0·05 µl H_2O, 0·2 µl Q5 buffer (5×; NEB), 0·67 µl PCR primer (corresponds to T7 oligo 1, 10 µM stock), 0·08 µl dNTPs (25 mM each). The reactions were successively incubated at 98 °C for 3 min, 60 °C for 2 min and 72 °C for 12 min. All PCR products were pooled together, purified with AMPure XP beads (ratio 2 : 1 with the sample; Beckman Coulter) and eluted with 100 µl of 10 mM Tris-HCl pH 8·5.

In vitro transcription and biotinylation

To transcribe the generated probes into biotinylated RNA, we followed Carpenter *et al*. (2013), using the HiScribe T7 High-Yield RNA Synthesis Kit (NEB) and assembled the following reaction: 3·1 µl nuclease-free water, 1·5 µl reaction buffer (10×), 1·5 µl of each dNTP (100 mM), 5 µl biotin-UTP (10 mM; Roche, Mannheim, Germany), 4·4 µl amplified probes (c. 500 ng) and 1·5 µl T7 RNA polymerase mix. The reaction was incubated at 37 °C for 2 h, treated with 1 µl of TURBO DNase (2 U μl^{-1}; Thermo Fisher Scientific) during 15 min at 37 °C, purified with RNeasy Mini Kit (Qiagen) following the manufacturer's guidelines and eluted twice in the same 30 µl of RNase-free water. To prevent RNA degradation, 1·5 µl of RNasin Plus RNase Inhibitor (40 U μl^{-1}; Promega, Dübendorf, Switzerland) was added to the biotinylated RNA probes and the mix was stored at −80 °C. One reaction generated about 12 µg of RNA, whose size, ranging from

100 bp to 500 bp, was checked on a Bioanalyzer 2100 (Agilent Technologies).

RNA adaptor-blocking oligonucleotides preparation

The generated aDNA libraries contained indexed multiplex adapters added during the PCR step (see Shotgun library preparation). To avoid unspecific binding during capture, we produced adaptor-blocking RNA oligonucleotides as described by Carpenter *et al.* (2013). We annealed T7 universal promoter (5′-AGTACTAATACGACT CACT ATAGG-3′) and multiplex block P5 (5′-AGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTA TCATTCCTATAGTGAGTCGTATTAGTACT-3′) by mixing 12·5 µl of each oligo stock (200 mM) with 5 ml of 10× buffer 2 (NEB) and 20 ml of water. We incubated the mix at 95 °C for 5 min, followed by 1 h at room temperature. The same was done with multiplex block P7 (5′-AGATCGGAAGAGCACACGTCTGAACTCCAGTCACN NNNNNATCTCGTATGCCGTCTTCTGCTTGCCTATAGTGA GTCGTATTAGTACT-3′), which contains random nucleotides at the level of the index sequence, allowing the use of the same aliquot of adaptor-blocking oligonucleotides for all libraries. We proceeded to the *in vitro* transcription of 700 ng of each double-stranded oligonucleotide solution using the T7 High-Yield RNA Synthesis Kit (NEB) following the manufacturer's guidelines. We mixed the resulting product with 1 ml of TURBO DNase (2 U μl^{-1}; Ambion Europe, Huntingdon, UK) and incubated it at 37 °C for 15 min. Then, we purified the RNA using the RNeasy Mini kit (Qiagen) according to the manufacturer's guidelines, but added 675 µl of ethanol at step 2 instead of 250 µl to recover also small RNA molecules. Finally, we eluted the RNA in 30 µl water and added 1·5 µl of SUPERase-In RNase inhibitor (20 U μl^{-1}; Ambion) for long-term storage at −80 °C.

ANCIENT DNA HYBRIDIZATION CAPTURE AND ENRICHMENT WITH RNA PROBES

The hyRAD-X hybridization capture method is adapted from both the hyRAD protocol (Suchan *et al.* 2016) and the in-solution capture protocol published by Carpenter *et al.* (2013).

In-solution hybridization with RNA probes

For the in-solution hybridization, three mixes were prepared. (i) The hybridization buffer (10× SSPE, 10× Denhardt's, 10 mM EDTA, 0·2% SDS and 0·01% Tween 20), prewarmed at 65 °C before use. (ii) The ancient DNA pond consisted in 22 µl of ancient DNA library (500–1000 ng) combined to 10 µl of SeqCap EZ Developer Reagent (Roche) – a universal reagent to block repetitive regions in the genome – and was incubated at 95 °C for 5 min, followed by 65 °C for 5 min. (iii) To generate the RNA probes pond, we assembled 1 µl biotinylated RNA probes (500 ng), 3 µl RNasin Plus RNase Inhibitor (40 U μl^{-1}; Promega), 2 µl P5 multiplex block RNA (100 µM) and 2 µl P7 multiplex block RNA (100 µM), prewarmed at 65 °C before use for 2·5 min. The ancient DNA pond was incubated at 95 °C for 5 min, followed by 65 °C for 5 min. After incubation, 26 µl prewarmed hybridization buffer was added to the 32 µl ancient DNA pond, followed by 8 µl of prewarmed RNA probes pond. The final 66 µl reaction was mixed by pipetting and incubated for 60 h at 65 °C. For each reaction, 50 µl of Dynabeads M-280 Streptavidin beads (Thermo Fisher Scientific) were magnetized and the supernatant was removed. Beads were resuspended in 200 µl of wash buffer (1 M NaCl, 10 mM Tris-HCl pH 7·5, 1 mM EDTA and 0·01% Tween 20), vortexed for 30 s, incubated 2 min on the magnetic rack and the supernatant was removed. This wash was performed two more times and the beads were ultimately resuspended in 164 µl of wash buffer. Then, 66 µl of the DNA/RNA hybrid mix was added to the 164 µl of washed streptavidin beads, vortexed 10 s and incubated 30 min at room temperature with occasional vortexing. The reaction was magnetized, the supernatant removed and the beads were resuspended in 165 µl of low-stringency buffer (1× SSC/0·1% SDS). After 15 min of incubation at room temperature, the reaction was magnetized, the supernatant removed, the beads were resuspended in 165 µl of prewarmed high-stringency buffer (0·1× SSC/0·1% SDS) and the reaction was incubated 10 min at 65 °C. This step was repeated two times more. Next, the solution was magnetized, the supernatant removed and 50 µl of NaOH (0·1 M) was added, followed by 10 min of incubation at room temperature. Finally, the supernatant was retained after magnetization and the solution was neutralized with 50 µl of Tris-HCl (1 M, pH 7·5). The captured DNA was concentrated with AMPure XP beads (ratio 1·8 : 1 with the sample; Beckman Coulter) and eluted in 30 µl H_2O.

Libraries enrichment

The 30 µl of captured libraries were divided into two replicates and each 15 µl replicate was mixed with the following reagents for PCR amplification: 18·6 µl H_2O, 10 µl Q5 reaction buffer (5×; NEB), 0·4 µl dNTPs (25 mM each), 0·5 µl Q5 Hot Start High-Fidelity DNA Polymerase (2U μl^{-1}; NEB), 0·5 µl RNase A (100 mg ml^{-1}; Qiagen), 5 µl IS5-IS6 primer mix (5 µM each; Meyer & Kircher 2010). The reactions were subject to the following cycling condition: 98 °C for 30 s; 25 cycles of 98 °C for 10 s, 60 °C for 20 s, 72 °C for 25 s; 72 °C for 2 min. The enriched libraries were purified with AMPure XP beads (ratio 1 : 1 with the sample; Beckman Coulter) and eluted in 30 µl of 10 mM Tris-HCl pH 8·5.

A schematized view of the whole hyRAD-X protocol is given in Fig. 2.

