MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 3647–3655 doi: 10.1111/j.1365-294X.2012.05545.x

FROM THE COVER

DNA from soil mirrors plant taxonomic and growth form diversity

N. G. YOCCOZ,* K. A. BRÅTHEN,* L. GIELLY,† J. HAILE,‡§ M. E. EDWARDS,¶ T. GOSLAR,** H. VON STEDINGK,¶ A. K. BRYSTING,†† E. COISSAC,† F. POMPANON,† J. H. SØNSTEBØ,†† C. MIQUEL,† A. VALENTINI,† F. DE BELLO,†,‡‡ J. CHAVE,§§ W. THUILLER,† P. WINCKER,¶¶ C. CRUAUD,¶¶ F. GAVORY,¶¶ M. RASMUSSEN,‡ M. T. P. GILBERT,‡ L. ORLANDO‡ C. BROCHMANN,††[1] E. WILLERSLEV,‡[1] and P. TABERLET,†[1]

*Department of Arctic and Marine Biology, University of Tromsø, NO-9037 Tromsø, Norway, †Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 43, F-38041 Grenoble Cedex 9, France, ‡Centre for GeoGenetics, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark, §Murdoch University, Perth, Western Australia 6150, Australia, ¶University of Southampton, Geography and Environment, Southampton SO17 1BJ, UK, **Faculty of Physics, Adam Mickiewicz University, ul. Umultowska 85, 61-614 Poznan, Poland, ††National Centre for Biosystematics, Natural History Museum, University of Oslo, PO Box 1172, Blindern, N-0318 Oslo, Norway, ‡‡Institute of Botany, Czech Academy of Sciences, Dukelská 135, CZ-379 82, Třeboň, Czech Republic, §§Laboratoire Evolution et Diversité Biologique, CNRS UMR 5174, Université Paul Sabatier, F-31062 Toulouse, France, ¶¶Genoscope, CEA, CNRS, UMR 8030, 2 rue Gaston Crémieux, BP 5706, F-91057 Evry cedex, France*

Abstract

Ecosystems across the globe are threatened by climate change and human activities. New rapid survey approaches for monitoring biodiversity would greatly advance assessment and understanding of these threats. Taking advantage of next-generation DNA sequencing, we tested an approach we call metabarcoding: high-throughput and simultaneous taxa identification based on a very short (usually <100 base pairs) but informative DNA fragment. Short DNA fragments allow the use of degraded DNA from environmental samples. All analyses included amplification using plant-specific versatile primers, sequencing and estimation of taxonomic diversity. We tested in three steps whether degraded DNA from dead material in soil has the potential of efficiently assessing biodiversity in different biomes. First, soil DNA from eight boreal plant communities located in two different vegetation types (meadow and heath) was amplified. Plant diversity detected from boreal soil was highly consistent with plant taxonomic and growth form diversity estimated from conventional above-ground surveys. Second, we assessed DNA persistence using samples from formerly cultivated soils in temperate environments. We found that the number of crop DNA sequences retrieved strongly varied with years since last cultivation, and crop sequences were absent from nearby, uncultivated plots. Third, we assessed the universal applicability of DNA metabarcoding using soil samples from tropical environments: a large proportion of species and families from the study site were efficiently recovered. The results open unprecedented opportunities for large-scale DNA-based biodiversity studies across a range of taxonomic groups using standardized metabarcoding approaches.

Keywords: biodiversity assessment, DNA metabarcoding, environmental sequencing, functional diversity, plant diversity

Received 30 November 2011; revision received 13 February 2012; accepted 16 February 2012

Correspondence: Pierre Taberlet, Fax: +33 476 51 42 79;
E-mail: pierre.taberlet@ujf-grenoble.fr

[1] These three authors share senior authorship.

Introduction

Climate change, habitat loss and invasive species are affecting biodiversity, ecosystem function and ecosystem services worldwide, placing ecosystems under increasing pressure (Sala *et al.* 2000). To assess and understand these changes requires both intensive studies focussing on mechanisms and extensive approaches allowing for rapid and economical assessment of biodiversity over large regions (McMahon *et al.* 2011). DNA barcoding—use of a standardized DNA sequence to tag and identify species—is becoming a popular tool for taxonomic identification of individual specimens (Hebert *et al.* 2003), but its usefulness as a tool for describing and analysing patterns of species diversity and abundance within ecosystems, without relying on individual specimens or individual parts (as e.g. in Kesanakurti *et al.* (2011)), has not been evaluated despite high potential (Valentini *et al.* 2009b). Existing approaches based on DNA that target multiple taxa have focussed on micro-organisms, for example, bacteria (Zinger *et al.* 2009), nematodes (Porazinska *et al.* 2009), micro-invertebrates (Chariton *et al.* 2010), and on determining diet composition based on samples of faeces (Valentini *et al.* 2009a; Kowalczyk *et al.* 2011; Rayé *et al.* 2011) or stomach content (Soininen *et al.* 2009).

