MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2017–2030 doi: 10.1111/j.1365-294X.2011.05407.x

Tracking earthworm communities from soil DNA

FRIEDERIKE BIENERT,* SÉBASTIEN DE DANIELI,† CHRISTIAN MIQUEL,* ERIC COISSAC,* CAROLE POILLOT,* JEAN-JACQUES BRUN† and PIERRE TABERLET*

**Laboratoire d'Ecologie Alpine, CNRS-UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble Cedex 9, France,*
†Mountain Ecosystems Research Unit, Cemagref Grenoble, 2, Rue de la Papeterie, BP 76, 38402 Saint-Martin-d'Hères, France

Abstract

Earthworms are known for their important role within the functioning of an ecosystem, and their diversity can be used as an indicator of ecosystem health. To date, earthworm diversity has been investigated through conventional extraction methods such as handsorting, soil washing or the application of a mustard solution. Such techniques are time consuming and often difficult to apply. We showed that combining DNA metabarcoding and next-generation sequencing facilitates the identification of earthworm species from soil samples. The first step of our experiments was to create a reference database of mitochondrial DNA (mtDNA) 16S gene for 14 earthworm species found in the French Alps. Using this database, we designed two new primer pairs targeting very short and informative DNA sequences (about 30 and 70 bp) that allow unambiguous species identification. Finally, we analysed extracellular DNA taken from soil samples in two localities (two plots per locality and eight samples per plot). The two short metabarcode regions led to the identification of a total of eight earthworm species. The earthworm communities identified by the DNA-based approach appeared to be well differentiated between the two localities and are consistent with results derived from inventories collected using the handsorting method. The possibility of assessing earthworm communities from hundreds or even thousands of localities through the use of extracellular soil DNA will undoubtedly stimulate further ecological research on these organisms. Using the same DNA extracts, our study also illustrates the potential of environmental DNA as a tool to assess the diversity of other soil-dwelling animal taxa.

Keywords: cytochrome *c* oxidase, DNA metabarcoding, extracellular soil DNA, mitochondrial 16S ribosomal gene, mitochondrial DNA, species identification

Received 27 June 2011; accepted 26 October 2011.

Introduction

Constituting up to 80% of total soil–animal biomass, earthworms are known for their important role in ecosystem functioning (Edwards & Bohlen 1996; Bardgett 2005). Their diversity can be used as an indicator for ecosystem health (e.g. Buckerfield *et al.* 1997; Römbke *et al.* 2005).

To date, earthworm inventories have been accomplished through passive and behavioural methods (Edwards & Bohlen 1996). Passive methods require the physical separation of earthworms from the soil through handsorting or soil washing (Bouché 1972; Edwards & Bohlen 1996). Behavioural methods use a physical or chemical stimulus to extract the worms. Such stimuli include the application of heat (Čoja *et al.* 2008), electricity (Schmidt 2001; Čoja *et al.* 2008; Eisenhauer *et al.* 2008), vibrations (Catania 2008; Mitra *et al.* 2009) and mustard (Chan & Munro 2001; Bartlett *et al.* 2006; Eisenhauer *et al.* 2008) or AITC (Alloisothiocyanate) solution (Zaborski 2003; Čoja *et al.* 2008; Pelosi *et al.* 2009). As yet, there is no universal and efficient method to realistically estimate the biomass and abundance of earthworm communities (Bartlett *et al.* 2006; Čoja *et al.* 2008). Biases such as physicochemical soil parameters (Čoja *et al.* 2008; Eisenhauer *et al.* 2008), earthworm activity (Callaham & Hendrix 1996), life

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

stage (Bartlett *et al.* 2006; Čoja *et al.* 2008; Pelosi *et al.* 2009) and spatial distribution (Zaborski 2003; Bartlett *et al.* 2006) influence the efficacy of all extraction procedures. The combination of different methods such as handsorting and washing (Edwards & Bohlen 1996), or handsorting in combination with the application of AITC solution (Bartlett *et al.* 2006; Čoja *et al.* 2008; Pelosi *et al.* 2009), seems to give the best estimates of earthworm communities. However, these methods are time consuming and sometimes difficult to apply because of logistic constraints (Čoja *et al.* 2008).

Recent conceptual and technical developments have the potential to revolutionize how earthworm communities are assessed. First, the use of a short DNA sequence, the DNA barcode, has transformed species identification (Valentini *et al.* 2009a). This approach offers a reliable earthworm identification method (e.g. Pop *et al.* 2003; Chang *et al.* 2009; but see King *et al.* 2008; Rougerie *et al.* 2009; James *et al.* 2010; Dupont *et al.* 2011 for a discussion of the potential pitfalls). Second, dramatic improvements in sequencing technologies allow the production of millions of sequence reads in a single experiment (Shendure & Ji 2008; Glenn 2011). The metabarcoding approach, used in conjunction with next-generation sequencers, should make it possible to implement a new high-throughput methodology to assess earthworm diversity using soil DNA, that is, extracellular earthworm DNA left behind in soil (Minamiya *et al.* 2011). Such DNA is highly degraded because of metabolic activities in the soil (Nielsen *et al.* 2007). As a consequence, we must concentrate on a short and variable DNA fragment that is likely to discriminate between different earthworm species. This variable fragment must be flanked by conserved regions; this will allow the design of primers that amplify most earthworm species. Previous DNA barcoding studies of earthworms have focused mainly on the mitochondrial cytochrome *c* oxidase subunit 1 (CO1) gene (Huang *et al.* 2007; King *et al.* 2008; Rougerie *et al.* 2009; Chang *et al.* 2009; Richard *et al.* 2010). The standardized DNA fragment coding for CO1 (658 bp; Hebert *et al.* 2003) is too long to use for amplifying highly degraded DNA from soils (Taberlet *et al.* in press; P. Taberlet, E. Coissac, F. Pompanon, C. Brochmann & E. Willerslev, unpublished), and the COI minibarcode (130 bp; Meusnier *et al.* 2008) does not work with earthworms (data not shown). Therefore, we will focus on mitochondrial 16S rDNA, which allows the design of very short metabarcodes because of the alternations of several conserved and variable regions.

The objective of this study was to adjust and validate an innovative methodology for assessing earthworm communities using soil DNA. More specifically, our goal was first to design new versatile primer pairs amplifying very short variable regions of the 16S gene (mtDNA) and then to test these short metabarcodes *in silico* and empirically. Finally, we compared the results obtained with inventories collected using traditional methods (i.e. handsorting, soil washing and chemical extraction). Another objective was to apply these techniques beyond earthworms and to investigate the possibility of using environmental DNA to analyse the whole spectrum of local biodiversity. This means that the same DNA extracts could potentially be used to assess the diversity of all organisms living in soil, including microorganisms, plants and animals.

Materials and methods

General strategy

Figure 1 presents our general strategy. As described in step 1 of that figure, our goal was first to create a reference database for mtDNA 16S genes using universal primers (ewA and ewF, see Table 1 and Fig. 2) taken from 70 earthworms that had been manually collected from different sampling sites in the Grenoble region. Individual earthworms had been identified through both morphological characteristics and cytochrome *c* oxidase subunit I (COI) barcodes.