HYBRIDIZATION CAPTURE BASED ON DNA PROBES

To examine the performance of hyRAD-X, we also captured a subset of the generated shotgun libraries using hyRAD-based DNA probes (total of 22 samples and 8 blank samples; Table S1), applying the protocol of Suchan *et al.* (2016) aiming at sequencing a representative fraction of the genome in a set of samples by generating a double-digest RADseq (vs. single-digest RADseq in hyRAD-X) genome-based library (vs. exome-based in hyRAD-X) that is used as DNA probes (vs. RNA probes in hyRAD-X) for hybridization capture of the shotgun libraries. ddRAD was performed on six fresh samples coming from the *A. alba* population in Derborence (Wallis, Switzerland). Then, we proceeded to a narrow size selection with a single peak at 270 bp, removed adaptor sequences and biotinylated the fragments to produce the final DNA probes. We employed the resulting probes to capture the shotgun libraries by hybridization. The hybridization lasted 66 h at 65 °C and repetitive regions were blocked with 10 µl of plant capture enhancer (PCE; Roche). The captured sequences were subsequently re-amplified. Detailed information on the protocol can be found in Suchan *et al.* (2016).

SEQUENCING

All captured aDNA libraries were sequenced on three lanes of a 100 paired-end Illumina HiSeq 2500 run, two lanes for samples treated with hyRAD-X and one lane for those processed with hyRAD. The subset

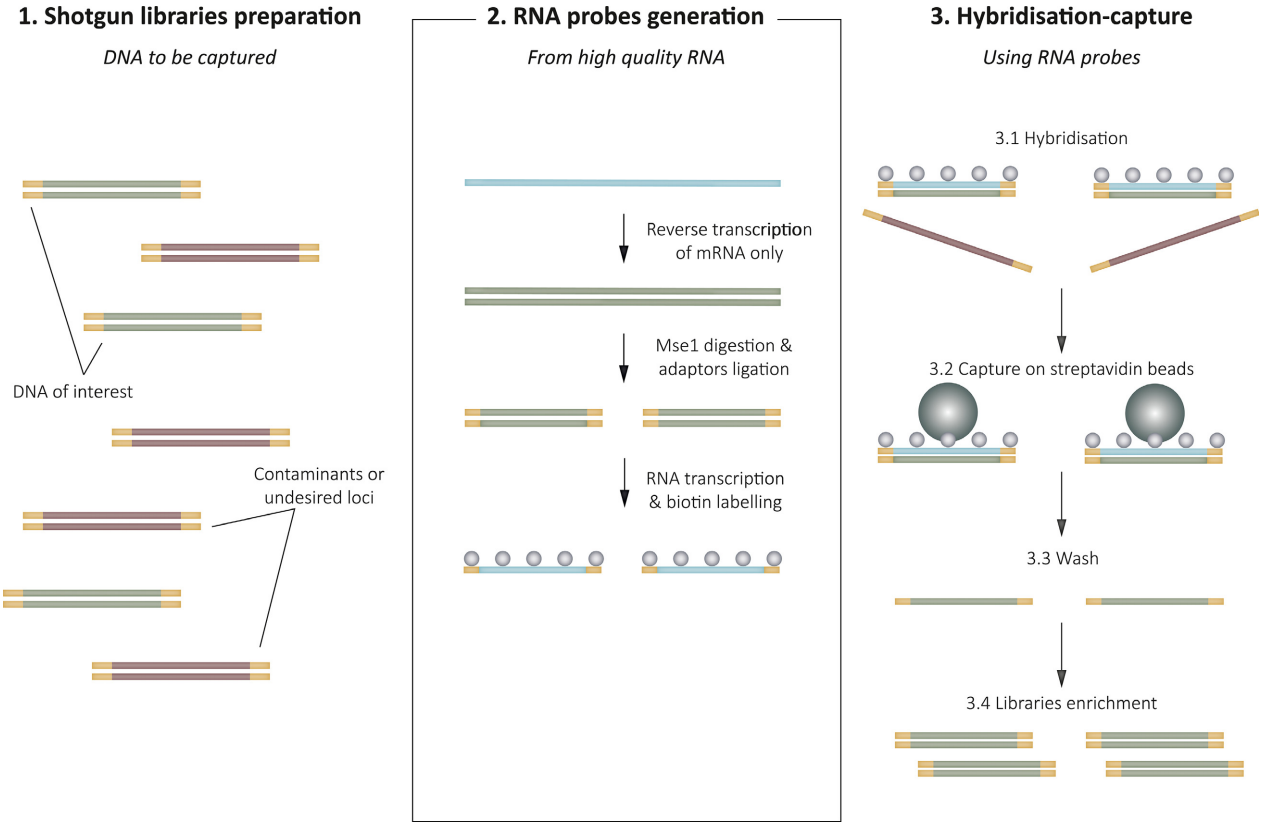

Fig. 2. HyRAD-X laboratory procedure. The pipeline consists in three steps: shotgun library generation, probes generation and hybridization. The main difference with hyRAD (Suchan *et al.* 2016) resides in the type of probes used for the capture.

of samples treated with HyRAD was also sequenced at the stage of raw shotgun library preparation – for comparison purposes – on one lane of a 100 paired-end Illumina HiSeq 2500 Rapid Run. The DNA probes derived from modern DNA were sequenced on one lane of Illumina MiSeq 250 paired end. Sequencing was carried out at the Lausanne Genomic Technologies Facility, University of Lausanne, Switzerland and at FASTERIS SA, NGS services, Switzerland.

ILLUMINA DATA BIOINFORMATICS PROCESSING

Demultiplexing and data preparation

We first performed demultiplexing and cleaning of the reads generated by Illumina HiSeq and MiSeq using a custom Perl script based on fastx-multx and fastq-clean tools from the *ea-utilis* package (Aronesty 2011; Mastretta-Yanes *et al.* 2015). Then, we removed leading Gs and tailing Cs generated during ribotailing reaction of the shotgun, paired the reads that still have a mate and combined orphan reads in one file using custom scripts.

Read mapping

Prior to read mapping, we aligned the processed reads against the *Picea abies* genome (Nystedt *et al.* 2013) with Bowtie 2 (Langmead & Salzberg 2012) to estimate the amount of endogenous DNA.

Read mapping was performed using the PALEOMIX BAM pipeline v1.2.5, which is dedicated to the analysis of ancient genomes (Schubert *et al.* 2014). In brief, the mentioned pipeline trims adapter sequences of demultiplexed reads, filters low-quality reads and collapses overlapping paired-end reads into one longer sequence. Then, the processed reads are mapped to the reference (see below), *post-mortem* DNA damages are assessed to rescale the quality scores of likely damaged positions in the reads and PCR duplicates are filtered. For the hyRAD-X and hyRAD libraries, we used as reference the contigs (Genbank accessions JV134525-JV157085) generated from the sequencing and assembly of a normalized transcriptome of a 1-year-old *A. alba* seedling (Roschanski *et al.* 2013). Additionally, for the hyRAD libraries, we implemented the mapping strategy used in the initial hyRAD protocol (Suchan *et al.* 2016) and assembled *de novo* the demultiplexed RAD-generated DNA probes into contigs using Rainbow v2.0.4 (default settings; Chong, Ruan & Wu 2012).

SNP calling and filtering

We applied SNP calling using FreeBayes v1.0.2-29, a Bayesian genetic variant detector dedicated to short-read sequencing (Garrison & Marth 2012). Subsequently, we filtered SNPs to conserve only high-quality and informative positions. Loci with a Phred quality score below 30 and indels were removed. Then, only biallelic loci with a minor allele count of six (i.e. minor allele frequency of 0·12), present in at least 50% of the samples and with a minimum depth of six were kept (VCFtools; Danecek *et al.* 2011). High coverage can lead to inflated locus quality score. To overcome such a bias, loci with a quality score below 1/4 of the depth were excluded (Puritz, Hollenbeck & Gold 2014). Finally, potential paralogous loci were removed by applying part of the dDocent filtering pipeline, which detects paralogs based on a coverage three standard deviations higher than the mean (Puritz, Hollenbeck & Gold 2014). Samples with more than 90% of missing data were removed from the final dataset.