In particular, DNA from soil samples might provide an efficient metric for components of above- and below-ground ecosystem diversity, including but not limited to vascular plant diversity. The approach has so far been restricted to ancient sediments revealing past animal and plant biodiversity (Willerslev *et al.* 2003, 2007). Current methods for estimating contemporary plant diversity rely on time-consuming above-ground sampling, usually abundance or biomass measurements of individuals and their taxonomic identification (Magurran 2004; Stohlgren 2007). While these methods have been and will remain invaluable, information contained in the soil could both complement above-ground data and be used to estimate components of plant diversity over longer temporal scales. Indeed, as DNA is likely to accumulate in the soil over more than a year, it should provide a temporally integrated view of plant community composition. Inability to detect particular species because of observer effects, plants entering dormancy, colonization/extinction dynamics at very small scales or rarity at early stage of invasion or late stage of extinction is increasingly recognized as affecting the accuracy of measures of diversity changes (Thompson 2004; Chabrerie *et al.* 2008; Chen *et al.* 2009; Vittoz *et al.* 2010). As DNA is probably more sensitive to small quantities of biomass, an approach based on soil DNA may alleviate some of these issues. At a practical level, sampling of soil is not limited to the growing season of plants; hence, the approach introduces flexibility to field campaign planning and reduces the likelihood for misidentification under field conditions. Furthermore, expert taxonomic knowledge can be invested in the development of large DNA reference databases and their analyses rather than routine identification of species at each site. Future studies can combine these databases with next-generation sequencing technologies in the analysis of environmental DNA samples. The sampling of soil DNA can be designed using well-known statistical principles [e.g. stratified random sampling based on known environmental and/or disturbance gradients (Thompson 2004; Albert *et al.* 2010)] that would make conclusions obtained from surveys and analytical studies statistically straightforward and valid over a range of spatial scales.

The use of soil-derived DNA to identify the presence and abundance of plants poses technical challenges, as much of the DNA is likely to be degraded. The CBoL Plant Working Group (2009) recently described standardized DNA barcodes for plants; these barcodes target relatively long DNA fragments (~550 bp for *rbc*L and ~950 bp for *mat*K). Using such long DNA sequences for taxonomic identification in environmental samples is likely to result in few positive matches and many species being missed, as experienced in ancient DNA studies (Willerslev *et al.* 2003). Shorter DNA markers for biodiversity assessment are therefore a compromise between the discriminatory power of sequences used to tag species and their ability to work on degraded soil DNA samples. While standardized barcodes cannot be used for environmental samples containing degraded DNA, short DNA markers such as the P6 loop of the plastid *trn*L (UAA) intron have the potential for such analyses (Taberlet *et al.* 2007). Here, we implement an approach that can be termed 'DNA metabarcoding', as it is based on the DNA barcoding concept, but refers to the analysis of complex substrates containing DNA from multiple species. We focussed on three related issues: (i) How well can soil DNA reflect the structural and functional diversity of the current vegetation? (ii) How long can DNA persist in soil? (iii) Can DNA metabarcoding be used effectively in warm, tropical environments where diversity is very high and DNA degradation is fast?

Material and methods

Study systems, DNA extraction and sequence analyses

Boreal site. We sampled four paired heath–meadow plots (15 m × 15 m) in a vegetation mosaic of dwarf shrub–dominated heath, forb- and grass-rich meadows (Varanger Peninsula; 70°N, 30°E; 110–290 m a.s.l.). The

minimum distance between pairs of plots was 1 km, and the mean distance between heath and meadow plots within a pair was 275 m. In July 2006, above-ground plant biomass was estimated in each plot by the point intercept method using 20 pins in 13 0.21 m^2 quadrats regularly spaced out across each plot. Records of plant species at each quadrat and plot were converted to biomass (Ravolainen *et al.* 2010) using equations given in the study conducted by Bråthen & Hagberg (2004). Three triplets of soil core samples per plot were taken in March 2007 by hammering sterile metal cylinders (5 cm diameter, 15 cm long) into frozen soil cleared of vegetation. Soil core samples were kept frozen until processed for DNA analyses.

Plant fragments and soil matrix organics extracted from 12 of the soil cores (six in meadow, six in heath) were radiocarbon-dated (see Goslar *et al.* 2005). Samples yielded pMC (per cent modern carbon) values between 111 and 138, indicating that they date to around the bomb peak 1950–2005.