Two new pairs of internal primers were subsequently designed (ewB/ewC and ewD/ewE; Table 1) in accordance with the 16S database. These primers target very short DNA sequences (about 30 and 70 bp, respectively), which allows efficient amplification from soil samples. We tested them using an *in silico* PCR (details below), before carrying out empirical experiments that produced sequence data from soil samples (step 2). Finally, we analysed the sequence data and compared the results with those derived from handsorting inventories (step 3).

Study sites and sampling

Earthworms analysed for the reference database were collected from numerous habitats in the Grenoble region (Chartreuse massif, Isère valley). Earthworms were collected manually and, because of logistical constraints, without any application of reagents. They were stored in a 70% ethanol solution before DNA extraction.

The first study site is situated in the Chartreuse massif in a dell near a village called 'les Cottaves' and route D57b [Geographic coordinates; Amont (Plot 1) Longitude: 05°46′59″ E, Latitude: 45°18′48″ N–Aval (Plot 2) Longitude: 05°46′52″ E, Latitude: 45°18′45″ N]. The second is on the Campus de Grenoble (Isère, France) (Geographic coordinates; Isère (Plot 3) Longitude:

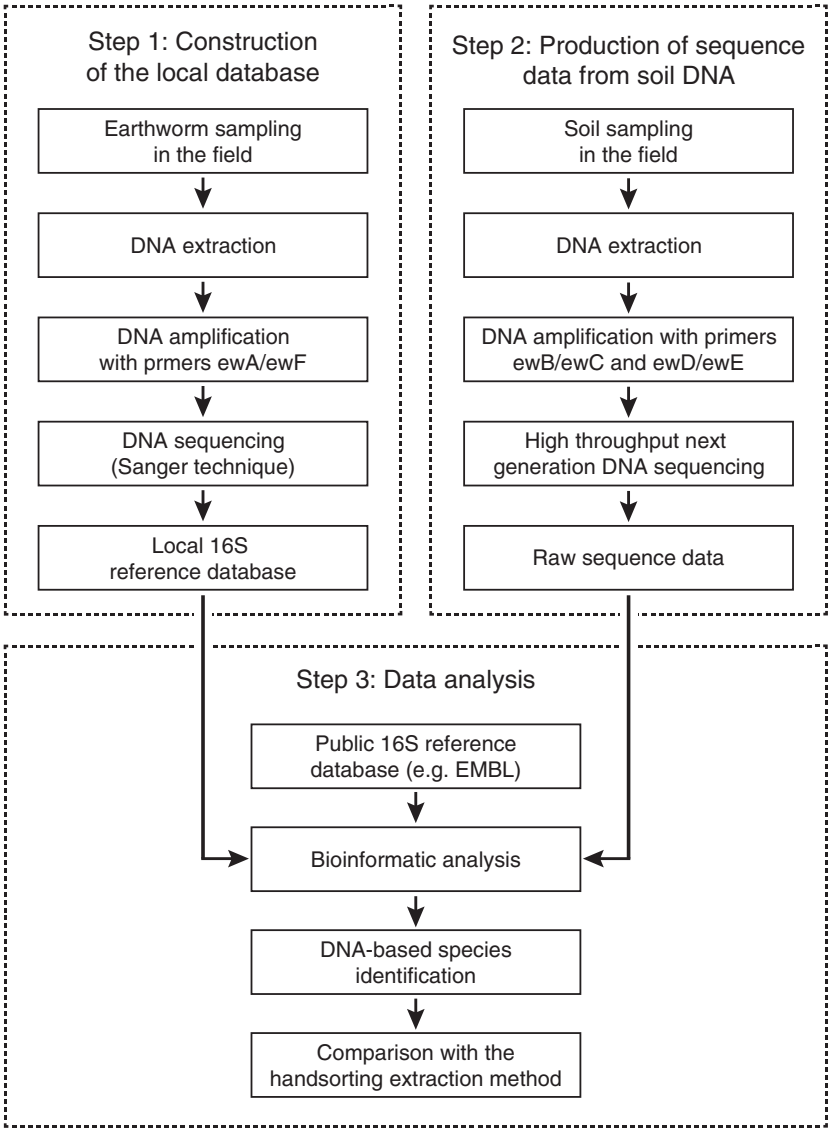

Fig. 1 Methodology applied for analyzing earthworm biodiversity from soil samples.

Table 1 Primers used in this study. The code reflects the position of the 3′ nucleotide in the *Lumbricus terrestris* mitochondrial DNA reference sequence (accession no.: U24570)

Name	Code	Length	Sequence 5′–3′
ewA	F11639	20	CGACTGTTTAACAAAAACAT
ewB	F11859	21	CAAGAAGACCCTATAGAGCTT
ewC	R11891	17	GGTCGCCCCAACCGAAT
ewD	F11907	17	ATTCGGTTGGGGCGACC
ewE	R11982	21	CTGTTATCCCTAAGGTAGCTT
ewF	R12121	22	CGCGGTCTGAACTCAGCTCATG

05°46′26″ E, Latitude: 45°11′59″ N–Vercors (Plot 4) Longitude: 05°46′26″ E, Latitude: 45°11′56″ N). The first site is in undisturbed woodland on cambisols that includes fir and beech trees (*Abies alba* and *Fagus sylvatica*). The environment of the second study site is a maintained meadow on fluvisols and dominated by *Lolium perenne* and *Festuca rubra*.

Handsorting inventories (Bouché 1972) were conducted by digging out earthworms on 0.5-m^2 sample plots without any application of reagents. The individuals collected from each sample plot were identified based on morphology. Because morphology-based identification of juveniles from the *Aporrectodea* genus is difficult, and many juveniles were collected, all of the individuals of this genus were grouped together and recorded as *Aporrectodea* sp.

The DNA-based approach was implemented in the same sampling sites that were used for the handsorting inventories. In each sampling site (Grenoble and

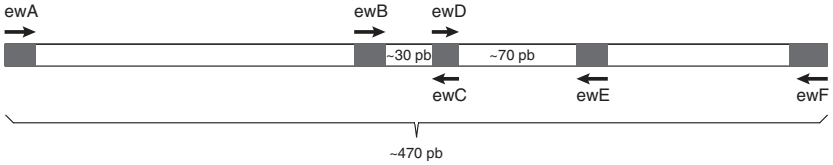

Fig. 2 Schema of 16S DNA sequence in earthworms showing variable (white) regions, conserved (grey) regions and positions of the primers designed for this study.

Chartreuse), two sampling spots were chosen for replication. The sampling procedure consisted of collecting four surface (0–20 cm depth) and four sub-surface (20–40 cm depth) soil layer samples per spot, contained within a radius of 10 m. We collected about 0.5 kg of soil per sample using a soil corer. We applied precautionary measures in the field to avoid cross-contamination between soil samples; this included sterilizing all material (drill, sieve, etc.) with a high-temperature flame. Soil sampling was carried out in January (Grenoble) and March (Chartreuse).

16S reference databases from earthworm tissues

All DNA extractions were performed in a room dedicated to nucleic acid extractions. Total DNA was extracted from about 25 mg of earthworm tissue with the QIAmp Tissue Kit (Qiagen GmbH), following the manufacturer's instructions. Mock extractions without samples were systematically performed to monitor for possible contaminations. The success of DNA extraction was confirmed by agarose gel electrophoresis (2%).

