MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1816–1820 doi: 10.1111/j.1365-294X.2011.05317.x

Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies

PIERRE TABERLET, SOPHIE M. PRUD'HOMME, ETIENNE CAMPIONE, JULIEN ROY, CHRISTIAN MIQUEL, WASIM SHEHZAD, LUDOVIC GIELLY, DELPHINE RIOUX, PHILIPPE CHOLER, JEAN-CHRISTOPHE CLÉMENT, CHRISTELLE MELODELIMA, FRANÇOIS POMPANON and ERIC COISSAC

Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble Cedex 9, France

Abstract

DNA metabarcoding refers to the DNA-based identification of multiple species from a single complex and degraded environmental sample. We developed new sampling and extraction protocols suitable for DNA metabarcoding analyses targeting soil extracellular DNA. The proposed sampling protocol has been designed to reduce, as much as possible, the influence of local heterogeneity by processing a large amount of soil resulting from the mixing of many different cores. The DNA extraction is based on the use of saturated phosphate buffer. The sampling and extraction protocols were validated first by analysing plant DNA from a set of 12 plots corresponding to four plant communities in alpine meadows, and, second, by conducting pilot experiments on fungi and earthworms. The results of the validation experiments clearly demonstrated that sound biological information can be retrieved when following these sampling and extraction procedures. Such a protocol can be implemented at any time of the year without any preliminary knowledge of specific types of organisms during the sampling. It offers the opportunity to analyse all groups of organisms using a single sampling/extraction procedure and opens the possibility to fully standardize biodiversity surveys.

Keywords: DNA extraction, DNA metabarcoding, extracellular DNA, soil

Received 20 July 2011; revision received 5 September 2011; accepted 9 September 2011

Introduction

DNA barcoding, the use of a standardized DNA sequence to identify species, is becoming a popular solution for taxonomic identification of individual specimens. For both animals and plants, the standardized DNA fragments are longer than 500 bp (Hebert *et al.* 2003; Hollingsworth *et al.* 2009). Besides the identification of individual specimens, the barcoding concept has been extended to the identification of multiple taxa based on a single experiment. Such an approach is mainly implemented for bacteria (e.g. Sogin *et al.* 2006), fungi (e.g. Zinger *et al.* 2009a), nematodes (e.g. Porazinska *et al.* 2009), herbivore and carnivore diet studies (e.g. Valentini *et al.* 2009; Kowalczyk *et al.* 2011; Shehzad *et al.* 2012), and solving ecological questions concerning river benthos (Hajibabaei *et al.* 2011), plants Yoccoz *et al.* 2012, L. Gielly, *et al.*, in revision) and earthworms (Bienert *et al.* 2012).

The approach consisting of identifying multiple species, in a single experiment, using complex and degraded environmental samples can be termed 'DNA metabarcoding'. As soil contains DNA remains from many organisms, it is tempting to use soil DNA to assess biodiversity using a DNA metabarcoding approach. Total soil DNA includes cellular DNA originating from living cells or from living multicellular organisms, and extracellular DNA (Levy-Booth *et al.* 2007; Pietramellara *et al.* 2009). Usually, extracellular DNA originates from cell lysis and represents a significant proportion of total soil DNA (Pietramellara *et al.*

Correspondence: Pierre Taberlet, Fax: +33(0)4 76 51 42 79; E-mail: pierre.taberlet@ujf-grenoble.fr

2009). Extracellular DNA can adsorb to negatively charged silica particles, clay and organic matter via their phosphates and cation bridging (England *et al.* 2004). Adsorbed DNA is much more resistant to DNase digestion than free DNA (Romanowski *et al.* 1991). It has also been demonstrated that the amount of adsorbed DNA molecules decreases in the presence of phosphates, indicating competition between DNA and phosphate (Saeki *et al.* 2010). As a consequence, soil extracellular DNA can be selectively extracted by a saturated phosphate buffer in absence of a cell lysis step (Lorenz & Wackernagel 1987; Ogram *et al.* 1987; Corinaldesi *et al.* 2005).