Total DNA was extracted from *c.* 6 g of soil without visible root fragments using the PowerMax Soil kit [MO BIO Laboratories, Inc., Carlsbad, CA, USA (Willerslev *et al.* 2003)]. The P6 loop of the plastid DNA *trn*L (UAA) intron was amplified using the *g* and *h* primers [5′-GGGCAATCCTGAGCCAA-3′ and 5′-CCATTGAGTC TCTGCACCTATC-3′, (Taberlet *et al.* 2007)], which had been 5′-labelled for each soil core sample with a unique six-nucleotide tag [with at least three differences among tags, a system modified from Binladen *et al.* (2007)]. Being short but variable, the P6 loop marker is suitable for analysing degraded DNA sequences. Further, it is sufficiently long to confidently assign a sequence to a species/genus if identification attempts are restricted to a local geographical scale (Taberlet *et al.* 2007). To prevent environmental contamination, both DNA extractions and PCR set-up were carried out in dedicated laboratories for analyses of low copy number DNA samples. The PCR products were sequenced on the Roche GS FLX platform following the manufacturer's instructions.

All sequences were identified using the soil core sample-specific 5′ variable primer tag after removing all sequences with a tag-sequence error (Binladen *et al.* 2007). We subsequently compared the sequences of each soil core sample to the *trn*L Arctic plant reference database (Sønstebø *et al.* 2010) (Table S1, see Data accessibility) or to GenBank using the ecoTag software (http://www.grenoble.prabi.fr/trac/OBITools). This program relies on an exact global alignment algorithm (Needelman & Wunch 1970) to align each sequence from the soil core sample with sequences in the database and calculate their similarity. Taxon identification required a 100% match on the whole sequence length [using the NCBI taxon identifier (taxid) (Benson *et al.* 2009)]. Several groups of closely related species had the same P6 loop sequence. As a consequence, the DNA profiling approach as implemented in this study has a lower taxonomic resolution compared with species lists based on above-ground surveys.

Temperate site. We selected 25 semi-natural meadows from subalpine vegetation in the French Alps (Le Pied du Col, Villar-d'Arêne; 45.04°N, 6.34E°; 1726–1847 m a.s.l.). Fourteen meadows were part of a formerly cultivated terrace (with a rotation of cereals—mainly barley—and potatoes) abandoned at different dates between 1810 and 1986. Date of abandonment and which crops had been planted were known from local registers. The remaining 11 meadows have never been cultivated and were within 1–4 km of the formerly cultivated meadows. Eight soil cores were taken at the time of the above-ground survey using a clean steel coring sampler. Soil samples were preserved dry in silica gel before DNA extraction. DNA was extracted from 3 g of dry soil using the PowerMax Soil kit. The P6 loop of the plastid DNA *trn*L (UAA) intron was amplified using the *g* and *h* primers that had been 5′-labelled for each soil core sample with a unique nine-nucleotide tag (with at least three differences among tags). All PCR products from the different samples were first titrated using capillary electrophoresis (QIAxel; QIAgen GmbH, Hilden, Germany) and then mixed together, in equimolar concentration, before the sequencing. The sequencing was carried out on an Illumina Genome Analyzer IIx (Illumina, San Diego, CA, USA), using the Paired-End Cluster Generation Kit V4 and the Sequencing Kit V4 (Illumina), following the manufacturer's instructions. A total of 108 nucleotides were sequenced on each extremity of the DNA fragments. The sequence reads were analysed using the OBITools (http://www.grenoble.prabi.fr/trac/OBITools). First, the direct and reverse reads corresponding to a single molecule were aligned and merged using the solexaPairEnd program, taking into account quality of data during the alignment and the consensus computation. Then, primers and tags were identified using the ngsfilter program. Only sequences with perfect match on tags and a maximum of two errors on primers were selected. The amplified regions, excluding primers and tags, were kept for further analysis. Strictly identical sequences were clustered together using the obiuniq program, keeping the information about their distribution among samples. Sequences shorter than 10 bp, or containing ambiguous nucleotides, or with occurrence lower or equal to 10 were excluded using the obigrep program. Finally, taxon identification was carried out by combining the results obtained with a BLAST search in the EMBL

database and the list of species known to be present in the area.