DNA amplifications of a 467–472 bp of mitochondrial 16S gene were carried out in a final volume of 25 μL, using 3 μL of DNA extract as template. The amplification mixture contained 0.6 U of AmpliTaq® Gold DNA Polymerase (Applied Biosystems), 10 mM Tris-HCl, 2 mM of MgCl2, 0.1 mM of each dNTP, 1 μM of each primer (ewA/ewF; Table 1 and Fig. 2) and 250 μg/mL of bovine serum albumin (BSA; Roche Diagnostic). The primers ewA and ewF were designed for this study. They target the exact homologous sequences of the universal primers for 16S (Palumbi 1996) but have been slightly modified according to the complete *Lumbricus terrestris* mitochondrial DNA sequence (Boore & Brown 1995). The mixture was denatured at 95 °C for 10 min, followed by 35 cycles of 30 s at 95 °C and 30 s at 50 °C. The elongation step was carried out at 72 °C for 60 s. Each PCR product was purified using the QIAquick PCR Purification Kit (Qiagen GmbH), and DNA was quantified by agarose gel electrophoresis (2%).

Earthworm species identifications for building the 16S reference database were secured by using the standardized DNA barcoding method (Hebert *et al.* 2003). The mitochondrial COI sequence was amplified using the universal primers HCO2198 (5′–TAAACTTCAGGGTGACCAAAAAATCA–3′) and LCO1490 (5′–GGTCAACAAATCATAAAGATATTGG–3′) (Folmer *et al.* 1994), generating DNA sequences of 658 bp. Initial denaturation was realized at 95 °C during 10 min for polymerase activation. Then, denaturation, hybridation and elongation were carried out at, respectively, 95 °C for 30 s, 52 °C for 45 s and 72 °C for 60 s. A total of 40 cycles were completed. Each PCR product was purified using the QIAquick PCR Purification Kit (Qiagen GmbH), and DNA was quantified by agarose gel electrophoresis (2%).

The sequencing reactions were performed as follows. We created a mixture that contained 4 μL of big dyes v3.1, 2 μL of a 10 μM solution of each primer (ewA/ewF for 16S gene fragments and HCO2198/LCO1490 for COI gene fragments), which were added separately in two different microplates. The mixture also contained the purification product, the amount of which varied according to its intensity on the agarose gel, and ultra-high-quality (UHQ) water to adjust DNA concentrations. Denaturation, hybridation and elongation were carried out at, respectively, 96 °C for 30 s, 50 °C for 30 s and 60 °C for 4–5 min. A total of 25 cycles were completed. Before electrophoresis, a purification step was performed using gel filtration on Sephadex G50/Sephacryl S200 columns. Electrophoresis was carried out using the 3130 XL Genetic Analyser (Applied Biosystems).

16S and COI sequences were corrected and aligned using SeqScape v2.5 (Applied Biosystems) and BioEdit v7.0.9.0 (Hall 1999) for correction and Seaview 4 (Galtier *et al.* 1996) for alignment. Species determination was based on morphological identification and on the comparison of CO1 sequences with those from the Barcode of Life Data Systems (BOLD; Ratnasingham & Hebert 2007) containing reference libraries for earthworms collected during a recent DNA barcoding campaign (Rougerie *et al.* 2009).

Design and in silico evaluation of short metabarcodes

The alignment of the 16S ewA/ewF sequences revealed three highly conserved regions flanking variable segments, which allowed the design of two versatile pairs

of primers that generate short sequences for application on soil samples (Table 1 and Fig. 2). The *in silico* efficiency of these primers and of their respective metabarcodes have been tested against all sequences available in public databases (EMBL, release 107). This was carried out using the ecoPCR program (Ficetola *et al.* 2010), applying the following parameters: length of amplified sequences from 5 to 200 bp, perfect match between the primer sequence and the target sequence for the two last nucleotides on the 3′ end of each primer and a maximum of three mismatches on the remaining sequence of the primers.

Extracellular soil DNA analysis

Eight DNA extractions corresponding to the eight samples collected per plot were carried out. For each sample, DNA was extracted from approximately 6 g of wet soil. We used the PowerMax® Soil DNA Isolation Kit (Mo Bio Laboratories) following the manufacturer's instructions. Eight DNA amplifications were performed per plot (one per sample), plus two amplifications using a mix of the extracts corresponding to the 0–20 and 20–40 cm depth samples, respectively. Thus, a total of 10 DNA amplifications were conducted per plot. DNA amplifications were realized in a final volume of 50 μL using 4 μL of DNA extract. The mixture contained 0.6 U/tube of AmpliTaq® Gold DNA Polymerase (Applied Biosystems), 2 mM $MgCl_2$, 0.2 μM of each dNTPs, 250 μg/mL of bovine serum albumin (BSA; Roche Diagnostic), 0.2 μM of each primer (Sigma) and finally UHQ water to bring each sample to the final volume. Initial denaturation was realized at 95 °C during 10 min for polymerase activation. Then, denaturation, hybridation and elongation were carried out, respectively, at 95 °C for 30 s, 58 °C for 30 s and 72 °C for 60 s. A total of 50 cycles were completed. Each sample was independently amplified with a pair of ewB/ewC and ewD/ewE primers (Fig. 2). The primers were modified by the addition of specific tags on the 5′ end to allow sequence reads to be assigned to the relevant sample (Valentini *et al.* 2009b). As a consequence, all of the PCR products were tagged identically on both ends. These tags were composed of CC on the 5′ end followed by nine variable nucleotides that were specific to each sample. The nine variable nucleotides were designed using the *oligoTag* program (http://www.prabi.grenoble.fr/trac/OBITools) with at least three differences among the tags, without homopolymers longer than two and avoiding a C on the 5′ end. All the PCR products from the different samples were first titrated using capillary electrophoresis (QIAxel; QIAgen GmbH) and then mixed together in equimolar concentration before the sequencing.

The sequencing was carried out on the Illumina/Solexa Genome Analyzer IIx (ILMN), using the Paired-End Cluster Generation Kit V4 and the Sequencing Kit V4 (ILMN), and following 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 program (http://www.prabi.grenoble.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 data quality during both the alignment and the consensus computation. Then, primers and tags were identified using the *ngsfilter* program. Only sequences with a 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. Strictly identical sequences were clustered together using the *obiuniq* program, keeping the information about their distribution among samples. Sequences shorter than 20 bp for the ewB/ewC fragment and 50 bp for the ewD/ewE fragment or containing ambiguous nucleotides or with occurrence lower or equal to 100 were excluded using the *obigrep* program. Taxon assignments were performed using following two methods. The first method used the *ecoTag* program (Pegard *et al.* 2009). *EcoTag* relies on an exact global alignment algorithm (Needleman & Wunsch 1970) to find highly similar sequences in the locally constituted reference database for our both markers. The second method is based on BLAST (Altschul *et al.* 1990, 1997) and on the whole nonredundant database (NR) through the public NCBI web site (http://blast.ncbi.nlm.nih.gov/). Only BLAST matches spanning the whole query sequence and with at least 95% of identity were considered. A taxon was assigned to each unique sequence. This unique taxon corresponds to the last common ancestor node in the NCBI taxonomic tree of all the taxids annotating the sequences of the reference database that matched against the query sequence. Automatically assigned taxonomic identifications were then manually curated to further eliminate those few sequences that were the probable result of PCR artifacts and those that did not correspond to any 16S sequences present in the reference database or in public databases. Finally, rarefaction curves per plot were drawn for both locations and for both fragments (ewB/ewC and ewD/ewE), showing the number of earthworm species detected according to the number of samples. We considered an earthworm species to be present if the number of its sequence reads was greater than 100 per sample. As the number of samples per plot was only eight, we estimated the number of species detected using an exhaustive approach, scanning all possible sample subsets with a cardinality ranging from one to eight.