Table 1. Sampling of *Abies alba*. Given are the coring depth, the estimated age and the number of needles found (*n*) for each 10-cm core slice used in the study

Depth (cm)	Age (year BP)	*n*
879–889	5800–5940	9
890–899	5940–6060	9
900–909	6060–6200	4
910–919	6200–6320	2
920–929	6320–6460	3
930–939	6460–6600	1
940–949	6600–6720	4
950–959	6720–6880	2
960–969	6880–7000	0
970–979	7000–7140	14

DATA ANALYSIS

Sequence composition of raw shotgun libraries and blanks

We estimated the amount of endogenous DNA in the raw shotgun libraries by aligning the reads to the genome of *P. abies* (Nystedt *et al.* 2013) using Bowtie 2 (Langmead & Salzberg 2012). To identify captured contaminant sequences, a BLAST alignment (Altschul *et al.* 1990) was run on random subsets of 20 000 reads from hyRAD and hyRAD-X blanks.

Identification of nucleotide misincorporations typical from ancient DNA

The Mapdamage 2.0 program (Jónsson *et al.* 2013) as implemented in the PALEOMIX pipeline was used to retrieve patterns of DNA misincorporation, typical of ancient DNA fragments. In particular, cytosine deamination is a *post-mortem* process that converts cytosines (C) into uracil – therefore retrieved as thymine (T) using DNA polymerases and on the complementary strand converting guanidine (G) into adenine (A) (Hofreiter *et al.* 2001). Such deamination processes occur more frequently at both 3′ and 5′ ends than in the centre of DNA fragments (Jónsson *et al.* 2013).

Comparison between RNA and DNA probes

To investigate how hyRAD-X performed compared to hyRAD, we determined the mean read depth per sample and per site, as well as the percentage of missing data per sample using VCFtools v0.1.14 (Danecek *et al.* 2011). Furthermore, we also compared the percentage of reads mapping on the reference and the percentage of PCR duplicates using PALEOMIX v1.2.5 (Schubert *et al.* 2014).

Genetic analysis based on hyRAD-X

We partitioned the data into four time slices (7200–7000 years BP, 6800–6500 years BP, 6500–6200 years BP, 6200–5800 years BP), according to the population intervals previously inferred by pollen analysis (Table 1; Tinner *et al.* 1999). While the 6500–6200 years BP time interval corresponds to a major decline in *A. alba* populations from the Southern Alps (including Origlio), the three other time slices show larger population sizes (Tinner *et al.* 1999). To compare genetic diversity indices among those time slices, SNPs with a fraction of missing data higher than 60% within each time slice were removed. We calculated observed heterozygosity (H_{obs}), expected heterozygosity (H_{exp}, also known as genetic diversity; Goudet 2005) and rarefied allelic richness (A_R) using HIERFSTAT (Goudet 2005). We tested for pairwise differences in mean H_{obs}, H_{exp} and A_R among each time slice by performing 10 000 permutations and we corrected for multiple comparisons by computing the expected false discovery rate (FDR), or *q*-value (Benjamini & Hochberg 1995). We estimated the number of private alleles before and after the 6500–6200 BP time slice by applying a rarefaction approach with ADZE V1.0 (Szpiech, Jakobsson & Rosenberg 2008). Finally, we compared genetic variation between time slices using the Weir and Cockerham fixation index (F_{ST}) to estimate population differentiation while controlling for population size using ADEGENET (Jombart & Ahmed 2011). We tested pairwise differences in F_{ST} with 10 000 permutations and corrected for multiple testing by calculating the expected FDR. We performed all statistical analyses in R V3.3.0 (R Core Team 2016).

Results

Sequencing of raw shotgun libraries yielded a total of 300 566 668 reads. Alignment of raw shotgun reads to the *P. abies* genome (Nystedt *et al.* 2013) showed that the estimated amount of endogenous DNA in the raw samples ranged from 0·01 to 0·33% (Table S1). HyRAD-X libraries based on 48 samples and 16 blank controls – eight extraction controls, four shotgun library controls and four subsampling controls [two water controls (i.e. 100 µl of water used for core slice sieving) and two aerosol controls (i.e. an Eppendorf tube that was left open during the subsampling step)] – yielded a total of 320 386 174 raw reads (c. 8 million reads per non-blank sample), of which 85·3% were retained after quality filtering. HyRAD libraries – which consisted in 24 samples and eight blank controls – generated 302 607 744 raw reads in total (c. 12·6 million reads per non-blank sample), among which 49·3% were kept after quality filtering. The overall level of chimeric fragments produced during shotgun library preparation (i.e. identified by reads characterized by a mismatch in the barcode + index combination) was 3·17% – all were discarded, that is, only reads characterized by the correct barcode + index composition were kept. In contrast to the low level of endogenous DNA retrieved from shotgun sequencing libraries, enriched hyRAD-X and hyRAD libraries showed a high proportion of reads aligning to the *P. abies* genome [46·9-fold average enrichment compared to the raw (pre-capture) shotgun library; see Table S1]. For the hyRAD-X library, 39 out of 48 samples yielded more reads than the 16 blanks. For the hyRAD library, 21 out of 24 samples yielded more reads than the eight blanks. However, after exogenous contaminant and PCR duplicates removal, blank controls contained an order of magnitude less reads than the worse performing sample (data not shown). In >90% of the cases, deamination patterns were typical of ancient DNA for both hyRAD-X and hyRAD libraries. In contrast, blank controls did not show such a pattern (Fig. S1). Blanks were mainly composed of synthetic sequences in both hyRAD-X and hyRAD libraries (Fig. S2).

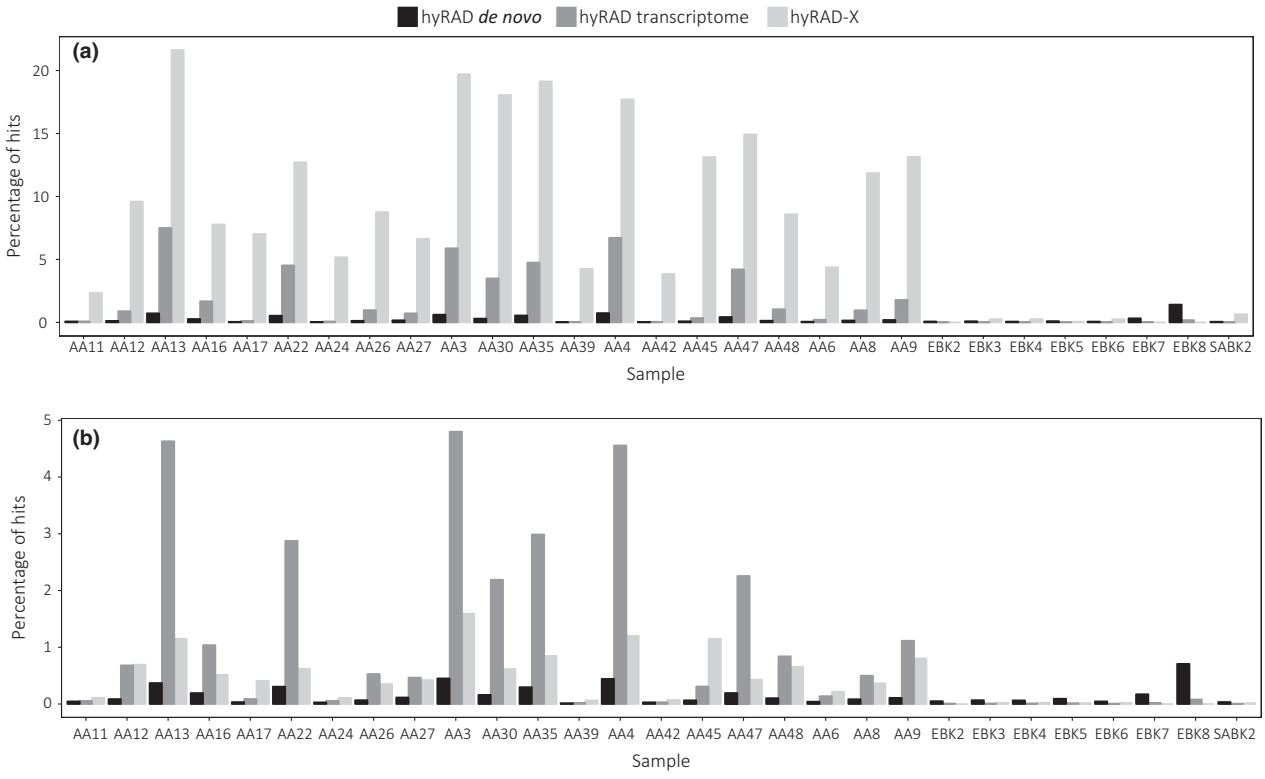

Fig. 3. Percentage of reads (hits) mapping against the reference genome including PCR duplicates (a) and excluding PCR duplicates (b) for a subset of 27 samples and blank controls. Black is for hyRAD reads mapped against the assembled RAD contigs (i.e. hyRAD *de novo*), dark grey is for hyRAD reads mapped against the *Abies alba* transcriptome (i.e. hyRAD transcriptome) and light grey represents hyRAD-X reads mapped against the *A. alba* transcriptome. EBK samples are extraction blank controls and SABK2 is a control for aerosols absence in the air during subsampling.