Tropical site. Our tropical site was located in central French Guiana, at the Nouragues Research Station, located North of the Arataie River, a tributary of the Approuague River, 100 km South of Cayenne (4°05′26 N, 52°40′48 W°; 114 m a.s.l.). The vegetation here is a pristine lowland tropical rain forest. Rainfall is 2824 mm per year (average 1988–2008) with a dry season averaging 2.5 month, from late August to early November, and a shorter dry season in March. The plant diversity of the 100 000 ha Nouragues Natural Reserve surrounding the study site is high, with a local flora exceeding 1700 angiosperm species. In August 2009, 49 soil cores were sampled across 25 hectares of the Grand Plateau permanent sampling tree plot (Bongers *et al.* 2001). This plot is covered with an old-growth tropical forest with canopy height around 35 m above ground, and it harbours a species diversity of *c.* 200 tree species per hectare. Within this plot, 100 litterfall collectors were set up in February 2001, and collections have been made twice monthly since this date as part of a long-term monitoring programme (see Norden *et al.* 2007, Chave *et al.* 2008 for more details). The soil cores were collected within 5 m of every other litterfall collector, at a random location. Sequencing and data analysis of the sequence reads obtained were identical to those implemented for temperate soil (see above), except that identification of the different taxa was carried out using the EMBL database. The chloroplast *trn*L (UAA) intron sequences of most tree species of the Nouragues research station (Gonzalez *et al.* 2009) have been uploaded to the EMBL database.

Statistical analyses

In the boreal site analysis, we used an ordination method, nonsymmetric correspondence analysis, which is consistent with the decomposition of Simpson diversity (Pélissier *et al.* 2003). Simpson diversity is a diversity measure giving more weight to frequent species (Sugihara 1982). The ordination was based on presence–absence data, for above-ground vegetation at the taxonomic level of species and for soil DNA at the taxonomic level of genus.

To investigate further how the relative abundance of DNA sequences retrieved from next-generation sequencing of the soil extracts reflected relative biomass, and as above- and below-ground biomass differences are likely to be related to functional groups, we grouped the boreal species according to plant growth forms (Chapin *et al.* 1996). We focussed on three dominant growth forms central to ecosystem processes and services (Bråthen *et al.* 2007; Wookey *et al.* 2009): woody plants, graminoids and forbs. We used linear models to relate the proportion of the total soil DNA fragment pool to the proportion of total above-ground biomass for each growth form. We logit-transformed proportions of woody growth forms, as some proportions were close to 0/1. Predicted values were back-transformed to the proportion scale.

Length of the DNA marker used and diversity of plant species recovered

To test the influence of marker length on the diversity of plant species recovered using the DNA approach, we performed an additional analysis of four samples from the boreal site and four samples from the tropical site. These eight DNA extracts were amplified with the *rbc*L primers recommended by the CBoL Plant Working Group (2009) (rbcLa_f 5′-ATGTCACCACAAAC AGA GACTAAAGC-3′ and rbcLa_rev 5′-GTAAAATCAAGTC CACCRCG-3′). These primers amplify a 553-bp fragment of the *rbc*L gene. The amplification products were sequenced on the Roche GS FLX platform following the manufacturer's instructions.

Results

Boreal site: taxonomic and growth form diversity of current vegetation

PCR amplification of the *trn*L P6 loop and high-throughput sequencing of the soil DNA resulted in 176 283 taxonomically identified sequences (Table 1; Table S1, see Data accessibility), corresponding to a total of 29 families, 66 genera (or sets of closely related genera with identical barcodes) and 79 species (or sets of closely related species with identical barcodes). Of the 71 plant species identified

Table 1 Summary of the DNA results obtained in the three different sites (boreal, temperate and tropical). Only the most common molecular operational taxonomic units representing a total of 80% of the sequence data sets were considered

Sites	No. of soil core samples	Sequencing platform	No. of sequence reads analysed	No. of families identified
Boreal (Varanger, Norway)	72	Roche 454 FLX	176 283	29
Temperate (French Alps, France)	200 (for 25 plots)	Illumina GA IIx	3 901 106	31
Tropical (Nouragues Field Station, French Guiana)	49	Illumina GA IIx	1 636 455	34

in the above-ground vegetation, 47 species and seven species sets (each consisting of from two to six closely related species) matched unique DNA barcodes. Only six species, representing 0.7% of the total biomass and 2.4% of occurrences, were not recovered from the soil DNA (Table S1, see Data accessibility). Multivariate ordinations gave very similar results for above-ground vegetation composition and soil DNA data (Fig. 1). Heath and meadow plant communities were clearly separated by both approaches; furthermore, the differences in above-ground vegetation composition of the meadow communities were also clearly identified by the soil DNA analysis.

Biomass and DNA sequence proportions were strongly related over the range of sample values, but the relationship differed among growth forms (Fig. 2). For graminoids, the relationship was approximately 1 : 1. Woody plants, which dominated the biomass of the heath habitats, constituted a substantially lower proportion of the soil DNA. Conversely, forb representation in soil DNA was greater than in above-ground biomass.

Temperate site: persistence of soil DNA

Crops cultivated 40–50 years ago could still be detected, albeit at low frequencies (Fig. 3). There was a positive relationship between the number of DNA sequences and the number of years since crop abandonment, reflecting the decay of DNA in soil. We never detected crop sequences in the 11 uncultivated plots.