Results

The reference database

The morphological determination of earthworm species is based on the taxonomy of the 'Museum National d'Histoire Naturelle' according to the 'Inventaire National du Patrimoine Naturel' (TAXREF v3.0; http://inpn.mnhn.fr/). Five of 70 individuals have not been identified to the species level. Overall, results from sequencing of the cytochrome *c* oxidase subunit one region were reliable. Nine of the 70 samples were excluded from the data set because of ambiguous electropherograms. Low sequence quality was observed for *Allolobophora icterica* and *Lumbricus friendi*. Each fragment has been compared using BLAST (Altschul *et al*. 1997) on the National Centre for Biotechnology Information (NCBI) website (http://blast.ncbi.nlm.nih.gov/) and using the default settings on the Consortium for the Barcode of Life (CBoL) (http://www.boldsystems.org/views/idrequest.php) website. Species identifications with a percentage of sequence similarity below 95% were considered to be insignificant (see e.g. Chang *et al*. (2009) for examples of intra- and interspecific genetic distances). CO1 sequence data allowed us to refine the morphological identifications of juvenile individuals, and when alcohol preservation prevented the analysis of morphological characteristics. We identified 14 earthworm taxa to the species level. However, for one tissue sample, the similarity to any available sequence in public database was less than 90%, with the closest matches being to several species of the *Aporrectodea* genus; this sample was thus identified as *Aporrectodea* sp.

Results of the 16S DNA sequencing using the ewA/ewF primer pair produced high sequence quality with fragment lengths ranging from 467 to 472 bp. Using public databases, we confirmed that ewB, ewC, ewD and ewE primers were conserved in all earthworms worldwide, with only slight variations in some species (not affecting the 3'-end of the primer). Clustering of ewB/ewC and ewD/ewE sequences provided an excellent discrimination between species, with a single exception: *Aporrectodea longa* and *Aporrectodea nocturna* have the same ewD/ewE metabarcode. We even observed intraspecific variation for five species, *Ap. icterica*, *Aporrectodea rosea*, *Lumbricus castaneus*, *Lumbricus rubellus* and *Lumbricus terrestris* (Table 2).

In silico assessment of the primers ewB/ewC and ewD/ewE and of their amplified metabarcodes

The four primers amplifying the ewB/ewC and ewD/ewE metabarcodes match extremely well with their target sequences, achieving no mismatches in a large majority of cases (97.19%, 96.00%, 95.99% and 89.75% for ewB, ewC, ewD and ewE, respectively). Of the 2698 primer/target sequence matches involving the four primers, we obtained 2556, 136, 5 and 1 cases with 0, 1, 2 and 3 mismatches, respectively.

For the ewB/ewC fragment, we obtained an *in silico* amplification for 676 earthworm (Lumbricina) sequences, with 361 unique sequences corresponding to 257 species and 66 genera. For the ewD/ewE fragment, we obtained 673 earthworm (Lumbricina) sequences, with 408 unique sequences corresponding to 254 species and 62 genera. The distribution of these unique sequences among the different Lumbricina families is given in Table 3.

Soil DNA analysis

We obtained a total of 12 711 955 and 2 611 172 aligned sequences with perfect matches on tags, and with no ambiguous nucleotides, which corresponds to 421 745 and 161 672 unique sequences for ewB/ewC and ewD/ewE, respectively. After removing sequences that corresponded to PCR artifacts (sequences shorter than expected) and sequences occurring less than 100 times in the whole dataset, we end up with 848 and 474 unique sequences for ewB/ewC and ewD/ewE, respectively. Only those sequences with proper species identification were taken into account. Sequence similarity had to be >95% when compared with the NCBI and the 16S reference databases. Finally, to further limit the impact of PCR and sequencing errors, the sequences that fell below the limit of 1/10 000 for ewB/ewC and 7/10 000 for ewD/ewE of the most common sequence were been excluded from the data set. As a consequence, only sequences with occurrences higher than 400 and 100 were kept for ewB/ewC and ewD/ewE, respectively.

Results from soil DNA analysis are displayed in Table 4 for the ewB/ewC metabarcode and in Table 5 for the ewD/ewE metabarcode. In these tables, the number of reads corresponds to the sum of all reads observed per plot for the ten DNA amplifications. It is interesting to note that the results provided by each extraction/amplification are relatively heterogeneous. When combining all the results of the 10 per plot amplifications, species identification using ewB/ewC and ewD/ewE primers provides almost identical results (Table 6). Eight species have been determined from both primer pairs: *Allolobophora chlorotica*, *Aporrectodea cupulifera*, *Ap. icterica*, *Ap. longa*, *Ap. rosea*, *Lumbricus terrestris*, *Octolasion cyaneum* and *Octolasion tyrtaeum*. Additionally, we obtained another MOTU (Molecular Operational Taxonomic Unit) for the ewB/ewC

Table 2 List of the two short metabarcodes (regions ewB/ewC and ewD/ewE) from 16S mtDNA for the 15 earthworm taxa analysed in this study to built a local reference database