In several pilot experiments for tracking plant or animal diversity using soil DNA (Yoccoz *et al.* 2012; Bienert *et al.* 2012), we observed a high heterogeneity among samples from the same location. These pilot experiments were carried out starting from three to six grams of soil using the PowerMax® Soil DNA Isolation Kit (Mo Bio Laboratories, Cambridge, UK) following manufacturer's instructions. Such extraction kits were initially developed for the analysis of microorganism DNA, and consequently, the first step of protocols consists of a cell lysis. Thus, by using such commercial kits, both intra- and extracellular DNA are extracted, but the amount of starting material is by far too low to properly assess plants or macro-invertebrate diversity. Furthermore, extracellular DNA might represent a better choice for targeting nonmicroorganism diversity, and might better integrate the local diversity over a longer period, avoiding rapid changes in microorganism composition according to seasonal or climatic parameters (Zinger *et al.* 2009b).

Owing to the local heterogeneity, it is of prime importance to extract DNA from a much larger amount of soil to represent as closely as possible the local biodiversity. Our goal was thus to develop new sampling and extraction procedures suitable for DNA metabarcoding analyses, and compatible with a large amount of starting material (several kilograms of soil). The extraction procedure we propose here is based on the use of saturated phosphate buffer, and target extracellular DNA. We also validated both the sampling and the extraction procedures on a set of 12 plots corresponding to four habitat types in alpine meadows by analysing plant DNA.

Materials and methods

Soil sampling

The goal of the sampling procedure is to obtain a soil sample as representative as possible of the local biodiversity. As plant and animal distribution is heterogeneous at a plot scale, we suggest collecting and mixing together several dozens of small core samples (20–50 g of soil per coring). These sub-samples can either be collected randomly or on a regular grid and should include material from different depths. The size of the study area should be designed to cope with the local heterogeneity. To properly monitor the variability introduced by the sampling strategy, we strongly recommend collecting at least two samples (each composed of many cores mixed together) per sampling location. The soil cores are collected with a coring sampler that is properly cleaned between each plot, by removing any soil remains before a final step at high temperature (flame cleaning). We usually collect the soil samples in wide-neck barrels (15.4 L; Cat. Number: 0789.1; Roth Sochiel E.U.R.L., Lauterbourg, France) that will be subsequently used for the DNA extraction.

Extraction of extracellular DNA

The first step is to add the saturated phosphate buffer (Na_2HPO_4; 0.12 M; pH ≈8) to the soil sample. After weighing the soil sample, we usually add the same weight of phosphate buffer. In some cases, with soil containing many organic matters, we increase the amount of phosphate buffer up to twice the weight of the soil. As it is not recommended to keep the phosphate buffer owing to possible bacterial contamination, we prepare it just before the DNA extraction. We obtain the correct pH and concentration by adding 1.97 g of NaH_2PO_4 and 14.7 g of Na_2HPO_4 per litre of sterile distilled water. After properly closing the container, the soil and the phosphate buffer are thoroughly mixed together during 15–30 min. A shorter time will reduce the concentration of extracellular DNA in the phosphate buffer, while a longer time will increase the relative proportion of PCR inhibitors (mainly humic acids). This mixing step is important to properly homogenize the sample. An aliquot of the soil/phosphate buffer mixture is then centrifuged for ten minutes at 10 000 rcf, and 500 μL of the resulting supernatant containing extracellular DNA is recovered for the next extraction steps that are carried out with a commercial kit for soil DNA (NucleoSpin® Soil; Macherey-Nagel, Düren, Germany), skipping the lysis step and following manufacturer's instructions. The DNA extract was recovered in 100 μL and diluted 10 times before being used as PCR template.

