MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1821–1833 doi: 10.1111/j.1365-294X.2012.05537.x

New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems

LAURA S. EPP,*[1] SANNE BOESSENKOOL,*[1] EVA P. BELLEMAIN,* JAMES HAILE,†‡ ALFONSO ESPOSITO,*¶¶ TIAYYBA RIAZ,§ CHRISTER ERSÉUS,¶ VLADIMIR I. GUSAROV,* MARY E. EDWARDS,** ARILD JOHNSEN,* HANS K. STENØIEN,†† KRISTIAN HASSEL,†† HÅVARD KAUSERUD,‡‡ NIGEL G. YOCCOZ,§§ KARI ANNE BRÅTHEN,§§ ESKE WILLERSLEV,† PIERRE TABERLET,§ ERIC COISSAC§[2] and CHRISTIAN BROCHMANN*[2]

*National Centre for Biosystematics, Natural History Museum, University of Oslo, PO Box 1172, Blindern, NO-0318 Oslo, Norway, †The Centre of Excellence for GeoGenetics, Natural History Museum of Denmark, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark, ‡Ancient DNA Research Laboratory, Murdoch University, South Street, Perth, WA 6150 Australia, §Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 9, France, ¶Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden, **Geography and Environment, University of Southampton, University Road, Southampton SO17 1BJ, UK, ††Systematics and Evolution Group, Museum of Natural History and Archeology, Norwegian University of Science and Technology, N-7491 Trondheim, Norway, ‡‡Microbial Evolution Research Group (MERG), Department of Biology, University of Oslo, PO Box 1066 Blindern, N-0316, Oslo, Norway, §§Department of Arctic and Marine Biology, University of Tromsø, NO-9037 Tromsø, Norway

Abstract

Metabarcoding approaches use total and typically degraded DNA from environmental samples to analyse biotic assemblages and can potentially be carried out for any kinds of organisms in an ecosystem. These analyses rely on specific markers, here called metabarcodes, which should be optimized for taxonomic resolution, minimal bias in amplification of the target organism group and short sequence length. Using bioinformatic tools, we developed metabarcodes for several groups of organisms: fungi, bryophytes, enchytraeids, beetles and birds. The ability of these metabarcodes to amplify the target groups was systematically evaluated by (i) *in silico* PCRs using all standard sequences in the EMBL public database as templates, (ii) *in vitro* PCRs of DNA extracts from surface soil samples from a site in Varanger, northern Norway and (iii) *in vitro* PCRs of DNA extracts from permanently frozen sediment samples of late-Pleistocene age (~16 000–50 000 years BP) from two Siberian sites, Duvanny Yar and Main River. Comparison of the results from the *in silico* PCR with those obtained *in vitro* showed that the *in silico* approach offered a reliable estimate of the suitability of a marker. All target groups were detected in the environmental DNA, but we found large variation in the level of detection among the groups and between modern and ancient samples. Success rates for the Pleistocene samples were highest for fungal DNA, whereas bryophyte, beetle and bird sequences could also be retrieved, but to a much lesser degree. The metabarcoding approach has considerable potential for biodiversity screening of modern samples and also as a palaeoecological tool.

Keywords: ancient DNA, Arctic, environmental DNA, metabarcoding, primers

Received 9 February 2012; revision accepted 10 February 2012

Correspondence: Laura Epp, Fax: +47 22851835;
E-mail: laura.epp@nhm.uio.no
¶¶Present address: Faculty of Science and Technology, Free University of Bozen, Sernesistrasse, 1, I-39100 Bozen, Italy.
[1]Contributed equally.
[2]Shared senior authorship.

Introduction

Sequencing of environmental DNA retrieved from soils and sediments plays an important role in the efforts to explore the biodiversity of prokaryotes (Stackebrandt

et al. 1993; Dunbar *et al.* 1999; Rappé & Giovannoni 2003). The targeted retrieval of DNA from environmental samples also promises great potential for the study of eukaryote biodiversity of recent as well as past environments (Hofreiter *et al.* 2003; Willerslev *et al.* 2003, 2007; Coolen *et al.* 2009; Sønstebø *et al.* 2010). In particular, DNA from vascular plants (Sønstebø *et al.* 2010), mammals (Haile *et al.* 2009) and fungi (Lydolph *et al.* 2005) targeted within total soil DNA has yielded promising results. This approach is particularly interesting for organisms that do not fossilize readily or for which only few fossils are found. However, its potential for ecosystem-wide biodiversity and palaeoecological reconstructions that include, for example, invertebrates and vertebrates other than mammals, is currently unclear and requires further evaluation.

Extracts from soil contain DNA from organisms living in the soil at the time of sampling as well as DNA from dead cells and DNA deposited from the surrounding environment (Levy-Booth *et al.* 2007). While the fraction derived from live organisms is largely intracellular and intact, the other fractions will be extracellular and probably highly degraded (Pietramellara *et al.* 2009; Valentini *et al.* 2009b). Such degradation accrues over time, and, particularly for ancient environmental DNA, analyses are typically restricted to very short fragments from multi-copy loci (Pääbo *et al.* 2004).

Genetic markers suitable for diversity analyses through taxonomic identification of DNA preserved in environmental samples (a form of DNA barcoding *sensu lato*; Valentini *et al.* 2009b) must fulfil requirements which partly differ from those of DNA barcodes used for the identification of single specimens (barcoding *sensu stricto*; Valentini *et al.* 2009b). First, they should be short enough to allow amplification from degraded DNA in environmental samples. Second, a diagnostic DNA sequence that is more or less identical within but variable between species is required for optimal taxonomic resolution. Third, this variable DNA marker has to be flanked by highly conserved stretches to which amplification primers can bind. These priming sites should be conserved enough to amplify DNA from a mixture of species belonging to the target organism group with minimal bias (Bellemain *et al.* 2010). Finally, the amplification primers should be highly specific to the target organism group to avoid amplification of nontarget DNA preserved in the sample.

Criterion two, taxonomic resolution to the species level, is of paramount importance for barcoding single specimens (Hebert *et al.* 2003), and the standard marker used for animals is a fragment of the mitochondrial cytochrome c oxidase subunit I gene (COI) with a length of 658 bp. For the analysis of degraded DNA from environmental samples, criteria one, three and four are the most important ones (Valentini *et al.* 2009b), while the fragment size of the classical COI barcode with approx. 650 bp prevents its routine use for the analysis of environmental samples. Primers to target a shorter fragment of the COI gene in all major eukaryotic groups have been suggested (Meusnier *et al.* 2008), but these primers are not conserved even within vertebrates (Ficetola *et al.* 2010), and the originally proposed primer set was not used in consecutive studies (Hajibabaei *et al.* 2011; Shokralla *et al.* 2011). There is thus a clear need for further development and evaluation of primers and markers for analysis of degraded DNA from environmental samples. With reference to the terms metagenomics and barcoding, we designate barcoding markers specifically designed for environmental samples as 'metabarcodes' (Pompanon *et al.* 2011).