Species	Metabarcode ewB/ewC	Metabarcode ewD/ewE
Allolobophora chlorotica (1) h1	CATTTTAATAAAGATATAGACTTTACTAAA	AGGGAAACCCAATCATCCCTAACTAAAAGATAAATAAATCTCATTACTGACC CTTAATTAAGATCACAAAAAC
Aporrectodea caliginosa (13) h1	TATTTTAATAAAAAATATAAATTTTTAATAA	AGGGAAGTTACTAATCATCCCTAAATAAAAGATTAATTAATCTAAATTCTGA CCCTTATTCAAGATCAATAGATC
Aporrectodea cupulifera (2) h1	CATTTTAATAAAAATTATAAATTTTACTAAA	TGGGAAGTAATTAATCATCCCACTATAACAGATAAATAAATCTACCCTCTGA CCCTTAATAAAGATCATCAAAAT
Aporrectodea icterica (3) h1	CATCTTAATGAAGACTAAAACTTCACTAAA	GGGGAAAATCAATCATCCCTAATAAATAGATAAATAAATCTAAACGCTGACC CTTAACAAAGATCACCAGAAT
Ap. icterica (2) h2	CATCTTAATGAAGACTAAAACTTCACTAAA	GGGGAAAATCAATCATCCCTAATAAATAGATAAGTAAATCTAAACACTGACC CTTAACAAAGATCACCAGAAT
Aporrectodea longa (1) h1	TATTTTAACAAAAACCCAAAAATTTTCAATAAA	AGGGAATATAAATCATCCCCTATAAACAGATAACTTAATCTAGACTCTGACC CTTAAACAAGATCAACAAATC
Aporrectodea nocturna (3) h1	TATTTTAACAAAATCCAAAAATTTTCAATAAA	AGGGAATATAAATCATCCCCTATAAACAGATAACTTAATCTAGACTCTGACC CTTAAACAAGATCAACAAATC
Aporrectodea rosea (4) h1	CATTTTAATTAAAACTGACTATTTAAATAAG	AGGGAAAAATCAATCATCCCCTACAAAAAGATTAATTAATCTACTTTATGAC CCTTAATTAAGATCACCAGAAC
Ap. rosea (1) h2	CATTTTAATTAAAACTGACTATTTAAATAAG	AGGGAAAATCAATCATCCCCTACAAAAAGACTAATTAATCTACTTTATGACC CTTAATTAAGATCACCAGAAC
Ap. rosea (3) h3	CATTTTAATAAAAACTAATATTTTAATAAA	GGGGAAGACCAATCATCCCCTACAAAAAGATTAATTAATCTTCTCTATGACC CTTAATTAAGATCACCAGAAT
Ap. rosea (3) h4	CATTTTAATAAAAACCTATATTTTAATAAG	GGGGAAATATCCATCATCCCTTACAAAAAGATTAATTAATCTCCTGTATGAT CCTTAATTAAGATCACCAGAAT
Aporrectodea sp. (1) h1	TATTTTAATAAAAACTTATATTTTTAATAA	AGGGAAATTTCAAATCATCCCTCAATAAAATATAAATTAATCAAACCTCGGA CCCTTAATCAAGATCAACAGATC
Lumbricus castaneus (1) h1	AATTTAAACAAATATAAAAAATTTACTAAA	GAGGAAATGACTATCATCCTCAGATAAAAGATAAATAAATCTAACTAATGAC CCTTAACTAAGATCACTAAAAC
L. castaneus (1) h2	AATTTAAACAAATATAAAAAATTTACTAAA	GAGGAAATAACTATCATCCTCAGATAAAAGATAAATAAATCTAACTAATGAC CCTTAACTAAGATCACTAAAAC
Lumbricus friendi (3) h1	AATTAAATTAAATACTAAAAATTTACTAAA	AGGGAAATTACTTATCATCCCTAAACTAAAGATAAATCAATCAAAATTCTGA CCCTTATACAAGATCATTAAATC
Lumbricus meliboeus (2) h1	AATTTAAACAAATACTAAAAATTTACTAAA	TGGGAATATCAATCATCCCATGATAAAAGATAAATTTATCTAGCTACTGACC CTTATTTAAGATCACCAAAAC
Lumbricus rubellus (4) h1	AATTAAAATAAATACCTATAATTTAAATAA	AGGGAAATTATAAATCATCCCTAATTAAAAAGATAAATTAATCTGACTACTG ACCCTTAATTAAGATCACTAAAAC
L. rubellus (4) h2	AATTAAAATAAATACCTAAAATTTACTAAA	AGGGAAATCATTTATCATCCCTAAACAAAAGATAAATTAATCCAACTATTGA CCCTTAATTAAGATCACTAAAAC
Lumbricus terrestris (1) h1	AATTTAAATAAATGTAAAAAATTTACTAAA	AGGGAATTACTTATCATCCCTAAACAAAAGATAAGTATATCAAAATACTGAC CCTTCTACAAGATCACTAAAAC
Lumbricus terrestris (2) h2	AATTTAAATAAATGTAAAAAATTTACTAAA	AGGGAATTACTTATCATCCCTAAACAAAAGATAAATATATCAAAATACTGAC CCTTCTACAAGATCACTAAAAC

Table 2 *Continued*

Species	Metabarcode ewB/ewC	Metabarcode ewD/ewE
Lumbricus terrestris (1) h3	AATTTAAATAAATGTAAAAAATTTACTAAA	AGGGAATTACTCATCATCCCTAAGCAAAAGATAATTATATCAAAATACTGACCCT TCTACAAGATCATTTAAAAC
Lumbricus terrestris (1) h4	AATTTAAATAAATATAAAAAATTTACTAAA	AGGGAATTACTCATCATCCCTAAGCAAAAGATAAATATGTCAAAATACTGACCCT TCCACAAGATCATTTAAAAC
Octodrilus juvyi (3) h1	CATTTTTAACAAATTTATTAAATTTATTAAG	AAGGAAAATAAATCATCCTGCTAAATATAGATTAATTAATCTAACTTCTGACCCTT AATAAAGATCATTTAAAAC
Octolasion cyaneum (10) h1	CATTTTTAATAGAAGCTTACTATTCTAATTAAA	CGGGAAACACACATCATCCCGCAAAATTAGATAAATACATCCTTATATTGACCCT TAGTTAAGATCAAAAGAAC

The number of specimens with the same metabarcodes and the haplotype (h) is indicated after the scientific name. The ewA/ewF sequences containing these two short metabarcodes have been uploaded in EMBL database under the accession nos HE611648–HE611676.

Table 3 Results of the *in silico* amplification experiment using all data from public database (EMBL release 107)

Family	Region ewB/ewC	Region ewD/ewE
Acanthodrilidae	67	66
Aeolosomatidae	0	0
Almidae	1	1
Eudrilidae	0	0
Glossoscolecidae	1	1
Hormogastridae	54	54
Lumbricidae	119	120
Megascolecidae	415	412
Microchaetidae	0	0
Ocnerodrilidae	5	5
Octochaetidae	3	3
Unidentified family	11	11
Total	676	673

Number of retrieved sequences from the different earthworm families for the two DNA fragment BC and DE. Aeolosomatidae, Eudrilidae and Microchaetidae do not show any retrieved sequences because EMBL (release 107) does not contain any 16S homologous sequences from these three families.

fragment, which we identified as *Ap. longa* or *Ap. nocturna* (recorded as *Ap.* sp. in Table 6). This sequence matches perfectly with an *Ap. longa* sequence in public database (accession no.: FJ967632) but has a single mismatch with both *Ap. longa* (h2) and *Ap. nocturna* from our reference database. This ewB/ewC fragment could correspond to the unidentified *Lumbricidae* MOTU ewD/ewE fragment (Table 5). Rarefaction curves showing the number of earthworm species detected according to the number of samples are shown in Fig. 3.

Handsorting inventory and comparison with the DNA approach

Results of handsorting inventories are given in Table 6, with species identification based only on morphological characteristics. The handsorting inventories are consistent with the DNA-based method, except for *L. friendi* and probably for *L. castaneus,* which were missed by the DNA-based approach, and for *A. chlorotica* and *O. tyrtaeum*, which were missed by the handsorting method.