Starting DNA concentration as well as the age of the sample were not good predictors of whether a library would contain less than 90% of missing data when processed with hyRAD-X ($\chi^2 = 0{\cdot}63$, d.f. = 1, $P = 0{\cdot}43$ and $\chi^2 = 1{\cdot}58$, d.f. = 1, $P = 0{\cdot}21$ respectively). However, the number of raw (i.e. pre-capture) sequences matching against the *P. abies* genome (Table S1) is significantly correlated to whether or not a library would contain less or more than 90% of missing data ($\chi^2 = 5{\cdot}81$, d.f. = 1, $P = 0{\cdot}01$).

Comparison between hyRAD-X and hyRAD

After SNP calling, we retrieved 42 095, 43 381 and 4435 variable sites (i.e. indels and SNPs), respectively, for hyRAD-X, hyRAD aligned against the transcriptome (hereafter, hyRAD transcriptome) and hyRAD aligned against the *de novo* reference (hereafter, hyRAD *de novo*). Following filtering and removal of samples with more than 90% of missing data (see Materials and methods), we retained 571 SNPs in 27 (67%) samples with hyRAD-X, 1738 SNPs in 13 (54%) samples with hyRAD transcriptome and 97 SNPs in 14 (58%) samples with hyRAD *de novo*. The matrix completeness was 73·1% for hyRAD-X, 77·4% for hyRAD transcriptome and 82·4% for hyRAD *de novo*.

Before removal of PCR duplicates, the percentage of reads mapping against the reference was up to 20 times higher when using hyRAD-X (with the reference consisting of a published transcriptome) compared to hyRAD *de novo* (whose reference was *de novo* generated using a dedicated RADseq assembler) and up to three times higher compared to hyRAD transcriptome (using the same reference as hyRAD-X) (Fig. 3a). However, after PCR duplicates removal, the percentage of reads mapping against the reference was higher for hyRAD transcriptome compared to hyRAD-X and hyRAD *de novo* (Fig. 3b). The percentage of PCR duplicates was between two to eight times higher when the capture was based on hyRAD-X (Table S1). No major difference was highlighted in the percentage of missing data and the mean read depth per sample between hyRAD-X, hyRAD transcriptome and hyRAD *de novo* after removal of PCR duplicates (Fig. 4a and b). However, mean read depth per site was higher with hyRAD-X, with a mean depth across all sites of 37·16, compared to 23·44 with hyRAD *de novo* and 15·7 with hyRAD transcriptome (Fig. 4c).

Change in genetic diversity over time and population structure

Due to the lower number of individuals analysed with hyRAD vs. hyRAD-X protocols – the latter allowing a larger multiplex pool due to higher level of genomic (here, exomic) reduction – genetic dynamics was inferred using the dataset produced with hyRAD-X (see Table 2 for details on analysed samples). SNP positions with a fraction of missing data higher than 60% within each time slice were removed, resulting in a final dataset

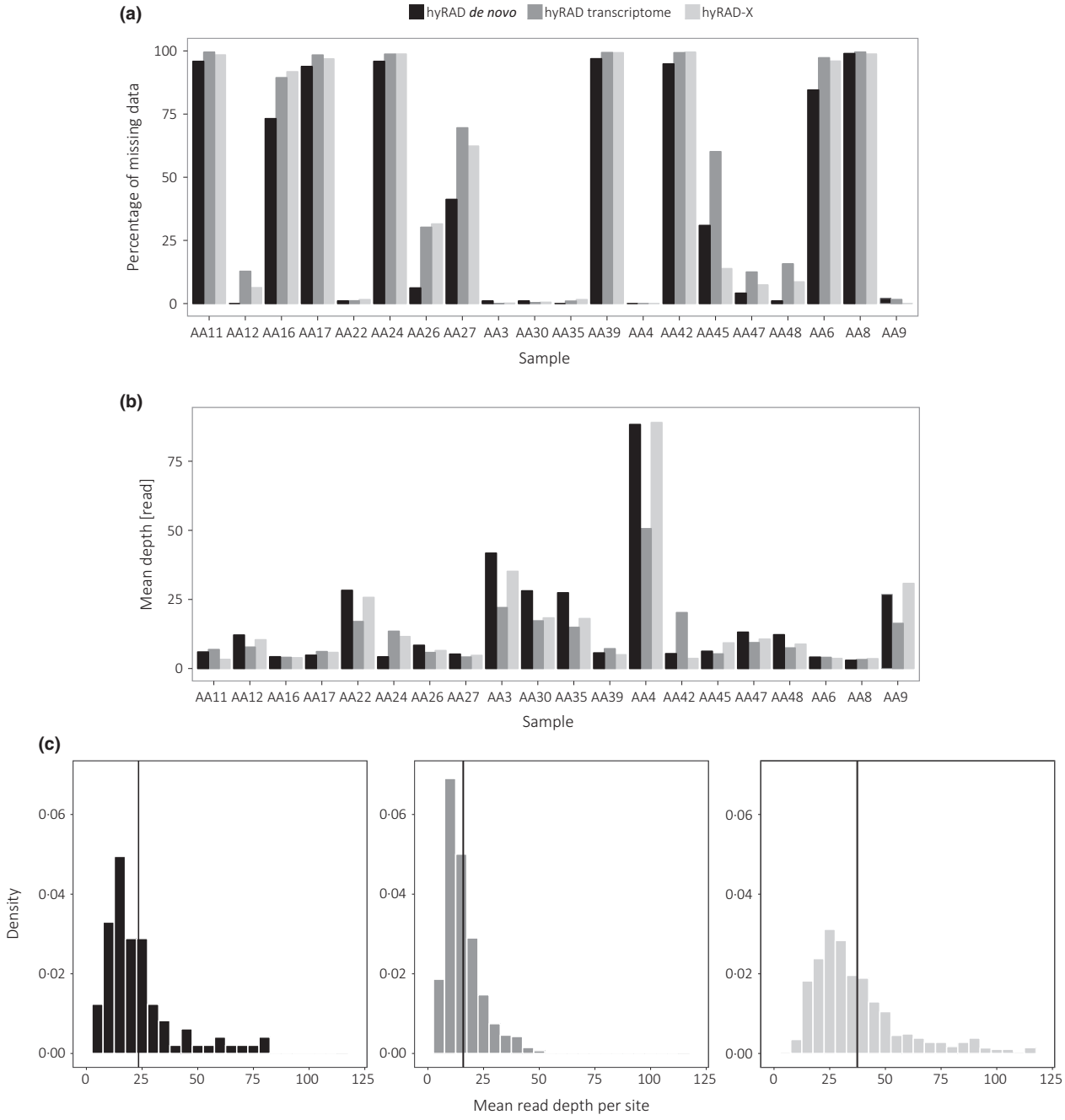

Fig. 4. Depth and missing data comparison between hyRAD and hyRAD-X in a subset of 24 samples. Black is for hyRAD reads mapped against the assembled RAD contigs (i.e. hyRAD *de novo*), dark grey is for hyRAD reads mapped against the *Abies alba* transcriptome (i.e. hyRAD transcriptome) and light grey represents hyRAD-X reads mapped against the *A. alba* transcriptome. (a) Percentage of missing data per sample after SNP filtering. (b) Mean read depth per sample after SNP filtering. (c) Distribution of mean read depth per site after SNP filtering. The vertical black line represents the read depth averaged across all sites.

of 524 SNPs in 27 samples. Expected heterozygosity was higher than observed heterozygosity for each time slice (Table 3). Observed and expected heterozygosities were significantly lower for the 6500–6200 BP time slice compared to all others and allelic richness was significantly higher for the 6800–6500 BP time slice ($P < 0{\cdot}05$ after FDR correction; Table 3). Mean number of private alleles per site was higher before than after the population decline, that is, between 6800 and 6500 years BP compared to 5800 and 6200 years BP (Fig. 5). Finally, pairwise F_{ST} estimates were significant between the 6500–6200 BP and 6200–5800 BP time slices, and marginally non-significant between the 6800–6600 BP and 6500–6200 BP time slices (Table 4).