In the present study, we designed metabarcoding markers and evaluated their potential for studying the biodiversity of different organism groups in past and present arctic ecosystems using DNA from soils and sediments. We targeted a range of phylogenetically and ecologically distinct groups that have received relatively little or no attention in previous studies of (ancient) environmental DNA: bryophytes, enchytraeids, beetles and birds. These four groups represent ecologically important and abundant taxa in the Arctic (Callaghan *et al.* 2004). We also designed a new metabarcoding primer for fungi targeting a part of the ITS, the proposed standard barcoding marker for fungi (IV International Barcode of Life conference, 2011). The new primer set amplifies a shorter fragment compared with the widely used fungi-specific ITS markers (see Bellemain *et al.* 2010). Some of the selected groups are closely associated with the soil, such as fungi, enchytraeids and various beetle larvae, while others live above ground, such as birds.

For the animal groups, we used mitochondrial DNA, which is well suited for work with degraded samples because of its high copy number per cell (Pääbo *et al.* 2004). Rather than using the standard COI region as a starting point, for the reasons outlined above, we screened either complete mitochondrial genomes or focused on the mitochondrial 12S and 16S rRNA genes. The latter evolve more slowly than the COI gene (Hebert *et al.* 2003), but display stem-loop structures leading to a variation of short stretches of highly conserved and stretches of highly variable DNA (Hickson *et al.* 1996; De Rijk *et al.* 1999). Nuclear rRNA genes also contain such structures, which have been shown to be valuable as markers for species identification (Sonnenberg *et al.* 2007; Raupach *et al.* 2010) because of their hypervariable regions. Short fragments of mitochondrial rRNA genes are good candidate regions for metabarcodes and have previously been identified for

vertebrate identification in degraded samples (Riaz *et al.* 2011).

In this study, we used bioinformatic approaches (Ficetola *et al.* 2010; Riaz *et al.* 2011) to first design a set of metabarcodes suitable for detection of these ecologically important groups and evaluate their performance *in silico*. Second, we evaluated the success of our newly designed markers in retrieving DNA of the target organism groups in recent soils (from the Varanger Peninsula in northern Norway) and in frozen late-Pleistocene sediment samples (Main River and Duvanny Yar, northeast Siberia). Arctic permafrost sediments have previously been a main focus of ancient environmental DNA studies (Willerslev *et al.* 2003), as DNA degradation is retarded under cold conditions (Pääbo *et al.* 2004). The Siberian samples range in age from ~16 000 to 50 000 years BP, a period encompassing the Last Glacial Maximum cold climatic interval and characterized by diverse ecosystems with no contemporary analogues (Blinnikov *et al.* 2011). We compare the predictions obtained from the bioinformatic analyses (*in silico* analyses) with our tests on soil DNA (*in vitro* analyses) and discuss the potential of metabarcoding approaches for the analysis of present and past biodiversity of the different groups studied here.

Material and methods

Study sites and samples

Recent soil samples were obtained from four pairs of heath and meadow plots used in other ecological studies (Ravolainen *et al.* 2011) and located on the Varanger Peninsula in northern Norway (70°19′ N, 30°01′ E and 70°18′ N, 29°06′ E; 110–290 m a. s. l.). The area is characterized by a mosaic of dwarf shrub heath, herb- and grass-rich meadows and willow thickets. A total of eight samples were analysed from the meadow plots (sample names beginning with ENG, Table 2), and seven samples from the heath plots (sample names beginning with HEI, Table 2). Samples were taken in March 2007 by hammering 15-cm-long and 5-cm-wide metal cylinders, which had been thoroughly cleaned and treated with sodium hypochlorite to remove DNA prior to use, into frozen soil cleared of surface vegetation and kept frozen until processed for DNA analyses.

Samples from ancient permafrost were obtained from two key palaeoecological sites in Eastern Siberia: (i) an exposure at Duvanny Yar, on the Kolyma River, northern Sakha (Yakutia) Republic, Russia (68°40′ N, 159°05′ E) and (ii) an exposure on the Main River, a tributary of the Anadyr River in southern Chukotka, Russia (64°17′ N, 171°15′ E; Kuzmina *et al.* 2011). At both sites, samples were taken at different depths along the exposures by drilling cores horizontally with equipment that had been cleaned thoroughly and treated with sodium hypochlorite prior to use. The sampled cores were stored intact and frozen until processing. A total of 14 samples were analysed from each site. Organic material (usually plant macrofossils) extracted from the soil/sediment samples by sieving was radiocarbon dated at facilities in Poznan, Poland or Oxford, UK. Two of the Varanger soil samples used here were also radiocarbon dated and both were confirmed to be 'modern' (i.e. from the past several decades, see Table 2).

Design and optimization of metabarcodes

Metabarcoding markers were designed and evaluated using a bioinformatic approach. Detailed information about settings and databases used are compiled in the Appendix S1 (Supporting information). Searches of potentially suitable metabarcodes were carried out using the program ECOPRIMERS (Riaz *et al.* 2011). This program uses a defined input database of homogeneous sequences (e.g. full mitochondrial genomes) to search for conserved stretches of DNA suitable to be used as primers that flank a region of a specified length (in this study 20–500 bp excluding primers, see Appendix S1, Supporting information). To simulate more realistic PCR conditions (similar to suggestions by Dieffenbach *et al.* 1993), we allowed a maximum of three mismatches between the primer and the target sequences, but no mismatches in the two last bases on the 3′ end of the primer. If not otherwise specified (see Appendix S1, Supporting information), primers were required to strictly (i.e. without mismatches) match 70% of target sequences (option −q 0.70), to match 90% of target sequences allowing a specified number of mismatches (option −s 0.90) and not to match more than 10% of nontarget sequences (option −x 0.10).

For each primer pair, ECOPRIMERS calculates two quality indices: taxonomic coverage (coverage index B_c) and taxonomic resolution capacity (specificity index B_s), as defined by Ficetola *et al.* (2010). Taxonomic coverage is the number of amplified target species relative to the total number of target species in the input database. Taxonomic resolution capacity is the number of unambiguously identified species relative to the total number of amplified target species. From the ECOPRIMERS output, we selected primer pairs with the highest taxonomic coverage and resolution capacity and a relatively short amplicon length. Primer characteristics were optimized using the programs PRIMER3 (http://frodo.wi.mit.edu/primer3/) or FASTPCR (Kalendar *et al.* 2009) and by visual inspection of an alignment containing a subset of sequences from the target taxa. The primers were also

evaluated using the program ECOPCR (Ficetola *et al.* 2010), which performs *in silico* PCRs on a specified database, such as one compiled from all standard sequences in the EMBL Nucleotide Sequence Database (Cochrane *et al.* 2009). The output of ECOPCR contains a list of all sequence entries matching the respective primer pair in a way allowing PCR amplification. This can be used to calculate the taxonomic resolution capacity and coverage for the target group, as defined above, and primer specificity to the target group (number of target species relative to all amplified species). For all final metabarcoding primers, specificity and taxonomic resolution capacity were evaluated using ECOPCR on a database constructed from the standard sequences in the release 107 of the EMBL database (March 2011) with the following parameters: (i) amplicon lengths between 20 and 1000 base pairs and (ii) a maximum of three mismatches between the primer and the target sequence, but no mismatches in the last two bases on the 3′ end. The coverage of finalized primer pairs was calculated by performing an *in silico* PCR on homogeneous databases containing only sequences of the target group (details in Appendix S1, Supporting information).