Tropical site: high diversity recovered

We were able to extract 4343 unique sequences. These sequences corresponded to a large number of taxa: 216 molecular taxonomic units and 34 families have been identified. Because of the relatively low sequence variation of the P6 loop of the *trn*L intron in some families (e.g. Sapotaceae, Lauraceae), some of the molecular taxonomic units identified might correspond to several species (Table 1 and Table S2, see Data accessibility). A partial botanical census is available for this 25-ha plot, and it currently lists over 600 tree species and 100 families, thus even our small sample of 49 soil cores included up to a third of the plant species in the plot.

Comparison between the P6 loop of the trn*L intron and* rbc*L*

For *rbc*L, the DNA amplification was weaker in tropical soils than in boreal soils, and we obtained a total of 70 469 and 7855 sequence reads for the boreal and the tropical samples, respectively. Table 2 summarizes the results obtained. Clearly, the short P6 loop of the *trn*L

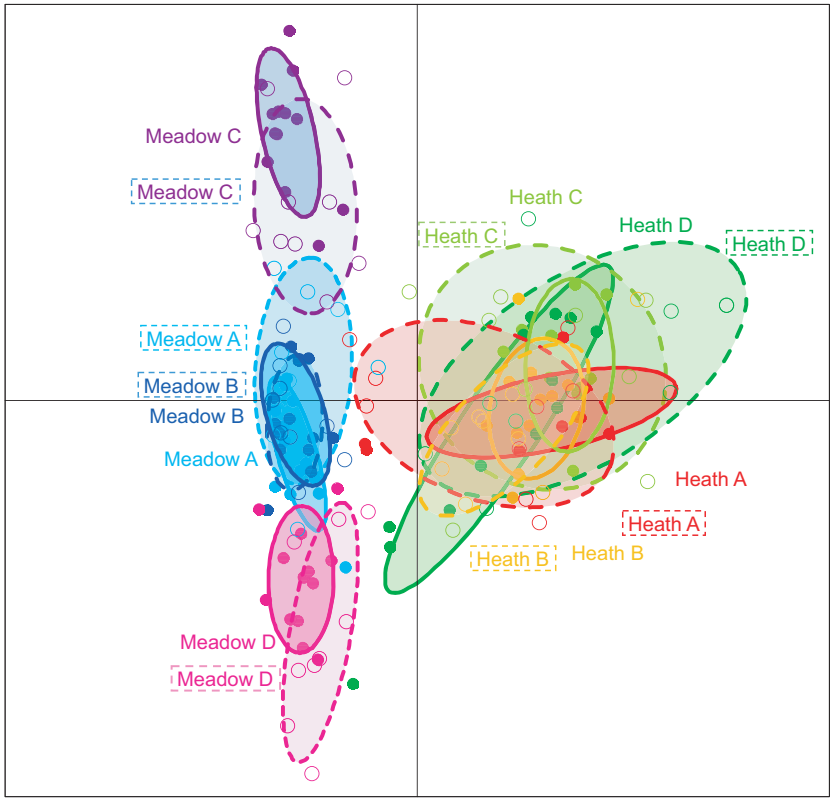

Fig. 1 Multivariate ordination of plant diversity of heath (A to D) and meadow (A to D) communities based on above-ground vegetation biomass (full circles) and soil DNA (open circles) for the boreal site. Ellipses show within-community variability of vegetation (continuous) and soil DNA (dotted) samples.

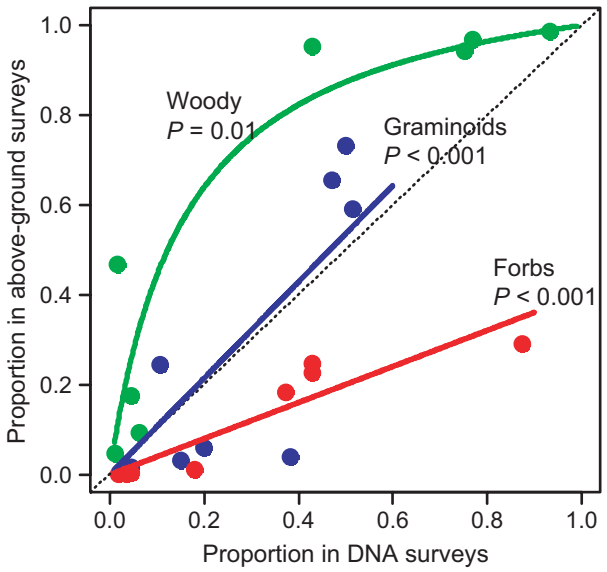

Fig. 2 Relationship between proportion of soil DNA and proportion of above-ground biomass for three functional groups, woody plants, graminoids and forbs (boreal site). The equivalence line (dotted) and fitted models (continuous lines) are shown together with significance level P. Models relate proportions (forbs and graminoids) or logit-transfomed proportions (woody).