Discussion

Hybridization capture methods involving anonymous probes obtained from reduced representation of the genome

Table 2. List of non-blank samples analysed using hyRAD-X. Samples in bold contained <90% of missing data and were used in the genetic analysis. Given are the concentrations after shotgun library generation and the time slice considered for pooling successful samples for genetic analyses

Sample	Age (years BP)	Conc. (ng per μl)	Time slice (years BP)
AA48	**7000–7140**	**20·18**	**7000–7140**
AA47	**7000–7140**	**N/A**	**7000–7140**
AA46	7000–7140	14·85	–
AA45	**7000–7140**	**36·94**	**7000–7140**
AA44	7000–7140	15·56	–
AA43	7000–7140	104·48	–
AA42	7000–7140	35·40	–
AA41	**7000–7140**	**78·66**	**7000–7140**
AA40	**7000–7140**	**40·36**	**7000–7140**
AA39	7000–7140	36·94	–
AA38	**7000–7140**	**21·89**	**7000–7140**
AA37	7000–7140	17·78	–
AA36	7000–7140	11·12	–
AA35	**7000–7140**	**28·39**	**7000–7140**
AA34	6720–6880	15·30	–
AA33	6720–6880	23·09	–
AA32	**6600–6720**	**46·85**	**6600–6800**
AA31	**6600–6720**	**31·46**	**6600–6800**
AA30	**6600–6720**	**21·20**	**6600–6800**
AA29	6600–6720	31·29	–
AA28	**6460–6600**	**38·30**	**6200–6500**
AA27	**6320–6460**	**8·38**	**6200–6500**
AA26	**6320–6460**	**26·33**	**6200–6500**
AA25	**6320–6460**	**56·43**	**6200–6500**
AA24	6200–6320	16·93	–
AA23	**6200–6320**	**42·41**	**6200–6500**
AA22	**6060–6200**	**17·44**	**5800–6200**
AA21	**6060–6200**	**35·06**	**5800–6200**
AA20	6060–6200	75·58	–
AA19	6060–6200	9·00	–
AA18	**5940–6060**	**27·02**	**5800–6200**
AA17	5960–6060	17·96	–
AA16	5960–6060	7·87	–
AA15	**5940–6060**	**57·80**	**5800–6200**
AA14	5940–6060	33·69	–
AA13	**5940–6060**	**14·02**	**5800–6200**
AA12	**5940–6060**	**19·32**	**5800–6200**
AA11	5940–6060	1·03	–
AA10	5940–6060	0·68	–
AA9	**5800–5940**	**38·30**	**5800–6200**
AA8	5800–5940	21·03	–
AA7	**5800–5940**	**18·30**	**5800–6200**
AA6	5800–5940	18·81	–
AA5	**5800–5940**	**5·64**	**5800–6200**
AA4	**5800–5940**	**23·60**	**5800–6200**
AA3	**5800–5940**	**13·34**	**5800–6200**
AA2	**5800–5940**	**27·87**	**5800–6200**
AA1	5800–5940	5·85	–

Table 3. Summary statistics for each time slice. Given are the sample size *n* (with the number of samples left after removing those with more than 90% of missing data), observed heterozygosity (H_{obs}), expected heterozygosity (H_{exp}) and rarefied allelic richness (A_r). Different letters indicate significant pairwise differences after adjustment for multiple testing with the Benjamini & Hochberg (1995) false discovery rate method (FDR; $P < 0·05$)

Age (years BP)	*n*	H_{obs}	H_{exp}	A_r
7000–7140	14 (7)	$0·757^{a}$	$0·458^{a}$	$1·495^{a}$
6600–6800	6 (3)	$0·830^{c}$	$0·469^{a}$	$1·521^{b}$
6200–6500 (decline)	5 (5)	$0·568^{b}$	$0·417^{b}$	$1·469^{a}$
5800–6200	22 (12)	$0·783^{a}$	$0·465^{a}$	$1·484^{a}$

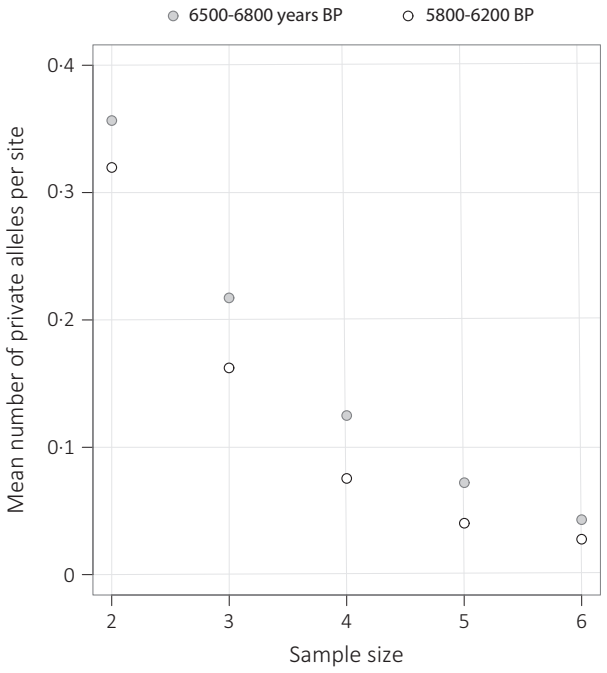

Fig. 5. Mean number of private alleles per site using a rarefaction approach applied to the hyRAD-X dataset, compared between time slices before and after the 6200–6500 years ago decline. White dots: 5800–6200 years BP; grey dots: 6500–6800 years BP. The *x*-axis shows the number of samples considered in the rarefaction approach.

Table 4. Pairwise genetic distance (F_{ST}) among four time slices. Significant values after standard false discovery rate correction (FDR; Benjamini & Hochberg 1995) are indicated with an asterisk ($\alpha = 0·05$) – marginally non-significant values ($0·05 < P < 0·1$) are indicated with ^{MNS}

	7000–7140 BP	6600–6800 BP	6200–6500 BP
6600–6800 BP	−0·004		
6200–6500 BP	−0·0003	$0·069^{MNS}$	
5800–6200 BP	0·037	0·006	0·054*

complexity – GBS or RADseq – have recently bloomed in the population genomics community (Ali *et al*. 2016; Hoffberg *et al*. 2016; Sánchez Barreiro *et al*. 2016; Suchan *et al*. 2016). However, so far, most published techniques (but see Suchan *et al*. 2016) have used commercially synthesized probes and are thus rather costly. In addition, those methods have targeted the whole genome and are therefore at bias of incorporating multiple repeated regions (e.g. transposable elements) in one single locus – despite attempts have been made to partly prevent non-specific hybridizations caused by such repetitive sequences (e.g. by using Cot-I; Thermo Fisher Scientific). Here, we present *hybridization capture from RAD-derived probes obtained from a reduced eXome template* (hyRAD-X), a hybridization capture method employing probes obtained from the transcriptome, produced from messenger RNA

isolated in fresh specimens. By applying hyRAD-X to a set of subfossil samples from a coniferous species (i.e. the silver fir tree, *A. alba*), we tested the applicability of this technique to ancient specimens with highly degraded DNA and its efficiency compared to its sister technique, namely, hyRAD (Suchan *et al.* 2016), which also uses liquid-phase capture with inexpensive bench-top produced probes, but with the latter obtained from the whole genome (not the reduced exome), and being DNA (not RNA) based.