All primers were tested and their annealing temperatures optimized in the laboratory on DNA extracts from single specimens of each target organism group (details in Appendix S1, Supporting information). Additionally, primers were tested on human and chicken DNA, as these are common laboratory contaminants (Leonard *et al.* 2007), and we aimed to exclude their amplification. PCRs were carried out in 10 μL volumes containing 1 μL of DNA, 0.5 μM of each primer, 1 mM dNTPs, 2.5 mM $MgCl_2$, 1× PCR buffer and 0.4 U Ampli*Taq* DNA Polymerase (Applied Biosystems). PCR conditions were 2 min at 94 °C, followed by 55–60 cycles of 94 °C for 30 s, T_a (see Table 1) for 30 s, 72 °C for 30 s, and a final extension of 10 min at 72 °C.

Molecular genetic work on sediment and soil samples

The intact frozen cores were subsampled from within the cores with sterile scalpels in the ancient DNA laboratory at the Centre for Geogenetics in Copenhagen. Extraction of total DNA was carried out from 7 to 10 g of material. DNA was extracted using the PowerMax™ Soil DNA Isolation kit (MOBIO), with the Powerbead solution replaced by 12 mL of the following buffer: 0.96 mL C1 buffer (from PowerMax™ Soil DNA Isolation kit), 50 mM Tris/HCl, 20 mM EDTA, 150 mM NaCl, 50 mM DDT, 2 mM PTB and 0.8 mg proteinase K. Samples were digested with rotation overnight at 56 °C, and the remainder of the extraction was carried out according to the manufacturer's instructions. The samples were eluted in 2 mL elution buffer.

PCRs were set up in the ancient DNA laboratory at the Natural History Museum in Oslo. For the modern soil samples, DNA was added to the PCR mix in a pre-PCR laboratory dedicated to sensitive (but not ancient) samples. PCR reactions were performed in 12.5 or 25 μL volumes containing 0.625 or 1.25 U Platinum® *Taq* High Fidelity DNA Polymerase (Invitrogen), 1× PCR buffer, 2 mM $MgSO_4$, 1 mM dNTPs, 0.2 μM of each primer, 0.8 mg/mL bovine serum albumin and 1–3 μL of DNA extract. Initial laboratory tests of the primers designed for Coleoptera showed that these primers also amplify human DNA when a high cycle number is used (here 55 cycles). Therefore, a blocking primer (5′–3′ TTCTCGTCTTGCTGTGTCATGCC) was added to the PCR at a 10-fold concentration of the amplification primers (Vestheim & Jarman 2008). Thermal profiles were as described above for primer testing, but the

Table 1 Primer characteristics of the metabarcoding markers

Taxon	Primer name	Primer sequence (5′–3′)	Genomic region	T_a (°C)	Taxonomic coverage (%)
Fungi	ITS5	GGAAGTAAAAGTCGTAACAAGG	ITS1	55	95.2
	5.8S_fungi	CAAGAGATCCGTTGTTGAAAGTT			
Bryophytes	bryo_P6F	GATTCAGGGAAACTTAGGTTG	*trn*L P6-loop	51	86
	bryo_P6R	CCATTGAGTCTCTGCACC			
Enchytraeidae	Ench_12Sa	GCTGCACTTTGACTTGAC	12S	56	98
	Ench_12Sc	AGCCTGTGTACTGCTGTC			
Coleoptera	Coleop_16Sc	TGCAAAGGTAGCATAATMATTAG	16S	55	98.5
	Coleop_16Sd	TCCATAGGGTCTTCTCGTC			
Aves	Aves_12Sa	GATTAGATACCCCACTATGC	12S	58	100
	Aves_12Sc	GTTTTAAGCGTTTGTGCTCG			

Taxonomic coverage is calculated as % of amplified target species of the total number of target species in the database, using *in silico* PCR.

T_a, annealing temperature based on optimization in the laboratory.

extension during each cycle was performed at 68 °C as recommended for the polymerase used. A subset of the positive PCRs was cloned using the Topo TA cloning kit (Invitrogen), and up to 12 clones from each cloning reaction were sequenced on an ABI 3730 sequencer.

Sequence analysis and taxonomic assignments

Clone sequences were analysed using CODONCODE ALIGNER (version 3.6.1). To exclude errors and artefacts from being counted as true variation, clone sequences were only considered if they differed from other sequences by more than one base, or if they were present in three or more clones. If a single clone sequence displayed only a single substitution from an otherwise more common sequence, this substitution was considered likely to be an artefact and the sequence was included in a common consensus sequence.

Taxonomic assignment of the sequences was carried out using the following two approaches:

1 The best matching sequence was determined using the program ECOTAG (http://www.grenoble.prabi.fr/trac/OBITools). This program determines identity between the query sequence and each sequence in a specified reference database through calculating the length of the longest common subsequence (LCS) by an exact algorithm corresponding to a global alignment algorithm (Ullman *et al.* 1976). Identity percentage is subsequently computed by dividing the LCS length by the length of the longest sequence involved in the alignment. For each of the primer pairs, a respective reference database for ECOTAG was created by *in silico* PCR on the EMBL standard sequences, release 107, allowing five mismatches between the primer and the target sequences. To create databases with a secure taxonomy, the ECOPCR output was filtered to merge unique sequences for each taxon in the database, and to include only sequences for which complete taxonomic information is available. For the fungi primers, this filtering was not performed to retain database sequences obtained from uncultured organisms. With the settings used, ECOTAG displays the taxon with the single closest similarity to the query sequence – either a species, if there is a single best matching sequence, or a higher-level taxon if there are multiple sequences with an equally good match. Taxonomic assignment obtained with ECOTAG was only considered if the similarity was over 75%, and no assignment was considered for sequences with a length below 18 bp (the minimum length for a sequence to be used as a specific primer; Dieffenbach *et al.* 1993).

2 As only the best matches were considered in ECOTAG, we performed analyses using the Statistical Assignment Package (Munch *et al.* 2008) to obtain measures of confidence for the assignment of sequences to taxonomic groups. This program compiles a set of homologues to the query sequence using NetBlast searches against GenBank and then uses a Bayesian approach to assign a probability that a sequence belongs to a specific taxonomic group. For some sequences, taxonomic assignment was not possible, either because an insufficient number of homologues could be retrieved to proceed with the Bayesian analysis or because the closest matching sequences were from uncultured organisms with no associated taxonomical information.

Results

Characteristics of the identified metabarcodes

The metabarcoding markers designed for each group are listed in Table 1. For fungi, the marker includes one novel primer (5.8S_fungi) used in combination with one previously published primer that is recommended by the International Fungal Barcoding Group (ITS5; White *et al.* 1990). All other primers are newly designed. The most promising metabarcoding markers for the animal groups were found to be located on the mitochondrial 12S or 16S rRNA genes, not in the standard barcoding region for animals (COI; Hebert *et al.* 2003). For birds, for which we screened complete mitochondrial genomes, the 12S fragment selected was the most optimal mitochondrial metabarcoding marker discovered. The bryophyte marker flanks the P6 loop of the *trn*L chloroplast intron, as previously selected for vascular plants (Taberlet *et al.* 2007). Our markers are hereafter referred to as Fungi_ITS (fungi), Bryo_P6 (bryophytes), Ench_12S (enchytraeids), Coleop_16S (Coleoptera) and Aves_12S (birds).