Validation experiments

Sampling for the validation experiments was performed in the central French Alps (45°2′ to 45°3′ N; 6°22′ to 6°24′ E, 2060–2700 m a.s.l). A total of 12 plots were sampled, three plots in each of four different plant commu-

nities: (i) subalpine grasslands dominated by *Festuca paniculata*, (ii) subalpine heath dominated by *Vaccinium vitis-idae* and *V. myrtillus*, (iii) low alpine meadows dominated by *Carex sempervirens* and (iv) dry high alpine meadows dominated by *Kobresia myosuroides*. Plant diversity and soil properties of plant communities differed according to altitude, aspect and snow cover duration (Choler & Michalet 2002). The geological substratum of the four communities consists mainly of carbonated flysch (sand-stone and mixed schist). In the subalpine grassland, this parent material is covered by thick calcic brown soils, while the subalpine heath and alpine meadows are located on calcic brown soils, thinner and richer in soil organic matter (Table 1). In each plot, soil sampling was carried out within a five metres radius circle, following a regular grid of 1 × 1 m, and collecting material in the first 10 cm of depth. Thus, a total of 80 soil cores were collected within this circle, with about 50 g of soil per core, and grouped together to produce the sample that will be subsequently used for the DNA extraction. This sampling procedure was repeated in each plot, leading to a total of 24 soil samples of about 4 kg each. Two DNA extractions were carried out per sample using the same weight of soil and of phosphate buffer and following the protocol given above, leading to 48 DNA extracts (four DNA extracts per plot). Finally, two DNA amplifications were carried out per extract, leading to 96 PCR products (eight PCRs per plot).

DNA amplifications were carried out in a final volume of 30 μL, using 2 μL of the diluted DNA extract. The amplification mixture contained 1 U of AmpliTaq® Gold DNA Polymerase (Applied Biosystems, Foster City, CA, USA), 10 mM Tris–HCl, 50 mM KCl, 2 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.25 μM of each primer and 0.005 mg of bovine serum albumin (BSA; Roche Diagnostic, Basel, Switzerland). The mixture was denatured at 95 °C for 10 min, followed by 45 cycles of 30 s at 95 °C and 30 s at 50 °C (no elongation step). The primer pair used (*g*: 5′-GGGCAATCCTGAGCCAA-3′, and *h*: 5′-CCATTGAGTCTCTGCACCTATC-3′) corresponds to a universal approach that targets the P6 loop region of the chloroplast *trn*L (UAA) intron (Taberlet *et al.* 2007). Both primers were modified by the addition of specific 9 bp tags on the 5′ end to allow the assignment of sequence reads to the relevant sample (Valentini *et al.* 2009). All the PCR products from the different samples were first titrated using capillary electrophoresis (QIAxcel, Qiagen GmbH, Hilden, Germany) and then mixed together, in equimolar concentrations. This mix underwent sequencing using Titanium chemistry (GS Rapid Library Prep, GS Junior Titanium emPCR Lib-L and GS Junior Titanium Sequencing Kits) on a GS Junior pyrosequencing system (Roche 454 Life Sciences, Branford, CT, USA), following manufacturer's instructions.

The sequence reads were analysed using the OBITools software (http://www.prabi.grenoble.fr/trac/OBITools). First, 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 taken into account. The amplified regions, excluding primers and tags, were kept for further analysis. Then, strictly identical sequences were clustered together using the obiuniq program, keeping the information about their distribution among the 96 amplifications. Sequences with a single occurrence were excluded using the obigrep program, as they probably correspond to amplification/sequencing errors. Each remaining sequence was further considered as a molecular operational taxonomic unit (MOTU, Blaxter *et al.* 2004), and the resulting sequence file was converted into an occurrence table (MOTUs × amplifications) with the obitab program. This sequence analysis protocol was designed to be as simple as possible for revealing the biodiversity differences among amplifications, but without further cleaning for allowing a good correspondence between MOTUs and plant species. The occur-

Table 1 Characteristics and top soil (0–10 cm) properties (mean ± SE) of the four plant communities

	Subalpine grasslands	Subalpine heath	Low alpine meadows	Dry high alpine meadows
Altitude (m a.s.l.)	2060–2210	2230–2350	2370–2560	2650–2700
Aspect	S-SE	W-NW	Flat	W-NW
Dominant species	*Festuca paniculata*	*Vaccinium* sp.	*Carex sempervirens*	*Kobresia myosuroides*
Vegetation cover (%)	95	85	95	75
Parent material	Flysch	Flysch	Flysch	Flysch
Soil organic matter content (%)	23.9 ± 1.9	33.9 ± 2.2	36.9 ± 11.3	37.1 ± 8.4
pH	6.9 ± 0.1	6.5 ± 0.2	5.8 ± 0.1	7.7 ± 0.3
C/N	12.9 ± 0.4	18.7 ± 0.0	13.1 ± 0.2	13.1 ± 0.1
$N\text{-}NH_4$ (μg/g)	15.9 ± 6.3	1.6 ± 1.6	18.2 ± 3.4	19.5 ± 4.8
$N\text{-}NO_3$ (μg/g)	1.7 ± 0.9	1.2 ± 0.3	9.1 ± 4.4	3.1 ± 0.5
N (%)	0.7 ± 0.0	0.8 ± 0.1	1.0 ± 0.1	1.1 ± 0.2