Discussion

The DNA approach for tracking earthworm communities delivered extremely promising results. Using extracellular DNA from the soil and very short metabarcodes, we were able to properly describe earthworm communities from two sampling locations. It is remarkable that earthworm diversity can be revealed by

Table 4 Fifteen most abundant mtDNA 16S sequences obtained in two different sampling spots in the Chartreuse massif (Amont, Aval) and in Grenoble (Isère, Vercors) and the corresponding number of sequence reads obtained in the two locations. These sequences were amplified with the ewB/ewC primers targeting earthworms DNA in soil samples

		Location			
		Chartreuse		Grenoble	
Sequence BC	Species	Amont	Aval	Isère	Vercors
CATCTTAATGAAGACTAAAACTTCACTAAA	*Aporrectodea icterica*	836 954	649 677	834 031	1 359 355
TATTTTAACAAAAACCCAAAAATTTTCAATAAA	*Aporrectodea longa*	2	6	244 463	271 829
CATTTTAATAAAAATTATAAATTTTACTAAA	*Aporrectodea cupulifera*	0	0	236 024	236 678
CATTTTAATAGAAGCTTACTATTCTAATAAA	*Octolasion cyaneum*	468 462	3823	0	2
TATTTTAATAAAATAGTAAATTTTACTAAA	Unidentified	334 804	96 337	0	1
TATTATAAATCAATTAATAATTGAGCATA	Unidentified	0	372 828	0	0
AATTTAAATAAATATAAAAAATTTACTAAA	*Lumbricus terrestris*	0	0	174 286	143 682
CATTTTAATAGAAAAATAATATCCTAATAAA	*Octolasion tyrtaeum* (P)	306 476	0	0	2
TATCACAATATTTATACAATAAATATTATG	*Achaeta unibulba*; Enchytraeidae (P)	183 116	68 615	0	0
TATTTTTCTTATACTTTAGTAAACAAAAA	Unidentified	96 924	42 148	0	0
AATTTAAATAAATATAAAAAAATTTACTAAA	*Lumbricus terrestris /castaneus* (1 mismatch)	0	0	56	131 001
TATTTTAACAAAACCCAAAAATTTTCAATAAA	*Aporrectodea longa* (P) or *longa/nocturna* (1 mismatch)	2469	105 312	159	145
CATTTTAATAAAGATATAAACTTTACTAAA	*Allobophora chlorotica* (P) (1 mismatch)	0	0	51 953	43 196
TATTTTATTTACCTAAAACAGTAACAAAA	*Marionina communis*; Enchytraeidae (P)	0	0	62 901	0
TATTTTTCTTATACTTTAGTAAATAAAAA	Unidentified	592	61 802	0	0

(P) indicates that the identification was done with public databases. All other identifications rely on the reference database built for this study.

so short metabarcodes and by analyzing only less than 50 g of soil per sampling plot (i.e. <100 g per sampling location).

The two short metabarcodes (regions ewB/ewC and ewD/ewE) were designed using only mtDNA 16S sequences from earthworms sampled in the Grenoble region. However, the *in silico* assessment of the two primer pairs showed that our method should work worldwide, for virtually all earthworms (*Lumbricina*). The four primers target highly conserved regions, with more than 90% of the cases registering no mismatches between the primers and the target sequences. Both the *in silico* approach and the reference database from the Grenoble region showed that the two amplified regions have a surprisingly high variability, not only between, but also within species (Table 2). As a consequence, our DNA-based approach has the potential to properly identify the large majority of earthworm species around the globe.

Overall, the two metabarcodes ewB/ewC and ewD/ewE (Tables 4 and 5) provided consistent results when taking into account only records with a high number of sequence reads. The 2 and 6 reads obtained in the Chartreuse locality for *Ap. longa* (Table 4) clearly correspond to amplification or sequencing artifacts of the most common MOTU (identified as *Ap. longa* with public database, or as *Ap. longa/nocturna* with our reference database). These two MOTUs are identical, except for an A homopolymer of 4 bp in the MOTU with high frequency, and 5 bp in the rare MOTU. In the same way, the 159 and 145 reads obtained in the Grenoble locality are artifacts from the 244 463 and 271 829 reads, with an homopolymer of 5 bp in the most common MOTU, and of 4 bp in the rare MOTU. As a consequence, MOTUs with a low frequency correspond in most of the cases to experimental artifacts, and cannot be interpreted, even if they perfectly match with a sequence in a reference database. This is a drawback of using very short metabarcodes: a single-artifactual nucleotide can lead to an incorrect species identification. However, there is no alternative to using very short metabarcodes when working with highly degraded extracellular DNA from soil. Another potential difficulty is posed by the incompleteness of the mitochondrial 16S reference databases. At the moment,

Table 5 Twenty most abundant mtDNA 16S sequences obtained in two different sampling spots in the Chartreuse massif (Amont, Aval) and in Grenoble (Isère, Vercors) and the corresponding number of sequence reads obtained in the two locations. These sequences were amplified with the ewD/ewE primers targeting earthworms DNA in soil samples

		Location			
		Chartreuse		Grenoble	
Sequence DE	Species	Amont	Aval	Isère	Vercors
GGGGAAAATCAATCATCCCTAATAAATAGATAAGTAAATCTAAACACTGACCCT TAACAAAGATCACCAGAAT	*Aporrectodea icterica*	49 437	73 767	11 200	0
GGGGAAAATCAATCATCCCTAATAAATAGATAAATAAATCTAAACGCTGACCCT TAACAAAGATCACCAGAAT	*Ap. icterica*	217	263	8246	76 395
TGGGAAGTAATTAATCATCCCACTATAACAGATAAATAAATCTACCCTCTGACC CTTAATAAAGATCATCAAAAT	*Aporrectodea cupulifera*	0	0	35 761	48 456
TAGGAAAAATCATCCTTTATAAATAAGATCTACACATCACCCATCAGACCCTTA TCCAAGATCAAAGAATC	Unidentified Enchytraeidae	2	60 656	0	0
TGGGAAGTAATTAATCATCCTACTTTAATAGATAAATAAATCTTCTTTTTGACC CTTAATCAAGATCACTAAAAC	Unidentified Lumbricidae	45 962	13 292	0	2
AGGGAATTACTCATCATCCCTAAGCAAAAGATAAATATGTCAAAATACTGACCC TTCTACAAGATCATTAAAAC	*Lumbricus terrestris* (1 mismatch)	0	0	17 499	38 824
CAGGATAACTCATCCTGTATATCAAAGACAAATATGTCAATTTATATGAACCTT TACAGATCACAGATC	Unidentified Enchytraeidae	46 308	5985	0	0
AGGGAATATAAATCATCCCCTATAAACAGATAACTTAATCTAAACTCTGACCCT TAAACAAGATCAACAAATC	*Aporrectodea longa* (P)	0	0	1127	51 098
AAGGATAAATCATCCTTTATAATTAGATTAACAAATCATACTAATGATCCTATA CTAGATCTTCGAATC	Unidentified Enchytraeidae	2570	42 661	0	0
TTGGATAAATCATCCATAATAAATAAGACTAACTAGTCAAACTATAGATCCTTA TAAGATCACAGAACC	Unidentified Enchytraeidae	7640	24 265	0	0
CGGGAAACACATATCATCCCGCAAAATCAGATAAATAAATCCTCATATTGACCC TTAGTTAAGATCAAAAGAAC	*Octolasion cyaneum* (3 mismatches)	29 482	0	0	2
AGGGAATTACCCATCATCCCTAAGCAAAAGATAAATATATCAAAATACTGACCC TTCTACAAGATCATTAAAAC	*Lumbricus terrestris/castaneus* (2 mismatches)	0	0	0	21 102
AAGGATAAATCATCCTTTAAACTAAAGATATACCCATCAATCCATAGATCCATA TATGATCACAGAAGC	Unidentified Enchytraeidae	19 338	1637	0	0
TAAGAAAATTCATCTTAAAAAAAAAGATTAACTAATCACAATTAAAGATCCAAA TATGATCAACGAACT	Unidentified Enchytraeidae	20 763	6	0	0
AAGGATAAATCATCCTTTAAATAAAAGATCAATTAATCAATCCAAAGAGCTTAT ACAATAAACTCCCAAATT	Unidentified Enchytraeidae	19 329	0	0	0
AAGGATAAATCATCCTTAATAATAAAGATTTATTAATCTTATAAAAGAACTTAT TAACTAAACTCCCAAATA	Unidentified Enchytraeidae	15 678	2459	0	0
GGGGAAATATCCATCATCCCTTACAAAAAGATTAATTAATCTCCTGTATGATCC TTAATTAAGATCACCAGAAT	*Aporrectodea rosea*	0	17 017	0	0