The coverage of the primers in the target organism groups calculated from *in silico* PCR was high, ranging from 86% (Bryo_P6) to 100% (Aves_12S; Table 1). The taxonomic resolution capacity was more variable among the markers: at the species level, the highest resolution is shown by Ench_12S (100%), followed by Coleop_16S (71.73%), Fungi_ITS (62.84%), Aves_12S (56.36%) and Bryo_P6 (30.57%) (Fig. 1). The taxonomic resolution capacity increased successively from the species to the family level for bryophytes and birds, but not for fungi and Coleoptera, for which the resolution at the genus level (72.11% and 93.60%, respectively) was higher than at the family level (60.36% and 91.27%, respectively).

The median amplicon length of the metabarcodes ranged between 50 and 100 bp excluding primers, except for the Fungi_ITS amplicons, which had a median length close to 200 bp and more variation (Fig. 2; only ~2.5% of the amplicons were more than 300 bp long).

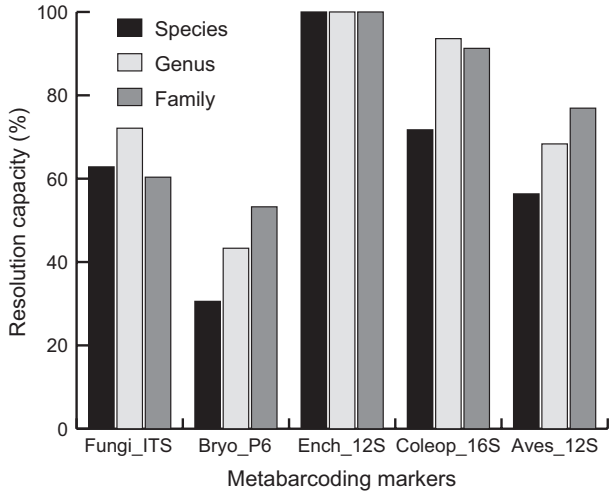

Fig. 1 Taxonomic resolution capacity (specificity index B_s, Ficetola *et al.* 2010) of the metabarcoding markers Fungi_ITS, Bryo_P6, Ench_12S, Coleop_16S, Aves_12S within the respective target taxa. This was calculated from the output obtained using *in silico* PCR on all standard sequences in the EMBL database, release 107.

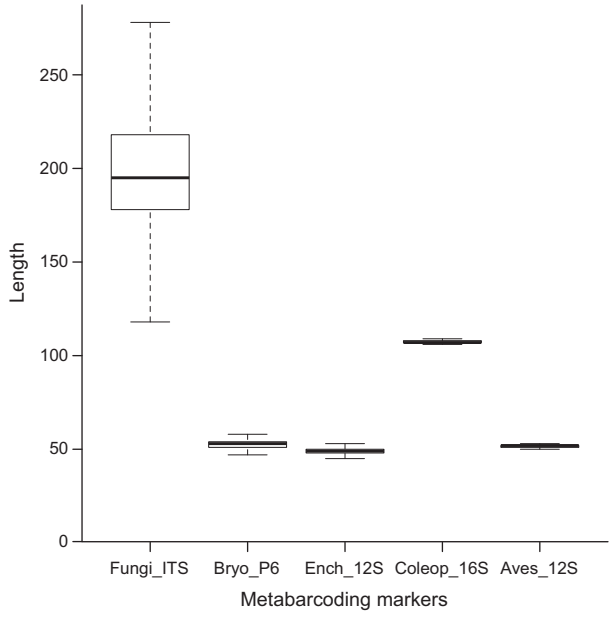

Fig. 2 Boxplots of the amplicon length variation in the metabarcoding markers Fungi_ITS, Bryo_P6, Ench_12S, Coleop_16S and Aves_12S. This was determined from the output obtained using *in silico* PCR on all standard sequences in the EMBL database, release 107. Length is given excluding primer sequences, and outliers are not shown.

Amplification success in recent and ancient arctic soil and sediment samples

Considerable variation in *in vitro* amplification success from soil and sediment samples was observed among the metabarcoding markers (Table 2). In the modern soils from the Varanger Peninsula, all samples amplified using the markers Fungi_ITS, Bryo_P6 and Coleop_16S, and 67% and 13% amplified using Ench_12S and Aves_12S, respectively. Amplification success was substantially lower in the ancient samples from Duvanny Yar and Main River. The highest amplification success was achieved with the marker Fungi_ITS, which yielded positive amplifications in 50% of the ancient permafrost samples. By contrast, no positive amplification of ancient samples was achieved for Ench_12S. None of the four Main River samples older than 26 590 ± 180 years BP gave positive results with any of the primers, but Duvanny Yar samples beyond this age could be amplified using Fungi_ITS, Bryo_P6 and Coleop_16S. Both Fungi_ITS and Bryo_P6 showed positive amplification in one of two extraction blanks, but none of the clone sequences retrieved corresponded to any of the sample sequences and thus did not compromise our results (Table S6, Supporting information).

Marker specificity as evaluated from in silico *and* in vitro *amplifications*

The specificity of the metabarcodes was evaluated and compared for the *in silico* and the *in vitro* amplifications. For this evaluation, the results from the modern and ancient soils were merged (Fig. 3). The *in silico* PCR results are based on the number of species amplified, while the *in vitro* PCR results are based on the number of clone sequences retrieved. Only clones with inserts that differed from primer dimers are reported (details on identification of the clones in the Appendix S1, Supporting information).

For three of the five metabarcoding markers (Fungi_ITS, Bryo_P6, Aves_12S), both the *in silico* PCR and the *in vitro* PCR primarily amplified the target organism groups (Fig. 3, Table S6, Supporting information). However, eight out of 10 clone sequences obtained with the Aves_12S primers were identical to those of chicken (*Gallus gallus*; Table S6, Supporting information), a common laboratory contaminant (Leonard *et al.* 2007). The remaining two sequences, retrieved from a Main River sample with an age of 26 590 ± 180 years BP, were identified as passeriformes. Multiple cloning attempts of the two positive Aves_12S products from the modern soil samples (Table 2) failed to yield any sequences other than a primer multimer in one of the two samples.