COMPARISON OF HYRAD-X AND HYRAD OUTPUTS

Examination of both hyRAD-X and hyRAD results showed evidence that the analysed reads are from ancient origin and not from contemporaneous contaminants, as revealed by deamination patterns, showing a clear cytosine to thymine conversion at both ends of the DNA reads over at least ten base pairs (Fig. S1). We assumed that the use of DNA probes as in hyRAD might potentially produce DNA chimeras during capture and further amplification step, whereas the application of RNA probes as in hyRAD-X might erase the possibility of chimera production during these steps. Nevertheless, based on the results from the blank samples, blank content in both types of libraries is not significantly different although slightly larger in hyRAD: the maximum number of sequenced reads found in the blank samples was lower than in 92·5% of the subfossil samples with hyRAD-X (Table S1), whereas examination of hyRAD results showed that the number of reads found in the blank controls is lower than 63·6% of the subfossil samples. When blasting the blank content, a large percentage of sequences hit against synthetic sequences (see Fig. S2), suggesting that exogenous contamination is rather low. Nonetheless, after synthetic and exogenous contaminant as well as PCR replicates removal, those slight differences vanished and the number of analysed reads in the poorest sample (either with hyRAD-X or hyRAD) was still larger than in any blank, by an order of magnitude.

During the alignment step of hyRAD-X reads against the published transcriptome of *A. alba* (Roschanski *et al.* 2013), it is by essence ensured that none of the loci proceeds from (ancient) exogenous contaminants. In contrast, when hyRAD reads are aligned against a *de novo*-assembled catalogue (i.e. hyRAD *de novo*, using the pipeline described by Suchan *et al.* 2016), one cannot rule out that some loci may have an (ancient) exogenous origin (e.g. bacterial or fungal). By aligning hyRAD reads against the transcriptome as well (i.e. hyRAD transcriptome), we were able to control for this potential bias. By doing so, we were, however, expecting a lower coverage given that most of the hyRAD probes should derive from non-exomic DNA regions and those matching to the exome-associated probes representing a small fraction of the aligned reads. In contrast, one could expect a higher number of low-coverage SNPs retrieved with hyRAD transcriptome – as well as a larger number of hits against the catalogue – as hyRAD-X probes were produced by using messenger RNA from one single individual at a single time (i.e. from the fraction of the messenger RNA produced in the buds of a 3-years-old seedling at the single moment of collection), therefore representing a reduced representation of the whole exome, at least quantitatively, compared to the hyRAD probes, which were produced from total genomic DNA.

Our results confirmed those assumptions, by showing that both hyRAD-X and hyRAD transcriptome retrieved a larger number of SNPs (42 095 and 43 381, respectively) than hyRAD *de novo* (4435). This pattern was maintained after SNP filtering; interestingly, the total number of SNPs retrieved was the highest with hyRAD transcriptome (1738), followed by hyRAD-X (571) and hyRAD *de novo* (97), indicating that hyRAD transcriptome has a good potential for retrieving exomic polymorphisms. However, this is mainly due to the high level of exome complexity reduction associated with hyRAD-X probes as performed in this experiment, as attested by the very large number of PCR duplicates found in hyRAD-X outputs, that is, two to eight times larger than hyRAD transcriptome and hyRAD *de novo* outputs (see Supplementary Information, Table S2). With 95% of PCR duplicates, hyRAD-X results demonstrate that enrichment was largely saturated and that there is a potential to pool a larger number of samples in one single library, or alternatively, to use messenger RNA from additional organs and/or development stages. This was also confirmed by the mean depth of coverage, which was more than twice larger in the hyRAD-X (37·16) than in the hyRAD transcriptome (15·7) outputs. In order to optimize the samples load per library, one could assess the optimal number of PCR cycles with a qPCR, for both shotgun libraries and enrichment steps. A good estimation of the number of cycles needed for sequencing would prevent against a high number of PCR duplicates. In addition, it would decrease the stochastic over-representation of amplicons that occur when performing a too large number of PCR cycles.

HyRAD-X and hyRAD transcriptome perform better than hyRAD *de novo* at several criteria such as the number of SNPs retrieved (only 97 for hyRAD *de novo*) and the percentage of hits against the reference (see Fig. 3). It is only for the mean depth per site that hyRAD *de novo* performs better than hyRAD transcriptome, although not as well as hyRAD-X (Fig. 4c). The generally larger number of hits against the reference in hyRAD-X and hyRAD transcriptome should be, at least in part, the consequence of using an assembled transcriptome as reference rather than a *de novo* reference generated from fresh samples. For other metrics, such as the matrix fullness (i.e. the inverse of the percentage of missing data; Fig. 4a), all three methods revealed equivalent. The main difference between hyRAD-X and hyRAD transcriptome, both performing well at all retained criteria, is that reads from the former align to the transcriptome only (due to the reduced representation of the exome). As a consequence, coverage is higher with hyRAD-X, meaning that reads are expected to match on many different places of a catalogue contig and therefore decreasing the error rate and increasing the potential length of loci. If required, an additional complexity reduction step could even be performed by applying size selection after

ligation of (MseI-)digested fragments as described in Suchan *et al*. (2016).

HyRAD-X is therefore a suitable method to conduct population genomic studies involving large numbers of subfossil samples showing highly degraded DNA patterns. To our knowledge, this is the first time that a population genomic study incorporating ancient DNA samples of non-mammal species as ancient as 7200 years ago is conducted at a moderate cost using reproducible exome-reduced complexity, with 48 specimens pooled on only two lanes of HiSeq paired end sequencing (i.e. 24 samples per lane). The only exception to date, although involving slightly more recent samples than those from our study and at a likely higher economic cost, is the work by da Fonseca *et al*. (2015), in which domesticated maize varieties as ancient as 6000 years ago were analysed using 318 commercially synthesized probes. With more than 500 SNPs and a matrix completeness of nearly three quarters, our approach outperforms the only other study aimed at exploring the potential of tree subfossil samples in population genomic studies, which unravelled a much lower number of SNPs (<15) from the chloroplast using classical PCR amplification techniques (Magyari *et al*. 2011). HyRAD-X can also be applied to any other non-model organism for a relatively modest cost, in contrast to methods with the same philosophy, but using commercially synthesized probes, which are rather expensive, particularly when large numbers of samples are considered (Ali *et al*. 2016; Hoffberg *et al*. 2016; Sánchez Barreiro *et al*. 2016).

CAN WE PREDICT SUCCESS OF A SUBFOSSIL SAMPLE ANALYSIS WITH HYBRIDIZATION CAPTURE METHODS?

Starting DNA concentration was a poor predictor of whether a sample would show less or more than 90% of missing data. Actually, 27 (68%) out of 48 specimens analysed with hyRAD-X showed less than 90% missing data. Similar proportions were retrieved with hyRAD *de novo* and hyRAD transcriptome with, respectively, 14 and 13 out of 24 specimens (58% and 54%) showing less than 90% missing data. However, the number of pre-captured sequences matching against the *P. abies* genome (Table S1) is significantly correlated to whether or not a library would contain less or more than 90% of missing data. Biologically, we suggest that those specimens not yielding a decent number of SNPs could be needles, which did not reach the sediment at the lake bottom rapidly enough to avoid major DNA decay before being protected by the anoxic conditions in the sediments. During this time before reaching the sediment, biotic decomposition already erased most endogenous DNA from the sample, for example, by high activity of fungal or bacterial nucleases; the half-life of DNA being a complex function of the interplay between the physical, chemical and biological properties of the microenvironment (Pedersen *et al*. 2015). Furthermore, a high DNA concentration after extraction might reflect an elevated proportion of microbial DNA. Microbial DNA can represent up to 99% of the DNA extracted from ancient samples (Skoglund *et al*. 2012). In addition, co-extracted contaminants may give an inaccurate fluorescence signal when measured with Qubit. Thus, the initial amount of DNA measured does not necessarily correspond to the amount of exogenous DNA targeted.