Table 5 *Continued*

Sequence DE	Species	Location			
		Chartreuse		Grenoble	
		Amont	Aval	Isère	Vercors
AGGGAAGTTACTAATCATCCCTAAATAAAAGATTAATTAATCTAAATTCTGACC CTTATTCAAGATCAATAGATC	*Aporrectodea caliginosa*	0	16 062	0	0
CGGGAATCAAATAATCATCCCGCAAAACCAGATAAATAAATCTCCATAATGACC CTTAATTAAGATCAAGAGAAC	*Octolasion tyrtaeum* (P)	14 430	0	0	0
AAGGATAAATCAACCTATACTACATGAGATATACATATCTTTCCAATGATCCTA CTTTAGATCACCGAACT	Unidentified Enchytraeidae	0	13 300	0	0

(P) indicates that the identification was done with public databases. All other identifications rely on the reference database built for this study. The number of mismatches is indicated when the sequence does not perfectly match with the reference sequence.

less than 700 relevant mitochondrial 16S sequences, corresponding to about 250 species, are available in public databases (for a total of about 5000 earthworm species). Both the ewB/ewC and ewD/ewE primer pairs also amplify some *Enchytraeidae* (Tables 4 and 5), despite that the target region of primers ewC and ewD is not conserved in this other group of *Oligochaeta*. In complement to the ewB/ewC and ewD/ewE primer pairs, it would be interesting to try to amplify a longer fragment by using ewB/ewE primers, which produce an amplified fragment of about 120 bp (without primers). The disadvantage would be that, if the target DNA were highly degraded, a strong bias towards the less degraded template would be observed. The advantage would be that the two ewB/ewC and ewD/ewE fragments would be associated, leading to a better species identification. However, the use of ewB/ewE primers would also lead to the amplification of virtually all *Enchytraeidae*, as the two target regions are very well conserved.

The DNA-based inventories missed *L. friendi* (and probably *L. castaneus*), while the traditional handsorting missed *Allobophora chlorotica* and *O. tyrtaeum*. It is interesting to note that *L. friendi* and *L. castaneus* are two epigeic species, while *Al. chlorotica* and *O. tyrtaeum* are two endogeic species. This means that, using the experimental protocol implemented in this study, the DNA-based approach is better at detecting endogeic species, but can miss epigeic species when compared with handsorting. In fact, manual digging can influence earthworm behaviour, induce fleeing (Drewes 1984) and thus lead to underestimates of the diversity of endogeic species. The failure to detect two epigeic species by the DNA-based approach can easily be explained by our sampling protocol: only four samples per plot were taken at a depth of 0–20 cm, but without trying to collect the first few centimetres of soil. Taking into account the relative heterogeneity of the results obtained for each sample/extraction/amplification, and accounting for the tendency of the DNA-based approach to miss epigeic species, we can give some recommendations for the soil sampling step. In our study, we took only eight samples per plot. Our results (Tables 4–6) and the rarefaction curves (Fig. 3) suggest that a better representation of the actual earthworm diversity can be achieved by taking more samples. These samples should include surface soil to increase the probability of detecting epigeic species, as well as samples of several other types taken at different depths. Furthermore, the extraction method used does not specifically target extracellular DNA and is more appropriate for studying microorganisms because of a cell lysis step at the beginning of the procedure. An extraction method targeting only extracellular DNA and allowing

Table 6 Comparison between the DNA-based approach and the traditional handsorting method for assessing earthworm diversity

	Chartreuse			Grenoble		
	DNA (no. of sequence reads)			DNA (no. of sequence reads)		
Species	ewB/ewC	ewD/ewE	Handsorting (no. of individuals)	ewB/ewC	ewD/ewE	Handsorting (no. of individuals)
Allobophora chlorotica	—	—	—	95 149	3918	—
Aporrectodea cupulifera	—	—	—	472 702	84 217	5
Aporrectodea icterica	1 486 631	123 684	—	2 193 386	95 841	—
Aporrectodea longa	—	—	—	516 596	52 225	—
Aporrectodea rosea	2106 (?)	17 017	—	—	—	—
Aporrectodea sp.	107 789	—	32	—	—	9
Lumbricus castaneus	—	—	—	(?)	(?)	4
Lumbricus friendi	—	—	7	—	—	—
Lumbricus terrestris	—	—	—	449 025	77 425	116
Octolasion cyaneum	472 285	29 482	12	—	—	—
Octolasion tyrtaeum	306 476	14 430	—	—	—	—

homogenization of several kilograms of soil as a starting material (Taberlet *et al.* in press) would be much more efficient for tracking earthworm communities. Finally, to cope with highly diluted template DNA, it would be a valuable exercise to carry out several amplifications per extraction (multiple-tube approach; Navidi *et al.* 1992; Taberlet *et al.* 1996).

At the moment, a limiting factor for implementing large-scale experiments using our method is the only partial knowledge of earthworm molecular diversity, leading to incomplete mitochondrial 16S reference databases. Thanks to the rapid development of sequencing technology and to the DNA repository maintained by the Consortium for the Barcode of Life (http://

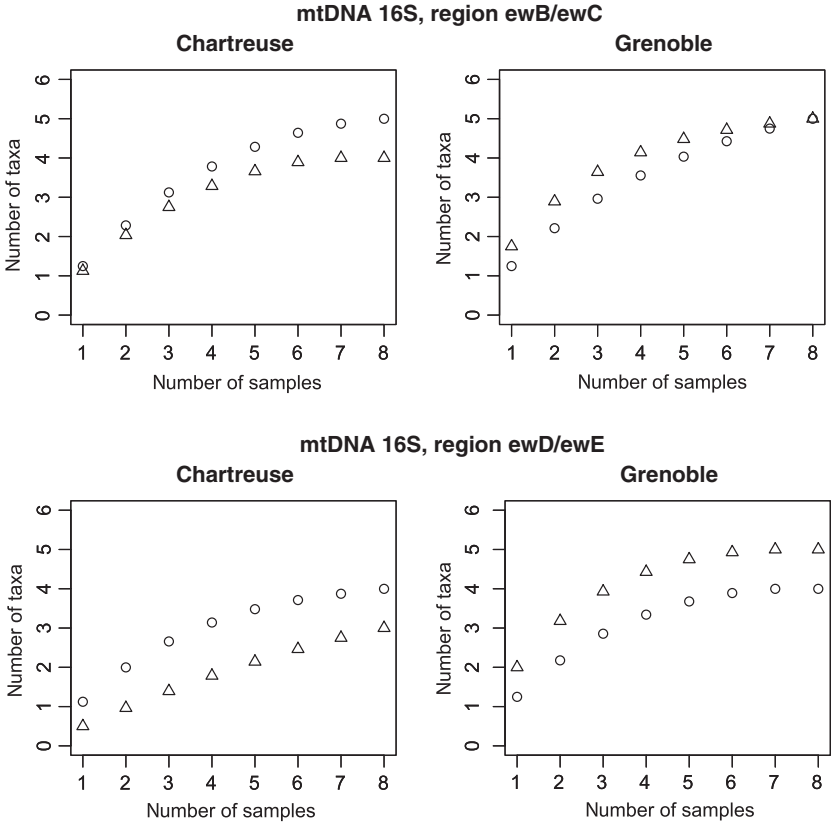

Fig. 3 Rarefaction curves of the number of earthworm species detected according to the number of soil samples collected per plot. For 'Chartreuse' location, circles correspond to 'amont' and triangles to 'aval'; for 'Grenoble' location, circles correspond to 'Vercors' and triangles to 'Isère' (see Table 4).