For the markers Ench_12S and Coleop_16S, the target organism group did not constitute the majority of amplified sequences *in silico* (11% and 33%, respectively; Fig. 3). Nonetheless, for Ench_12S, all recovered *in vitro* sequences were inferred to stem from enchytraeids. In this case, the majority of the species amplified *in*

Table 2 Amplification success on soil and sediment samples for the different markers

Field Sample	Sample age (years BP)	Lab. identifier	Fungi_ITS	Bryo _P6	Coleop_16S	Ench_12S	Aves_12S
Varanger							
ENG_A_1.2	Undated		+	+	+	−	+
ENG_A_2.2	Undated		+	+	+	−	−
ENG_B_1.2	Modern	Poz-26576	+	+	+	+	+
ENG_C_1.2	Undated		+	+	+	+	−
ENG_D_1.2	Undated		+	+	+	−	−
ENG_D_1.3	Undated		+	+	+	−	−
ENG_D_2.1	Undated		+	+	+	+	−
HEI_A_1.2	Undated		+	+	+	+	−
HEI_A_2.2	Undated		+	+	+	+	−
HEI_B_1.2	Undated		+	+	+	+	−
HEI_C_1.2	Undated		+	+	+	+	−
HEI_D_1.2	Undated		+	+	+	−	−
HEI_D_1.3	Undated		+	+	+	+	−
HEI_D_2.1	Modern	Poz-26591	+	+	+	+	−
HEI_D_2.2	Undated		+	+	+	+	−
Success rate (%)			100	100	100	67	13
Duvanny Yar							
69	16 850 ± 100	Poz-32563	−	−	−	−	−
57A	19 780 ± 130	Poz-32457	+	+	−	−	−
59	20 670 ± 120	Poz-32490	+	−	−	−	−
62	22 900 ± 170	Poz-32557	−	−	−	−	−
64	23 630 ± 190	Poz-32559	+	−	+	−	−
67	25 340 ± 220	Poz-32562	+	−	+	−	−
118	25 830 ± 630	Poz-32681	+	+	+	−	−
122	28 530 ± 800	Poz-32682	+	−	−	−	−
131	29 900 ± 300	Poz-32791	−	−	−	−	−
21A	>48 000	Poz-32301	+	−	+	−	−
24A	50 000 ± 2000	Poz-32232	−	−	+	−	−
28A	>45 000	Poz-32370	+	+	−	−	−
32A	>45 000	Poz-32374	+	−	−	−	−
26B	45 000 ± 2000	Poz-32368	+	−	+	−	−
Success rate (%)			71	21	50	0	0
Main River							
58A	15 810 ± 75	OxA-14930	−	−	−	−	−
55B	20 030 ± 110	Poz-28724	+	−	−	−	−
54B	20 160 ± 110	Poz-28723	+	−	−	−	−
47A	20 830 ± 90	OxA-15667	−	+	−	−	−
56B	20 900 ± 110	OxA-14958	+	−	+	−	−
39A	23 210 ± 130	OxA-15348	−	+	−	−	−
34B	23 880 ± 140	Poz-28680	+	−	−	−	+
36A	25 440 ± 130	OxA-14957	−	−	−	−	−
35B	25 450 ± 160	Poz-28681	−	+	−	−	−
37B	26 590 ± 180	Poz-28682	−	−	−	−	+
33A	28 190 ± 160	OxA-15349	−	−	−	−	−
28A	29 780 ± 210	OxA-14928	−	−	−	−	−
27A	30 900 ± 400	Poz-28653	−	−	−	−	−
19B	>47 000	Poz-28618	−	−	−	−	−
Success rate (%)			29	21	7	0	14

Sample ages, when available, are given with their laboratory identifier in uncalibrated ^{14}C yr BP and counting error.
+, a positive amplification was obtained; −, no amplification.

silico were amphibians, which are not expected widely in the Arctic. With the Coleop_16S primers, only a single of the 57 clone sequences (from a Duvanny Yar sample with an age of 45 000 ± 2000 yr BP) was inferred to stem from a beetle. Some sequences were highly divergent and identified as an enchytraeid, a reindeer

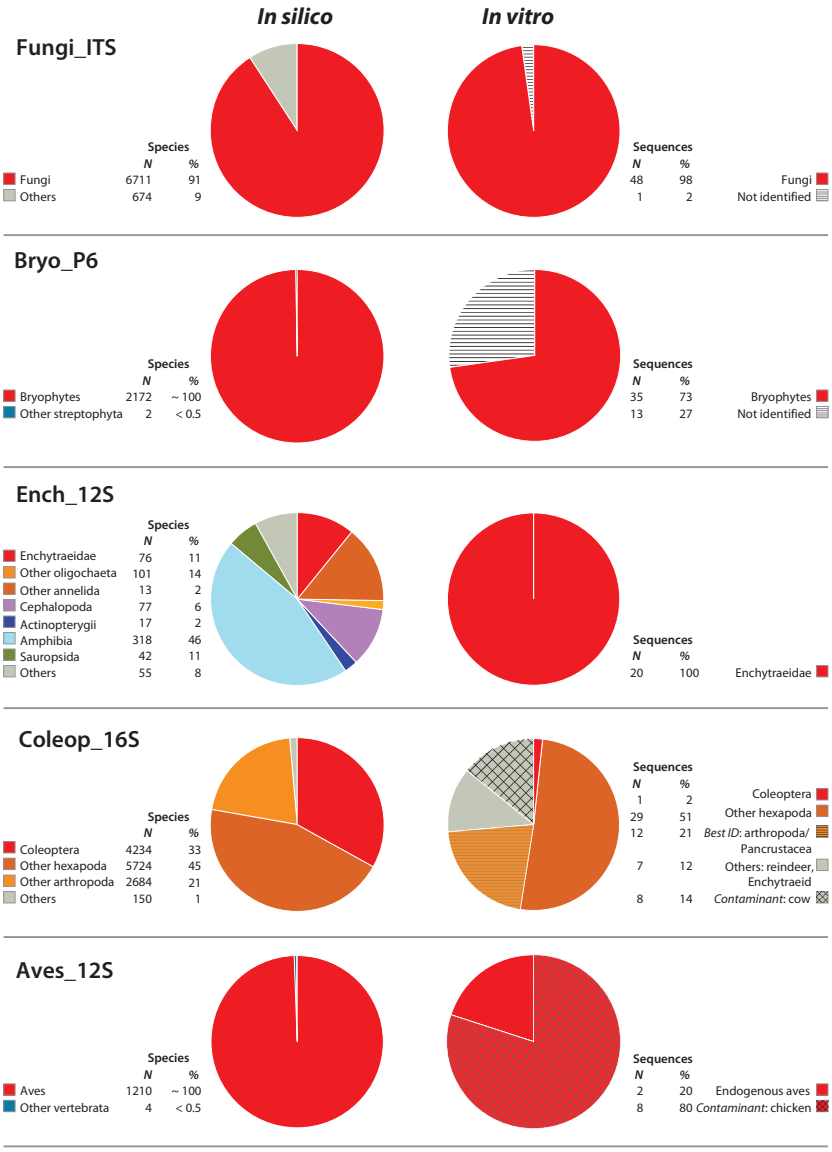

Fig. 3 Comparison of the *in silico* PCR (left) and *in vitro* PCR and cloning of arctic soil and sediment samples (right) using the different markers designed in this study. Actual numbers (*N*) and percentages (%) of species and clone sequences retrieved from the *in silico* PCRs and from the *in vitro* PCRs are given. The *in silico* PCR was performed on all standard sequences in the EMBL database, release 107. Results obtained from the recent and ancient arctic samples were merged.

(*Rangifer tarandus*) and a cow (*Bos taurus*). Notably, the single sample that amplified from Main River only yielded sequences that were identical to those of cow, a common laboratory contaminant (Leonard *et al.* 2007).