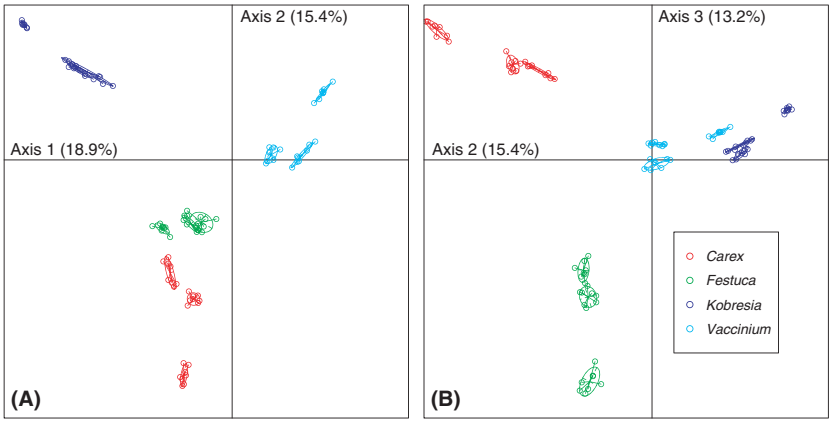

Fig. 1 Results of the validation experiments using DNA extracts from soil samples collected in 12 plots from four different alpine plant communities (low alpine meadows dominated by *Carex sempervirens*, subalpine grasslands dominated by *Festuca paniculata*, dry high alpine meadows dominated by *Kobresia myosuroides* and subalpine heath dominated by *Vaccinium* sp.). Graphics A and B represent projections of a between-class analysis on axis 1 vs. 2, and 2 vs. 3, respectively.

rence table was finally analysed using between-class analysis (BCA; Doledec & Chessel 1987) from the ade4 R package (Dray & Dufour 2007). In the between-class analysis (BCA), individuals are the 96 PCR products and variables used for the ordination are, for each sequences, the count of their occurrences in each PCR. The metric used is the chi square distance. The modalities of the discrete variable used for classification of the PCR are the 12 plots. According to this, individuals clustered when the description of the plant communities described by the occurrences of sequences is similar (same plants in same proportions).

Results of the validation experiments

We obtained a total of 83 827 usable *g/h* sequence reads as output of the ngsfilter program. This led to a total of 15 326 unique sequences, of which 3861, occurring more than once and representing 86.3 % of the initial reads, were selected for the between-class analysis. At this stage, we obtained a mean of 753 reads per amplification (minimum: 237; median: 463; maximum: 6400).

Projections over the three-first axes of the between-class analysis (Fig. 1) clearly show that the eight amplifications from the same plot group together. The plots corresponding to the same plant communities also group together, but are nevertheless slightly different in relation to known heterogeneity within each plant community.

Discussion

The sampling and the extraction protocols presented here are easy to implement both in the field and in the laboratory. The results of the validation experiments clearly demonstrated that sound biological information can be retrieved when following the sampling, extraction and analysis procedures, even without taxonomic assignation of the different MOTUs. Furthermore, our sampling/extraction procedure allows the analysis of different types of organisms using a single DNA extract. We already tested this possibility by analysing two other groups of organisms in a pilot experiment with DNA extracts from two plots. We first used the primers ewB/ewE (Bienert *et al*. 2012) amplifying a ~120 bp fragment of ice worms (Enchytraeidae) and earthworms (Lumbricina) mitochondrial 16S gene. We also amplified fungal internal transcribed spacer one (ITS1) of the nuclear ribosomal DNA using specific primers (L. Epp, S. Boessenkool, E. Bellemain, *et al*., in revision). In both experiments carried out specifically for this study and using the extraction protocol presented here, we obtained the expected sequences (data not shown).