A FIRST INSIGHT INTO EARLY-ANTHROPOCENE DYNAMICS OF CONIFEROUS SPECIES POPULATIONS

Because the wide exome reduction in hyRAD-X allowed to pool larger numbers of samples per Illumina lane than with hyRAD, population genomic analyses were exclusively based on the hyRAD-X output. After pooling the 27 samples into four distinct time slices, we calculated several genetic diversity estimators (Tables 2 and 3). Although the group sizes were modest, our results highlight that observed and expected heterozygosity dropped during the disturbance period 6200–6500 years ago, when *A. alba* populations declined (Table 3) following increased fire activity connected to land use, as attested by the palynological record (e.g. introduced weeds such as *Plantago lanceolata* preceding burning). Observed heterozygosity further recovered a higher value, yet not reaching the peak observed in the time slice just before the decline. Expected heterozygosity (i.e. genetic diversity; Goudet 2005) and allele richness showed a similar trend by also dropping during the decline. Interestingly, genetic differentiation between populations during the decline and, respectively, right before and right after were also significant (vs. time slice 5800–6200 years BP) or marginally non-significant (vs. 6600–6800 years BP) (Table 4). However, there was no genetic differentiation between time slices 5800–6200 years BP and 6600–6800 years BP (Table 4), meaning that the genetic variation present before the decline was recovered afterwards, most likely by *in situ* recolonization from scattered trees or tree stands in the Origlio area that escaped fires. In order to assess the maximum distance from which specimens encompassing the same gene pool recolonized the area, an accurate estimation of isolation by distance patterns in the silver fir should be performed. The absence of recolonization by genomes from other lineages is also attested by the examination of private alleles between the time slices before and after the decline, with a slightly higher number of private alleles in the former, which could suggest a definitive loss of a reduced portion of genetic diversity. The fact that there is also a non-negligible fraction of private alleles in the time slice after the decline should be seen in the light of the rarefaction pattern tending to smooth differences when considering sample sizes of six individuals (Fig. 5) – this may denote that private alleles results should be taken with caution as there is an evident lack of power in our analyses, due to the limited number of individuals analysed. Assuming that genotypes obtained at 524 SNPs are sensitive enough to identify demo-genetic processes, our results do not show a decrease in the genetic diversity of the Origlio *A. alba* population through time, before and after the 6200–6500 years ago decline. This main finding contradicts the hypothesis of Wick & Möhl (2006), who suggested that the extinction of lowland *A. alba* populations in the southern Prealps of Piedmont, Lombardy and Ticino c. 5100 years ago (Tinner *et al*. 1999; Gobet *et al*. 2000; Wick & Möhl 2006) was possibly caused by

reduced genetic variation that impeded adaptation to changing environments.

CURRENT LIMITATIONS IN HYRAD-X APPLICATION

Despite hyRAD-X represents a major step forward in the popularization of ancient DNA analysis, we have identified three ways to improve it:

1 One caveat of using SNPs from the exome-based RAD loci – which are putatively under selection – rather than neutrally evolving regions may be that several assumptions of population genetics models could be violated. However, given that most substitutions in the exome are synonymous, that is, with a proportion ranging from 0·90 to 0·95 (Wolfe, Sharp & Li 1989; Hurst 2002), results should not be heavily biased. In future studies, this limitation could, however, be circumvented by identifying the reading frame and filtering out non-synonymous SNPs.

2 Another possible drawback of our method is that some baits created from messenger RNA may not be able to capture the corresponding DNA molecule, at least near the splice site, given that messenger RNAs are produced from the concatenation of pre-messenger RNAs. Our conservative filtering has removed the possibility of misaligned reads to account for additional SNPs, but at the same time, has limited the total number of SNPs retrieved. In order to circumvent this limitation, first aligning the transcriptome against the genome might allow to split all messenger RNA loci into pre-messenger RNA sub-loci, compatible with the DNA frame, and suitable to serve as a catalogue for all reads. Sequencing of the *A. alba* genome is ongoing (see https://www.aforgen.org/documents/sfgp-initiative.pdf), and once available, we will be able to use splicing-aware aligners (e.g. TopHat; Trapnell, Pachter & Salzberg 2009) to refine the transcriptome catalogue. Availability of genome data will also make possible to accurately distinguish paralogous from orthologous loci and discard the former.

3 In hyRAD-X, the generation of the probes relies on RNA extractions from fresh samples, which allows to study species without previous knowledge about the transcriptome or genome sequences. Whereas using a reduced number of specimens × organs might allow reducing exomic complexity as it is the case in the current study, the selected tissue and the developmental stage of the sample might affect the composition of the probes. Global gene expression analyses in *Arabidospis thaliana* highlighted a similar number of messenger RNAs in distinct tissues, whereas level of expressions greatly varied among tissues (Schmid *et al.* 2005). Thus, using various tissues to generate the probes might increase the messenger RNA diversity and the total number of SNPs retrieved.

NEW AVENUES IN PALEO-POPULATION GENOMICS

Our case study on a coniferous tree population from the southern Alps demonstrates that hyRAD-X is well suited to analyse large sets of ancient DNA samples, here across a time series of tree subfossils. The generation of hyRAD-X-based probes costs c. 220 dollars, but allows to achieve up to 250 captures. In comparison, the price of commercial probes such as the ones used in the pre-designed MYbait capture kit (MYcroarray, Ann Arbor, MI, USA) is 100 dollars for one capture. Even when using dilutions of the MYbait probes up to 10 folds (e.g. Cruz-Dávalos *et al.* 2016; Sánchez Barreiro *et al.* 2016), hyRAD-X still reveals an order of magnitude cheaper than commercially synthesized baits. In addition, commercial probes can be long to synthesize and require prior knowledge of the exome. Overall, our application of hyRAD-X to silver fir subfossil needles shows the potential to incorporate such samples in population genomic studies at a moderate cost in order to compare genetic variation at different time points, and eventually unravel important environmental, ecological and evolutionary questions.

Authors' contributions

N.A., W.T. and C.S. designed the study. S.S., N.A., C.S. and T.S. elaborated the hyRAD-X protocol. E.G. and R.G. collected the samples. R.G. and T.S. tested the extraction protocol. S.S., E.G. and R.G. performed the subsampling. S.S. and C.S. performed ancient DNA extractions as well as RNA extractions. S.S. performed the wetlab and bioinformatic analyses. S.S. and N.A. drafted the manuscript. All authors reviewed the manuscript.

Acknowledgements

We are grateful to Tiziana Pedrotta, Paul Henne, Christoph Schwörer and Willi Tanner for fieldwork. We also thank Fabiola Bastian and Catherine Hänni for preliminary trials on aDNA, Bertalan Lendvay and Morten Rasmussen for advice on aDNA extraction, Catherine Berney, Dessislava Savova Bianchi for trials on RNA extraction, Ludovic Orlando for advising on bioinformatics processing of aDNA and the Lausanne Genomic Technologies Facility as well as FASTERIS SA, NGS services, Switzerland, for the sequencing work. This study was supported by the Swiss National Science Foundation (SNF grant PP00P3_144870 to N.A.).

Data accessibility

Trimmed and filtered reads, alignments and SNPs are deposited in the Dryad Digital Repository https://doi.org/10.5061/dryad.m94f4 (Schmid *et al.* 2017).

References

Ali, O.A., O'Rourke, S.M., Amish, S.J., Meek, M.H., Luikart, G., Jeffres, C. & Miller, M.R. (2016) RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics*, **202**, 389–400.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Aronesty, E. (2011) *ea-utils: Command-Line Tools for Processing Biological Sequencing Data*. Expression Analysis, A Quintiles Company, Durham, NC, USA. Available at: https://expressionanalysis.github.io/ea-utils/ (accessed 1 April 2017).

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. & Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Beckman Coulter (2013) *Agencourt AMPure XP Instructions for Use Guide – PCR Purification*. Beckman Coulter, Inc., Brea, CA, USA.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300.

Birks, H.H. (2003) The importance of plant macrofossils in the reconstruction of Late glacial vegetation and climate: examples from Scotland, western Norway, and Minnesota, USA. *Quaternary Science Reviews*, **22**, 453–473.

Birks, H.H. & Birks, H.J.B. (2000) Future uses of pollen analysis must include plant macrofossils. *Journal of Biogeography*, **27**, 31–35.

Boucher, F.C., Casazza, G., Szövényi, P. & Conti, E. (2016) Sequence capture using RAD probes clarifies phylogenetic relationships and species boundaries in *Primula* sect. *Auricula*. *Molecular Phylogenetics and Evolution*, **104**, 60–72.