Discussion

Total soil DNA is a largely untapped resource for recent and ancient biodiversity information of eukaryotic organisms, and its potential for the study of diverse organism groups has not been comparatively assessed before. For this purpose, we here present a suite of new metabarcoding markers for ecologically highly important organism groups (bryophytes, enchytraeids, beetles and birds), and a new primer for metabarcoding of fungi. We evaluate their performance *in silico* and *in vitro* by amplification of modern and ancient arctic soil and sediment samples.

Taxonomic resolution capacity of the metabarcodes

The large differences in taxonomic resolution capacity of the metabarcodes (Fig. 1) can partly be attributed to the fact that the target groups varied considerably in size, age and evolutionary divergence among their members. The high taxonomic resolution capacity for the Ench_12S primer set (100% at the species level) exemplifies that, even with very short amplicons (~50 bp), reliable taxonomic identification is possible. This has also been demonstrated for all major eukaryotic kingdoms for fragments of the COI gene with lengths between 100 and 250 bp (Hajibabaei *et al.* 2006;

Meusnier *et al.* 2008), but in this case, the suggested primer sites are not sufficiently conserved (Ficetola *et al.* 2010). The limiting factor for obtaining good metabarcodes is obviously not only the length of the sequence amplified, but the possibility to place primers specific to a target group around highly variable sequence fragments. Such conserved primers are less easily found for large groups with a high overall level of evolutionary divergence.

Among the new metabarcodes, the taxonomic resolution capacity at the species level is the lowest for Bryo_P6 (~30%). Nonetheless, this is ~10% higher than the resolution of the P6-loop in vascular plants (Taberlet *et al.* 2007). Like in vascular plants, the taxonomic resolution of the bryophyte P6 loop increases considerably when the set of possible sequences considered is reduced, for example, to the 400 most important arctic bryophyte species (~46% resolution to the species level, unpublished data).

The taxonomic resolution capacity of a barcoding marker calculated as the ratio of unambiguously identified taxa for a given taxonomic level (specificity index B_s, Ficetola *et al.* 2010) is expected to increase with increasing taxonomic level, such that more families than genera and species can be identified. The metabarcodes for bryophytes and birds showed this pattern, but not those for fungi and beetles (Fig. 1). This could be caused by difficulties with higher-level classifications in these two groups (Hibbett *et al.* 2007; Beutel *et al.* 2009) and/or by erroneous taxonomic assignments in GenBank. Alternatively, as these are both very large groups with a high level of evolutionary divergence, the lower resolution at the level of family could also be caused by homoplasy in the short marker sequences.

Suitability of the markers for amplifying the target organism groups

The *in silico* PCR approach offers a simple way to evaluate the performance of the primers, and we have shown here that such an evaluation provides quite a realistic prediction of the *in vitro* performance of the markers. The three primer sets designed for fungi, bryophytes and birds showed high specificity to the target groups, both in the *in silico* PCR and in the *in vitro* PCR. The other two primer sets, designed for beetles and enchytraeids, were not highly specific in the *in silico* PCR, but the *in vitro* products for Ench_12S yielded only enchytraeids. This can be explained by the fact that the other major groups potentially amplified by these primers (e.g. Amphibia, Cephalopoda) are not expected in the arctic environment.

In contrast, the *in silico* amplification of the marker Coleop_16S showed low specificity to Coleoptera (33%), but the specificity obtained from the *in vitro* experiments was even lower (2%). A number of non-arthropod sequences were retrieved, such as the putatively endogenous sequences of an enchytraeid, *Cognettia sphagnetorum*, and a mammal, *Rangifer tarandus* (reindeer), and sequences of the putative laboratory contaminant *Bos* (cow). A comparison of the closest matching sequence in GenBank with the primer sequences revealed that they displayed at most a single mismatch to any of the Coleop_16S primers – but in all cases one of these mismatches occurred at the second base on the 3′ end of the primer. With the restrictions imposed in the *in silico* PCR (three mismatches allowed, but no mismatches in the last two bases at the 3′ end), these amplifications were not predicted. The discrepancy between the specificity predicted by the ECOPCR results and that observed in the *in vitro* PCRs demonstrates the limitations of *in silico* PCR. Variation caused by PCR conditions (e.g. number of cycles, annealing temperature) and differences in template concentrations cannot obviously be taken into account by a pattern-matching algorithm. Hence, the amplification of taxa such as reindeer and cow is not entirely surprising when using a high number of cycles. Nonetheless, competition between templates should result in preferential amplification of coleopteran or other arthropod DNA. It is therefore notable that Coleoptera are not readily amplified from ancient sediment samples, despite the presence of Coleoptera exoskeleton remains in these environments (e.g. Sher *et al.* 2005; Elias 2006), also at the localities Main River (Kuzmina *et al.* 2011) and Duvanny Yar (Alfimov *et al.* 2003).

It should be noted that employing as many as 55 cycles in PCR can lead to the formation of chimeric sequences (Meyerhans *et al.* 1990) and other artefacts, and therefore cycle number should be minimized. However, high numbers of cycles are commonly used in ancient environmental DNA studies (e.g. Willerslev *et al.* 2003; Haile *et al.* 2009) because such samples have a mixed template pool with low initial copy numbers, causing the reactions to behave stochastically. It is however obvious that if included in further analyses, artefacts lead to erroneously increased estimates of the sequence diversity in the samples. Even if short sequences are less prone to chimera formation (Epp *et al.* 2011), retrieved sequence pools should be checked for chimeras (using approaches such as suggested by Creer *et al.* 2010; or specific programs, e.g. Edgar *et al.* 2011) and other artefacts.

Potential for soil metabarcoding studies of different target groups in the Arctic

Of the groups investigated here, fungi were amplified with the greatest success from the permafrost sediment

samples (71% and 29% success rate for Duvanny Yar and Main River, respectively; Table 2). A previous study on ancient fungal DNA preserved in permafrost successfully identified a wide diversity of fungal taxa in samples as old as 300 000–400 000 years (Lydolph *et al.* 2005). The generally high success for fungi is not unexpected as fungal DNA is a dominant component of total soil DNA (Pietramellara *et al.* 2009), present in a higher proportion than plant DNA or even than bacterial DNA in central European agricultural soils (Gangneux *et al.* 2011). Fungal species in the arctic ecosystem are furthermore cryoprotected by a range of different mechanisms (Ozerskaya *et al.* 2009), and it has even been suggested that microbial communities can show long-term viability in the permafrost (Lewis *et al.* 2008; Coolen *et al.* 2011). This could explain that the retrieval of good quality DNA from fungi is higher than for any of the other taxa studied here.

Amplification rates for bryophyte DNA on the other hand were high in the recent soil (100%), but this was contrasted by a low success rate for the Pleistocene samples, indicating that the amount of bryophyte DNA in the ancient sediment samples is very low. Given the ecological importance of bryophytes in the Arctic, this is surprising. However, it is possible that the bryophyte DNA content in soils is lower than that of vascular plants because of their growth form and anatomy. Most bryophyte growth and productivity occurs above ground (Lindo & Gonzalez 2010), and below ground they lack substantial somatic tissue such as true roots. These are thought to be one of the primary sources of vascular plant DNA in soil (Willerslev *et al.* 2003), particularly through the sloughing off of root cap cells (Levy-Booth *et al.* 2007). Furthermore, bryophytes contain secondary metabolites (e.g. Xie & Lou 2009), which are known to enhance DNA degradation. This could potentially cause proportionally higher DNA degradation directly after cell lysis in bryophytes compared to other organism groups.