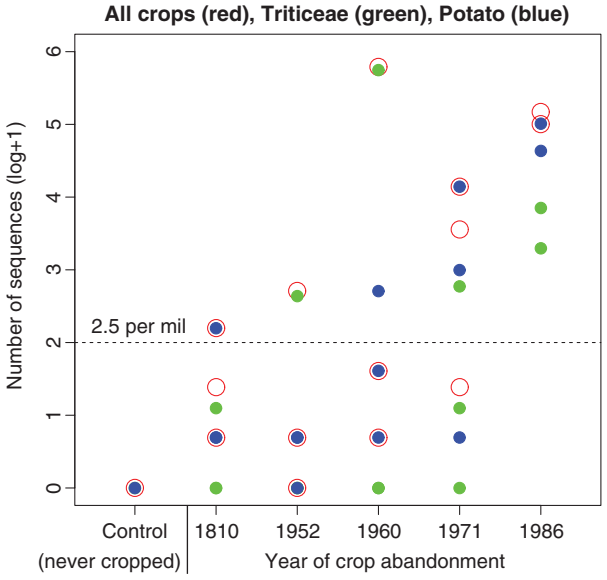

Fig. 3 Relationships between frequencies of soil DNA sequences (log(x + 1)) and year of crop abandonment as recorded in local registries. Two crops dominated: cereals [Triticeae, mainly barley, *Hordeum vulgare*, but also rye (*Secale cereale*) and oat (*Avena sativa*)] and potato (*Solanum tuberosum*). Different plots were sampled for each year of abandonment (2 plots for 1986, 3 for 1971, 3 for 1960, 3 for 1952 and 3 for 1810) and for meadows that were never cropped (i.e. control, 11 plots). The '2.5 per mil' line generally indicates the mean contribution of crop sequences in all plots that were previously cultivated.

intron is much more efficient than the longer *rbc*L marker for estimating plant diversity based on soil DNA.

Discussion

Our results show that soil DNA metabarcoding can be applied across a variety of biomes and reveals even taxa that are not detectable through traditional surveys. First, soil DNA metabarcoding led to an excellent match with above-ground diversity for the boreal plant communities. Second, crops cultivated a few decades ago were still detectable in modern soil, but their frequency declined to a very low level after *c.* 50 years. Third, even in moist tropical environments where DNA should quickly degrade, we were able to recover and assign to taxonomic units large number of sequences. We discuss below some of the main challenges or opportunities raised by the use of DNA metabarcoding.

Differences between DNA sequence frequencies and biomass are probably related to the absolute amount of DNA in the soil associated with different litter turnover rates (Cornelissen *et al.* 2007) or root : shoot biomass ratio differences among growth forms (Aerts & Chapin 2000). While the woody component of above-ground biomass of shrubs is highest, graminoids are richer in lignins than are forbs. This could explain the observed relationships between functional group biomass and DNA abundance and leads us to suspect that DNA values may relate more to biomass turnover than total biomass. These differences also suggest that relative DNA frequencies cannot be interpreted without a proper calibration and identification of the factors affecting both DNA quantities in soil and potential PCR bias.

We need to understand better the temporal and spatial representation of vegetation by soil samples. Our analyses show that soil DNA samples collected within boreal meadow and heath plots <100 m apart are clearly distinct in terms of diversity patterns based on functionally important species. As such the vegetation turnover, or beta diversity among samples, was well depicted by soil DNA. At the temperate site, meadows that were never cultivated for crops located 1 km from formerly cultivated meadows showed no crop DNA sequences. This matches our initial expectation that locally produced biomass contributes the large majority of the soil DNA. However, wind, above- and below-ground water flow or animals can transport DNA over longer distances, and hence, the occurrence of a few sequences at a location does not necessarily imply that the species is or has been present locally. Accordingly, among the 176 283 sequences extracted from the boreal samples, we identified 301 sequence reads matching 22 plant species that were not in the local flora but likely part of a more regional species pool (Table S1, see Data accessibility).

Table 2 Comparison between the P6 loop of the *trn*L (UAA) intron and *rbc*L for assessing the plant diversity using total DNA extracted from soil samples

	Length of the PCR product (without primers)	Number of sequences in public databases (release 103 of EMBL)	Number of MOTUs identified in four samples of the boreal experiment	Number of MOTUs identified in four samples of the tropical experiment
P6 loop of the *trn*L (UAA) intron	10–143 bp (99% <100 bp)	53 207	30	106
*rbc*L	553 bp	10 997	12	14

MOTUs, molecular operational taxonomic units.