Burrell, A.S., Disotell, T.R. & Bergey, C.M. (2015) The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, **79**, 35–44.
Carpenter, M.L., Buenrostro, J.D., Valdiosera, C. *et al.* (2013) Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *The American Journal of Human Genetics*, **93**, 852–864.
Chong, Z., Ruan, J. & Wu, C.-I. (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, **28**, 2732–2737.
Cruz-Dávalos, D.I., Llamas, B., Gaunitz, C. *et al.* (2016) Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources*. doi: https://doi.org/10.1111/1755-0998.12595.
Danecek, P., Auton, A., Abecasis, G. *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
Ellegren, H. (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.
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**, e19379.
Erickson, D.L., Smith, B.D., Clarke, A.C., Sandweiss, D.H. & Tuross, N. (2005) An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18315–18320.
da Fonseca, R.R., Smith, B.D., Wales, N. *et al.* (2015) The origin and evolution of maize in the Southwestern United States. *Nature Plants*, **1**, 14003.
Garrison, E. & Marth, G. (2012) Haplotype-based variant detection from short-read sequencing. *ArXiv Preprint*, **1207**, 3907.
Gobet, E., Tinner, W., Hubschmid, P., Jansen, I., Wehrli, M., Ammann, B. & Wick, L. (2000) Influence of human impact and bedrock differences on the vegetational history of the Insubrian Southern Alps. *Vegetation History and Archaeobotany*, **9**, 175–178.
Goudet, J. (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, **5**, 184–186.
Gyi, J.I., Lane, A.N., Conn, G.L. & Brown, T. (1998) The orientation and dynamics of the C2′-OH and hydration of RNA and DNA RNA hybrids. *Nucleic Acids Research*, **26**, 3104–3110.
Hoffberg, S.L., Kieran, T.J., Catchen, J.M., Devault, A., Faircloth, B.C., Mauricio, R. & Glenn, T.C. (2016) RADcap: sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, **16**, 1264–1278.
Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. & Pääbo, S. (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, **29**, 4793–4799.
Horn, S. (2012) Target enrichment via DNA hybridization capture. *Ancient DNA: Methods and Protocols, Methods in Molecular Biology*, Vol. 840 (eds B. Shapiro & M. Hofreiter), pp. 177–188. Humana Press, New York, NY, USA.
Hurst, L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics*, **18**, 486–487.
Jaenicke-Despres, V., Buckler, E.S., Smith, B.D., Gilbert, M.T.P., Cooper, A., Doebley, J. & Pääbo, S. (2003) Early allelic selection in maize as revealed by ancient DNA. *Science*, **302**, 1206–1208.
Jombart, T. & Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.
Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L. & Orlando, L. (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, **29**, 1682–1684.
Khaleel, S. (2012) *Pre-and post-processing tools for next-generation sequencing de novo* assemblies. Doctoral dissertation, University of Delaware, Newark, DE, USA.
Kistler, L. (2012) Ancient DNA extraction from plants. *Ancient DNA: Methods and Protocols, Methods in Molecular Biology*, Vol. 840 (eds B. Shapiro & M. Hofreiter), pp. 71–79. Humana Press, New York, NY, USA.
Langmead, B. & Salzberg, S. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.
Magyari, E.K., Major, Á., Bálint, M., Nédli, J., Braun, M., Rácz, I. & Parducci, L. (2011) Population dynamics and genetic changes of *Picea abies* in the South Carpathians revealed by pollen and ancient DNA analyses. *BMC Evolutionary Biology*, **11**, 1.
Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T.H., Piñero, D. & Emerson, B.C. (2015) Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.
Metcalfe, C.J. & Casane, D. (2013) Accommodating the load: the transposable element content of very large genomes. *Mobile Genetic Elements*, **3**, e24775.
Meyer, M. & Kircher, M. (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, t5448.
Nystedt, B., Street, N.R., Wetterbom, A. *et al.* (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature*, **497**, 579–584.
Orlando, L., Gilbert, M.T.P. & Willerslev, E. (2015) Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics*, **16**, 395–408.
Pedersen, M.W., Overballe-Petersen, S., Ermini, L. *et al.* (2015) Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society Series B*, **370**, 20130383.
Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.
Puritz, J.B., Hollenbeck, C.M. & Gold, J.R. (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e431.
R Core Team (2016) *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
Roschanski, A.M., Fady, B., Ziegenhagen, B. & Liepelt, S. (2013) Annotation and re-sequencing of genes from *de novo* transcriptome assembly of *Abies alba* (Pinaceae). *Applications in Plant Sciences*, **1**, apps.1200179.
Sánchez Barreiro, F., Vieira, F.G., Martin, M.D., Haile, J., Gilbert, M.T.P. & Wales, N. (2016) Characterizing restriction enzyme-associated loci in historic ragweed (*Ambrosia artemisiifolia*) voucher specimens using custom-designed RNA probes. *Molecular Ecology Resources*, **17**, 209–220.
Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M., Schölkopf, B. & Lohmann, J.U. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nature Genetics*, **37**, 501–506.
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. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.m94f4
Schubert, M., Ermini, L., Der Sarkissian, C. *et al.* (2014) Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature Protocols*, **9**, 1056–1082.
Skoglund, P., Malmström, H., Raghavan, M., Stora, J., Hall, P., Willerslev, E., Gilbert, M.T., Götherström, G.A. & Jakobsson, M. (2012) Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science*, **336**, 466–469.
Suchan, T., Pitteloud, C., Gerasimova, N.S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M. & Alvarez, N. (2016) Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS ONE*, **11**, e0151651.
Szpiech, Z.A., Jakobsson, M. & Rosenberg, N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics*, **24**, 2498–2504.
Tinner, W., Hubschmid, P., Wehrli, M., Ammann, B. & Conedera, M. (1999) Long-term forest fire ecology and dynamics in southern Switzerland. *Journal of Ecology*, **87**, 273–289.
Tinner, W., Lotter, A.F., Ammann, B., Conedera, M., Hubschmid, P., van Leeuwen, J. & Wehrli, M. (2003) Climatic change and contemporaneous land-use phases north and south of the Alps 2300 BC to 800 AD. *Quaternary Science Reviews*, **22**, 1447–1460.
Trapnell, C., Pachter, L. & Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
Wang, Z., Gerstein, M. & Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.
Wick, L. & Möhl, A. (2006) The mid-Holocene extinction of silver fir (*Abies alba*) in the Southern Alps: a consequence of forest fires? Palaeobotanical records and forest simulations. *Vegetation History and Archaeobotany*, **15**, 435–444.
Wolfe, K.H., Sharp, P.M. & Li, W.H. (1989) Rates of synonymous substitution in plant nuclear genes. *Journal of Molecular Evolution*, **29**, 208–211.
Yeates, D.K., Meusemann, K., Trautwein, M., Wiegmann, B. & Zwick, A. (2016) Power, resolution and bias: recent advances in insect phylogeny driven by the genomic revolution. *Current Opinion in Insect Science*, **13**, 16–23.

Received 8 December 2016; accepted 27 March 2017
Handling Editor: M. Gilbert

Supporting Information

Details of electronic Supporting Information are provided below.

Table S1. Percentage of pre-capture reads, reads captured with hyRAD-X and reads captured with hyRAD aligned to the *Picea abies*

(Nystedt *et al.* 2013) genome with Bowtie 2, as well as respective increments.

Table S2. Raw number of reads before contaminants and PCR duplicates removal, total number and percentage of reads matching against the reference, fraction of PCR duplicates, number and percentage of reads matching against the reference when PCR duplicates are removed for the subset of samples for which both hyRAD and hyRAD-X were applied, as well as post-capture increment in endogeneous sequences compared to raw reads.

Fig. S1. Example of DNA misincorporation profiles at 5′-termini obtained with the MapDamage 2.0 software.

Fig. S2. Percentage of matching sequences for different taxa in the extraction blank controls (EBK) and aerosol controls (SABK) for both hyRAD and hyRAD-X procedures.

Fig. S3. Percentage of sequences matching against GenBank public database in the extraction blank controls (EBK) and aerosol control (SABK).

Appendix S1. RNA extraction protocol.