The sampling/extraction protocol could be adjusted according to the question and to the target taxonomic groups. The size of the sampling area should be representative of the local diversity of the target groups. For example, to analyse tree diversity in a forest, one hectare should be appropriate. We advise adjusting the size of the sampling area according to the target taxonomic group having the coarsest spatial heterogeneity. In the same way, if the target taxonomic group corresponds to animals only living above ground, the soil sampling should only concentrate on the top soil layer, without sampling deepest layers.

The procedure presented here fulfils the needs of metabarcoding analyses using soil extracellular DNA, at least when using soils containing a high proportion of organic matter. Such a protocol can be implemented at any time of the year, provided that soil is accessible, and without having to look for specific types of organisms during the sampling. It offers the opportunity to analyse all groups of organisms using a single sampling/extraction procedure. Finally, it also opens the possibility to standardize biodiversity surveys; lack of such a standardization represents the main difficulty for carrying out large-scale meta-analyses.

Acknowledgements

This work is financially supported by the European Commission, under the Sixth Framework Programme (EcoChange project, contract no FP6-036866).

References

Bienert F, de Danieli S, Miquel C *et al.* (2012) Tracking earthworm communities from soil DNA. *Molecular Ecology*, **21**, doi: 10.1111/j.1365-294X.2011.05407.x.

Blaxter M, Elsworth B, Daub J (2004) DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **271**, S189–S192.

Choler P, Michalet R (2002) Niche differentiation and distribution of *Carex curvula* along a bioclimatic gradient in the southwestern Alps. *Journal of Vegetation Science*, **13**, 851–858.

Corinaldesi C, Danovaro R, Dell'Anno A (2005) Simultaneous recovery of extracellular and intracellular DNA suitable for molecular studies from marine sediments. *Applied and Environmental Microbiology*, **71**, 46–50.

Doledec S, Chessel D (1987) Seasonal successions and spatial variables in fresh-water environments. 1. Description of a complete 2-way layout by projection of variables. *Acta Oecologica-Oecologia Generalis*, **8**, 403–426.

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

England LS, Vincent ML, Trevors JT, Holmes SB (2004) Extraction, detection and persistence of extracellular DNA in forest litter microcosms. *Molecular and Cellular Probes*, **18**, 313–319.

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, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, S96–S99.

Hollingsworth PM, Forrest LL, Spouge JL *et al.* (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.

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

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

Lorenz MG, Wackernagel W (1987) Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. *Applied and Environmental Microbiology*, **53**, 2948–2952.

Ogram A, Sayler GS, Barkay T (1987) The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*, **7**, 57–66.

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

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.

Romanowski G, Lorenz MG, Wackernagel W (1991) Adsorption of plasmid DNA to mineral surfaces and protection against DNase-I. *Applied and Environmental Microbiology*, **57**, 1057–1061.

Saeki K, Kunito T, Sakai M (2010) Effects of pH, ionic strength, and solutes on DNA adsorption by andosols. *Biology and Fertility of Soils*, **46**, 531–535.

Shehzad W, Riaz T, Nawaz MA *et al.* (2012) Carnivore diet analysis based on next-generation sequencing: application to the leopard cat (*Prionailurus bengalensis*) in Pakistan. *Molecular Ecology*, **21**, doi: 10.1111/j.1365-294X.2011.05424.x.

Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored ''rare biosphere''. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12115–12120.

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

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

Yoccoz NG, Bråthen KA, Gielly L *et al.* (2012) DNA in soil mirrors plant functional and structural diversity. *Molecular Ecology*, **21**, in press.

Zinger L, Coissac E, Choler P, Geremia RA (2009a) Assessment of microbial communities by graph partitioning in a study of soil fungi in two alpine meadows. *Applied and Environmental Microbiology*, **75**, 5863–5870.

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

P.T., L.G., and C.M. 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

Data of the validation experiments: DRYAD entry doi:10.5061/dryad.ck8pg.