As for bryophytes, amplification success in ancient samples was not high for the invertebrate groups. Only a single beetle sequence could be retrieved from a Pleistocene sample. Notably, this sample had an age of 45 000 ± 2000 years BP, while no beetle DNA was found in any of the modern samples. This suggests that beetle DNA can potentially be preserved for long time periods, but the potential of sedimentary DNA for biodiversity screenings of beetles, both in modern and in ancient samples, currently seems limited. Information on beetle palaeocommunities can more easily be gained by identifying exoskeleton macrofossils (e.g. Kuzmina *et al.* 2011). Unfortunately, this cannot be done for enchytraeids, which often dominate soil faunal communities in the Arctic, particularly in terms of biomass (Briones *et al.* 2007), but leave no visible fossil traces. No amplification of enchytraeids was achieved from any of the Pleistocene samples, whereas for modern samples amplification success was quite high (67% in total), especially in the heath samples (87.5%). It is noteworthy that this amplification success could be obtained even from samples collected in winter, when population densities of enchytraeids are the lowest (Birkemoe *et al.* 2000).

Finally, both our study and previous studies using cave sediments (Hofreiter *et al.* 2003; Haile *et al.* 2007) have retrieved putatively endogenous avian sequences from soil, although the retrieval rate is not very high. The source of bird DNA in soil could be faeces, which has been shown to contain DNA in modern populations (Regnaut *et al.* 2006; Mäki-Petäys *et al.* 2007), or dead animals. Unfortunately, our results were also confounded by the presence of contaminant chicken DNA, which will readily amplify with the avian metabarcoding primers. This problem cannot be circumvented easily, for example by the inclusion of a blocking primer (Vestheim & Jarman 2008; Gigli *et al.* 2009), because of the lack of sufficient chicken-specific mutations in the amplicon compared to other arctic Galliformes (e.g. ptarmigan). Dominant contaminant DNA may bias the PCR, hampering the retrieval of endogenous DNA present in only low concentrations (Boessenkool *et al.* 2012). Therefore, we cannot fully evaluate the DNA preservation and subsequent potential for diversity reconstruction of bird diversity through sedimentary ancient DNA in the Arctic.

Concluding remarks

The metabarcoding markers developed here testify that high taxonomic resolution and high specificity to target groups are achievable and can be predicted quite well using bioinformatic tools. The approach was most promising for fungi, bryophytes and enchytraeids in the recent soil, but amplification success dropped substantially in the Pleistocene samples. As no ancient samples younger than 15 810 ± 75 years BP were tested here, the potential for historical studies on shorter timescales might nonetheless be considerably larger than our results might seem to indicate.

For modern samples, metabarcoding approaches have great practical potential as an efficient and cost-effective means to conduct biodiversity screenings for ecological surveys, diet studies (Valentini *et al.* 2009a) and biological monitoring programs. Bryophyte community composition, for example, can be used as bioindicator to monitor heavy metal pollution (Denayer *et al.* 1999; Nimis *et al.* 2002), and enchytraeid diversity is suitable for ecological soil classification and assessment schemes

(Jaensch *et al.* 2005). The full potential of metabarcoding can be exploited by coupling next-generation sequencing techniques with identification using reliable reference databases (e.g. Sønstebø *et al.* 2010).

Acknowledgements

This work is part of the BarFrost project supported by the Research Council of Norway (grant no. 191627/V40 to CB). Analyses were performed using the computing facilities provided by the Norwegian Metacenter for Computational Science (Notur). We thank Emilia Rota for identification of enchytraeid voucher specimens.

References

Alfimov AV, Berman DI, Sher AV (2003) Tundra-steppe insect assemblages and reconstruction of Late Pleistocene climate in the lower reaches of the Kolyma River. *Zoologichesky Zhurnal*, **82**, 281–300 (in Russian).

Bellemain EP, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an *in silico* approach reveals potential PCR biases. *BMC Microbiology*, **10**, 189.

Beutel RG, Friedrich F, Leschen RAB (2009) Charles Darwin, beetles and phylogenetics. *Naturwissenschaften*, **96**, 1293–1312.

Birkemoe T, Coulson SJ, Somme L (2000) Life cycles and population dynamics of enchytraeids (Oligochaeta) from the High Arctic. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, **78**, 2079–2086.

Blinnikov MS, Gaglioti BV, Walker DA, Wooller MJ, Zazula GD (2011) Pleistocene graminoid-dominated ecosystems in the Arctic. *Quaternary Science Reviews*, **30**, 2906–2929.

Boessenkool S, Epp LS, Haile J *et al.* (2012) Blocking human contaminant DNA during PCR allows amplification of rare mammal species from sedimentary ancient DNA. *Molecular Ecology*, **21**, 1806–1815.

Briones MJI, Ineson P, Heinemeyer A (2007) Predicting potential impacts of climate change on the geographical distribution of enchytraeids: a meta-analysis approach. *Global Change Biology*, **13**, 2252–2269.

Callaghan TV, Bjorn LO, Chernov Y *et al.* (2004) Biodiversity, distributions and adaptations of arctic species in the context of environmental change. *Ambio*, **33**, 404–417.

Cochrane G, Akhtar R, Bonfield J *et al.* (2009) Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Research*, **37**, D19–D25.

Coolen MJL, Saenz JP, Giosan L *et al.* (2009) DNA and lipid molecular stratigraphic records of haptophyte succession in the Black Sea during the Holocene. *Earth and Planetary Science Letters*, **284**, 610–621.

Coolen MJL, van de Giessen J, Zhu EY, Wuchter C (2011) Bioavailability of soil organic matter and microbial community dynamics upon permafrost thaw. *Environmental Microbiology*, **13**, 2299–2314.

Creer S, Fonseca VG, Porazinska DL *et al.* (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, **19**, 4–20.

De Rijk P, Robbrecht E, de Hoog S, Caers A, Van de Peer Y, De Wachter R (1999) Database on the structure of large subunit ribosomal RNA. *Nucleic Acids Research*, **27**, 174–178.

Denayer FO, Van Haluwyn C, de Foucault B, Schumacker R, Colein P (1999) Use of bryological communities as a diagnostic tool of heavy metal soil contamination (Cd, Pb, Zn) in northern France. *Plant Ecology*, **140**, 191–201.

Dieffenbach CW, Lowe TM, Dveksler GS (1993) General concepts for PCR primer design. *Genome Research*, **3**, S30–S37.

Dunbar J, Takala S, Barns SM, Davis JA, Kuske CR (1999) Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Applied and Environmental Microbiology*, **65**, 1662–1669.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, **27**, 2194–2200.

Elias SA (2006) Quaternary beetle research: the state of the art. *Quaternary Science Reviews*, **25**, 1731–1737.