Vegetation in northern environments is mainly composed of perennial species, and change is typically slow. Given the modern (i.e. 0–50 year) age of the soil samples, and the low rate of DNA degradation in cold environments, this can explain the exceptional match of the two data sets. The analyses of the soil samples from the temperate site show that DNA can persist over decades; the frequency of crop sequences was, however, very low after 40–50 years. DNA metabarcoding can therefore offer a more integrated and longer time perspective on plant communities than, for example, revisiting previously sampled plots (e.g. Kapfer *et al.* 2011), as this approach suffers from biases such as false absences (Kéry *et al.* 2006) and annual variability (Kéry *et al.* 2006). By integrating DNA degradation rates, relationships between above-ground biomass and DNA soil frequencies, and present vegetation surveys, we may infer vegetation changes that have occurred over the last decades. For example, it might be possible to assess which plant species were dominant before human influences profoundly modified vegetation composition.

This study is the first to explore the potential of DNA metabarcoding in assaying the plant diversity at a species-rich tropical forest site. We anticipated that our technique might be faced with a serious challenge in the tropics, because it was not sure that sufficient amounts of DNA could be extracted from a moist tropical soil where rapid DNA degradation could be expected. We nevertheless found that our approach was efficient, resulting in detection of more than 200 molecular operational taxonomic units. Finally, the best match of the sequences were in virtually all cases plant species that are known to occur at the site (Table S2, see Data accessibility), demonstrating that this technique may be used for rapid assessment programmes.

One potential limitation of DNA metabarcoding using soil DNA is the constraint concerning the length of the barcode. Only very short barcodes are efficient, as demonstrated by the comparison between the short P6 loop fragment and the longer *rbc*L marker (Table 1). The taxonomic resolution of the DNA barcode used here (the *trn*L P6 loop) allowed a sound comparison of above-ground and soil DNA plant diversity. However, resolution is relatively low for some families both in the boreal and temperate sites (Poaceae, Cyperaceae, Asteraceae and Salicaceae) and in the tropics (Sapotaceae, Lauraceae, genus *Inga* in the mimosoid legumes). Analysing additional short plastid or nuclear DNA fragments could reduce or remove this limitation. For instance, at the tropical site, a preliminary analysis of a short fragment (66–107 bp) from the ITS region in the ribosomal DNA was shown to be more effective in discriminating within the Sapotaceae than any of the plastid fragments (unpublished results).

Given the expected improvement of sequencing technology (Shendure & Ji 2008), the expected increase in DNA barcoding resolution by analysing additional DNA fragments, and the possibility of identifying other groups of organisms from the same DNA extracts (e.g. bacteria, insects, fungi, nematodes), analysing soil DNA should in the near future provide an efficient and economical approach to assess states and understand changes in ecosystem structure and function across the globe.

Acknowledgements

This work is financially supported by the European Commission, under the Sixth Framework Programme (EcoChange project, contract no FP6-036866), the Norwegian Research Council ('Arctic Predators' and 'Ecosystem Finnmark' projects), the Centre National de la Recherche Scientifique (AMAZONIE research programme for the Nouragues station), and the project GACR P505/12/1296.

References

Aerts R, Chapin FS (2000) The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. *Advances in Ecological Research*, **30**, 1–67.

Albert CH, Yoccoz NG, Edwards TC *et al.* (2010) Sampling in ecology and evolution – bridging the gap between theory and practice. *Ecography*, **33**, 1028–1037.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2009) GenBank. *Nucleic Acids Research*, **37**, D26–D31.

Binladen J, Gilbert MTP, Bollback JP *et al.* (2007) The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One*, **2**, e197.

Bongers F, Charles-Dominique P, Forget PM, Théry M (eds) (2001) *Nouragues: Dynamics and Plant-Animal Interactions in a Neotropical Rainforest*. Kluwer, Dordrecht.

Bråthen KA, Hagberg O (2004) More efficient estimation of plant biomass. *Journal of Vegetation Science*, **15**, 653–660.

Bråthen KA, Ims RA, Yoccoz NG *et al.* (2007) Induced shift in ecosystem productivity? Extensive scale effects of abundant large herbivores. *Ecosystems*, **10**, 773–789.

CBoL Plant Working Group (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.

Chabrerie O, Verheyen K, Saguez R, Decocq G (2008) Disentangling relationships between habitat conditions, disturbance history, plant diversity, and American black cherry (*Prunus serotina* Ehrh.) invasion in a European temperate forest. *Diversity and Distributions*, **14**, 204–212.

Chapin FS, Bret-Harte MS, Hobbie SE, Zhong HL (1996) Plant functional types as predictors of transient responses of arctic vegetation to global change. *Journal of Vegetation Science*, **7**, 347–358.

Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. *Frontiers in Ecology and the Environment*, **8**, 233–238.

Chave J, Olivier J, Bongers F *et al.* (2008) Above-ground biomass and productivity in a rain forest of eastern South America. *Journal of Tropical Ecology*, **24**, 355–366.

Chen GK, Kery M, Zhang JL, Ma KP (2009) Factors affecting detection probability in plant distribution studies. *Journal of Ecology*, **97**, 1383–1389.

Cornelissen JHC, van Bodegom PM, Aerts R *et al.* (2007) Global negative vegetation feedback to climate warming responses of leaf litter decomposition rates in cold biomes. *Ecology Letters*, **10**, 619–627.

Gonzalez MA, Baraloto C, Engel J *et al.* (2009) Identification of Amazonian trees with DNA barcodes. *PLoS One*, **4**, e7483.

Goslar T, van der Knaap WO, Hicks S *et al.* (2005) Radiocarbon dating of modern peat profiles: pre- and post-bomb C-14 variations in the construction of age-depth models. *Radiocarbon*, **47**, 115–134.

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.

Kapfer J, Grytnes JA, Gunnarsson U, Birks HJB (2011) Fine-scale changes in vegetation composition in a boreal mire over 50 years. *Journal of Ecology*, **99**, 1179–1189.

Kéry M, Spillmann JH, Truong C, Holderegger R (2006) How biased are estimates of extinction probability in revisitation studies? *Journal of Ecology*, **94**, 980–986.

Kesanakurti PR, Fazekas AJ, Burgess KS *et al.* (2011) Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Molecular Ecology*, **20**, 1289–1302.

Kowalczyk R, Taberlet P, Coissac E *et al.* (2011) Influence of management practices on large herbivore diet-Case of European bison in Bialowieza Primeval Forest (Poland). *Forest Ecology and Management*, **261**, 821–828.

Magurran AE (2004) *Measuring Biological Diversity*. Blackwell, Oxford.

McMahon SM, Harrison SP, Armbruster WS *et al.* (2011) Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. *Trends in Ecology & Evolution*, **26**, 249–259.

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, **48**, 443–453.

Norden N, Chave J, Caubere A *et al.* (2007) Is temporal variation of seedling communities determined by environment or by seed arrival? A test in a neotropical forest. *Journal of Ecology*, **95**, 507–516.

Pélissier R, Couteron P, Dray S, Sabatier D (2003) Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. *Ecology*, **84**, 242–251.

Porazinska DL, Giblin-Davis RM, Faller L *et al.* (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources*, **9**, 1439–1450.

Ravolainen VT, Yoccoz NG, Bråthen KA *et al.* (2010) Additive partitioning of diversity reveals no scale-dependent impacts of large ungulates on the structure of tundra plant communities. *Ecosystems*, **13**, 157–170.

Rayé G, Miquel C, Coissac E *et al.* (2011) New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. *Ecological Research*, **26**, 265–276.

Sala OE, Chapin FS, Armesto JJ *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Shendure J, Ji HL (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1135–1145.

Soininen EM, Valentini A, Coissac E *et al.* (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. *Frontiers in Zoology*, **6**, 16.

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.

Stohlgren TJ (2007) *Measuring Plant Diversity*. Oxford University Press, Oxford.

Sugihara G (1982) Diversity as a concept and its measurement – comment. *Journal of the American Statistical Association*, **77**, 564–565.

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.

Thompson WL (ed.) (2004) *Sampling Rare or Elusive Species*. Island Press, Washington.

Valentini A, Miquel C, Nawaz MA *et al.* (2009a) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the *trn*L 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.

Vittoz P, Bayfield N, Brooker R *et al.* (2010) Reproducibility of species lists, visual cover estimates and frequency methods

for recording high-mountain vegetation. *Journal of Vegetation Science*, **21**, 1035–1047.

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.

Wookey PA, Aerts R, Bardgett RD *et al*. (2009) Ecosystem feedbacks and cascade processes: understanding their role in the responses of Arctic and alpine ecosystems to environmental change. *Global Change Biology*, **15**, 1153–1172.

Zinger L, Shahnavaz B, Baptist F, Geremia RA, Choler P (2009) Microbial diversity in alpine tundra soils correlates with snow cover dynamics. *Isme Journal*, **3**, 850–859.

The authors note that L.G., C.M. and P.T. are coinventors of patents related to the g/h primers and the use of the P6 loop of the chloroplast *trn*L (UAA) intron for plant identification using degraded template DNA. These patents only restrict commercial applications and have no impact on the use of this locus by academic researchers.

Data accessibility

Tables S1 and S2, raw and filtered sequence data are deposited at DRYAD entry doi: 10.5061/dryad.m346b576.

The sequence accession nos are given in Tables S1 and S2.