Epp LS, Stoof-Leichsenring KR, Trauth MH, Tiedemann R (2011) Molecular profiling of diatom assemblages in tropical lake sediments using taxon-specific PCR and Denaturing High-Performance Liquid Chromatography (PCR-DHPLC). *Molecular Ecology Resources*, **11**, 842–853.

Ficetola GF, Coissac E, Zundel S *et al.* (2010) An *in silico* approach for the evaluation of DNA barcodes. *BMC Genomics*, **11**, 434.

Gangneux C, Akpa-Vinceslas M, Sauvage H, Desaire S, Houot S, Laval K (2011) Fungal, bacterial and plant dsDNA contributions to soil total DNA extracted from silty soils under different farming practices: relationships with chloroform-labile carbon. *Soil Biology & Biochemistry*, **43**, 431–437.

Gigli E, Rasmussen M, Civit S *et al.* (2009) An improved PCR method for endogenous DNA retrieval in contaminated Neandertal samples based on the use of blocking primers. *Journal of Archaeological Science*, **36**, 2676–2679.

Haile J, Holdaway R, Oliver K *et al.* (2007) Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? *Molecular Biology and Evolution*, **24**, 982–989.

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

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using River Benthos. *PLoS ONE*, **6**, e17497.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London Series B – Biological Sciences*, **270**, 313–321.

Hibbett DS, Binder M, Bischoff JF *et al.* (2007) A higher-level phylogenetic classification of the Fungi. *Mycological Research*, **111**, 509–547.

Hickson RE, Simon C, Copper A, Spicer GS, Sullivan J, Penny D (1996) Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Molecular Biology and Evolution*, **13**, 150–169.

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

Jaensch S, Rombke J, Didden W (2005) The use of enchytraeids in ecological soil classification and assessment concepts. *Ecotoxicology and Environmental Safety*, **62**, 266–277.

Kalendar R, Lee D, Schulman AH (2009) FastPCR software for PCR primer and probe design and repeat search. *Genes, Genomes and Genomics*, **3**, 1–14.

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

Leonard JA, Shanks O, Hofreiter M *et al.* (2007) Animal DNA in PCR reagents plagues ancient DNA research. *Journal of Archaeological Science*, **34**, 1361–1366.

Levy-Booth DJ, Campbell RG, Gulden RH *et al.* (2007) Cycling of extracellular DNA in the soil environment. *Soil Biology & Biochemistry*, **39**, 2977–2991.

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

Lindo Z, Gonzalez A (2010) The Bryosphere: an integral and influential component of the earth's biosphere. *Ecosystems*, **13**, 612–627.

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

Mäki-Petäys H, Corander J, Aalto J, Liukkonen T, Helle P, Orell M (2007) No genetic evidence of sex-biased dispersal in a lekking bird, the capercaillie (Tetrao urogallus). *Journal of Evolutionary Biology*, **20**, 865–873.

Meusnier I, Singer GAC, Landry JF, Hickey DA, Hebert PDN, Hajibabaei M (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Meyerhans A, Vartanian JP, Wainhobson S (1990) DNA recombination during PCR. *Nucleic Acids Research*, **18**, 1687–1691.

Munch K, Boomsma W, Huelsenbeck JP, Willerslev E, Nielsen R (2008) Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology*, **57**, 750–757.

Nimis PL, Fumagalli F, Bizzotto A, Codogno M, Skert N (2002) Bryophytes as indicators of trace metals pollution in the River Brenta (NE Italy). *Science of the Total Environment*, **286**, 233–242.

Ozerskaya S, Kochkina G, Ivanushkina N, Gilichinsky DA (2009) Fungi in permafrost. *Soil Biology*, **16**, 85–95.

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

Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P (2009) Extracellular DNA in soil and sediment: fate and ecological relevance. *Biology and Fertility of Soils*, **45**, 219–235.

Pompanon F, Coissac E, Taberlet P (2011) Metabarcoding, a new way to analyze biodiversity. *Biofutur*, **30**, 30–32.

Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annual Review of Microbiology*, **57**, 369–394.

Raupach MJ, Astrin JJ, Hannig K, Peters MK, Stoeckle MY, Wagele JW (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology*, **7**, 26.

Ravolainen VT, Bråthen KA, Ims RA, Yoccoz NG, Henden J-A, Killengreen ST (2011) Rapid, landscape scale responses in riparian tundra vegetation to exclusion of small and large mammalian herbivores. *Basic and Applied Ecology*, **12**, 643–653.

Regnaut S, Lucas FS, Fumagalli L (2006) DNA degradation in avian faecal samples and feasibility of non-invasive genetic studies of threatened capercaillie populations. *Conservation Genetics*, **7**, 449–453.

Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, **39**, e145.

Sher AV, Kuzmina SA, Kuznetsova TV, Sulerzhitsky LD (2005) New insights into the Weichselian environment and climate of the East Siberian Arctic, derived from fossil insects, plants, and mammals. *Quaternary Science Reviews*, **24**, 533–569.

Shokralla S, Zhou X, Janzen DH *et al.* (2011) Pyrosequencing for mini-barcoding of fresh and old museum specimens. *PLoS ONE*, **6**, e21252.

Sonnenberg R, Nolte AW, Tautz D (2007) An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology*, **4**, 6.

Sønstebø JH, Gielly L, Brysting AK *et al.* (2010) Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, **10**, 1009–1018.

Stackebrandt E, Liesack W, Goebel BM (1993) Bacterial diversity in a soil sample from a subtropical australian environment as determined by 16S rDNA analysis. *FASEB Journal*, **7**, 232–236.

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

Ullman JD, Aho AV, Hirschberg DS (1976) Bounds on the complexity of the longest common subsequence problem. *Journal of the ACM*, **23**, 1–12.

Valentini A, Miquel C, Nawaz MA *et al.* (2009a) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Molecular Ecology Resources*, **9**, 51–60.

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

Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples - a case study on prey DNA in Antarctic krill stomachs. *Frontiers in Zoology*, **5**, 12.

White T, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR-Protocols A Guide to Methods and Applications* (eds Innis MA, Gelfand DH, Sninski JJ and White TJ), pp. 315–322. Academic Press, San Diego.

Willerslev E, Hansen AJ, Binladen J *et al.* (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, **300**, 791–795.

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

Xie CF, Lou HX (2009) Secondary metabolites in Bryophytes: an ecological aspect. *Chemistry & Biodiversity*, **6**, 303–312.

Data accessibility

All environmental clone sequences are deposited in the Dryad database: doi: 10.5061/dryad.37s6bb42.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Databases for design of the fungi metabarcode, created by *in silico* PCR using ECOPCR.

Table S2 Fungi species used for primer testing in the laboratory.

Table S3 Bryophyte species used for primer testing in the laboratory.

Table S4 Specimen IDs and species names for Enchytraeidae samples that were used for primer testing in the laboratory.

Table S5 Coleoptera specimens used for primer testing in the laboratory.

Table S6 Taxonomic identities derived from the programs ecoTag and the Statistical Assignment Package (SAP) for cloned sequences.

Appendix S1 Supplementary material: Details of primer design and details of clone sequence identification.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.