www.barcodeoflife.org), we can assume that the available data on earthworm molecular diversity will increase in the near future, not only for the standardized COI barcode (Hebert *et al*. 2003), but also for the two short 16S fragments.

Traditional approaches for earthworm inventories are currently restricted by time-consuming fieldwork and the difficulties of morphological identification, especially of juveniles. A DNA-based approach circumvents this problem (Richard *et al*. 2010) and allows the identification of DNA remains from all development stages, from eggs to adults. Better reference databases, together with the possibility of assessing earthworm communities from hundreds or even thousands of localities using extracellular soil DNA, will undoubtedly stimulate further ecological research on these organisms.

Finally, looking beyond earthworms, our study also illustrates the potential of environmental DNA to assess the whole spectrum of biodiversity. The crucial advantage of our 'DNA metabarcoding' approach is that different taxonomic groups, including archaea, bacterial, fungi, plants and animals, can be analysed using the same DNA extracts. The potential of environmental DNA is impressive as, in addition to soil-dwelling organisms, the technique can even detect birds and mammals living above the soil (Andersen *et al*. 2012).

Acknowledgements

This study has been financially supported by a grant from the 'Fondation pour la Recherche sur la Biodiversité (FRB)'. We thank Marc Robinson-Rechavi and Mark Reynolds for rereading our manuscript.

Conflicts of interest

F. B., C. M., E. C. and P. T. are co-inventors of a pending French patent application on the ewB, ewC, ewD and ewE primers and on the use of the amplified fragments for identifying earthworms from environmental samples. This patent only restricts commercial applications and has no impact on the use of this method by academic researchers.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Altschul SF, Madden TL, Schäffer AA *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Andersen K, Bird KL, Rasmussen M *et al*. (2012) Metabarcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, **21**, doi: 10.1111/j.1365-294X.2011.05261.x.

Bardgett RD (2005) *The Biology of Soil: A Community and Ecosystem Approach*. Oxford University Press, Oxford, 254 p.

Bartlett MD, Harris JA, James IT, Ritz K (2006) Inefficiency of mustard extraction technique for assessing size and structure of earthworm communities in UK pasture. *Soil Biology and Biochemistry*, **38**, 2990–2992.

Boore JL, Brown WM (1995) Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. *Genetics*, **141**, 305–319.

Bouché MB (1972) *Lombriciens de France: Ecologie et Systématique*. Institut National de la Recherche Agronomique, Paris, 671 p.

Buckerfield JC, Lee KE, Davoren CW, Hannay JN (1997) Earthworms as indicators of sustainable production in dryland cropping in southern Australia. *Soil Biology & Biochemistry*, **29**, 547–554.

Callaham MA, Hendrix PF (1996) Relative abundance and seasonal activity of earthworms (Lumbricidae and Megascolidae) as determined by hand-sorting and formalin extraction in forest soils on the southern Appalachian piedmont. *Soil Biology & Biochemistry*, **29**, 317–321.

Catania KC (2008) Worm grunting, fiddling, and charming—humans unknowingly mimic a predator to harvest bait. *PLoS ONE*, **3**, e3472.

Chan KY, Munro K (2001) Evaluating mustard extracts for earthworm sampling. *Pedobiologia*, **45**, 272–278.

Chang CH, Rougerie R, Chen JH (2009) Identifying earthworms through DNA barcodes: pitfalls and promise. *Pedobiologia*, **52**, 171–180.

Čoja T, Zehetner K, Bruckner A, Watzinger A, Meyer E (2008) Efficacy and side effects of five sampling methods for soil earthworms (Annelida, Lumbricideae). *Ecotoxicology and Environmental Safety*, **71**, 552–565.

Drewes CD (1984) Escape reflexes in earthworms and other annelids. In: *Neural mechanisms of startle behavior* (ed. Eaton RC), pp. 43–91. Plenum, New York.

Dupont L, Lazrek F, Porco D *et al*. (2011) New insight into the genetic structure of the *Allobophora chlorotica* aggregate in Europe using microsatellite and mitochondrial data. *Pedobiologia*, **54**, 217–224.

Edwards CA, Bohlen PJ (1996) *Biology and Ecology of Earthworms*. Chapman and Hall, London, 426 p.

Eisenhauer N, Straube D, Scheu S (2008) Efficiency of two widespread non-destructive extraction methods under dry soil conditions for different ecological earthworm groups. *European Journal of Soil Biology*, **44**, 141–145.

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

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Computer Applications in the Biosciences*, **12**, 543–548.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759–769.

Hall TA (1999) BioEdit: a user—friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

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

Huang J, Xu Q, Sun ZJ, Tang GL, Su ZY (2007) Identifying earthworms through DNA barcodes. *Pedobiologia*, **51**, 301–309.

James SW, Porco D, Decaëns T *et al.* (2010) DNA barcoding reveals cryptic diversity in *Lumbricus terrestris* L., 1758 (Clitellata): resurrection of *L. herculeus* (Savigny, 1826). *PLoS ONE*, **5**, e15629.

King RA, Tibble AL, Symondson WOC (2008) Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, **17**, 4684–4698.

Meusnier I, Singer GAC, Landry JF *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.

Minamiya Y, Ohga K, Hayakawa H, Ito K, Fukuda T (2011) Coelomic fluid: a noninvasive source of DNA in earthworms. *Molecular Ecology Resources*, **11**, 645–649.

Mitra O, Callaham MA, Smith ML, Yack JE (2009) Grunting for worms: seismic vibrations cause Diplocardia earthworms to emerge from the soil. *Biology Letters*, **5**, 16–19.

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tube approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics*, **50**, 347–359.

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.

Nielsen KM, Johnson PJ, Bensasson D, Daffonchio D (2007) Release and persistence of extracellular DNA in the environment. *Environmental Biosafety Research*, **6**, 37–53.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular systematics* (eds Hillis DM, Moritz C and Mable BK), pp. 205–247. Sinauer & Associates Inc, Sunderland, Massachusetts.

Pegard A, Miquel C, Valentini A *et al.* (2009) Universal DNA based methods for assessing the diet of grazing livestock and wildlife from faeces. *Journal of Agricultural and Food Chemistry*, **57**, 5700–5706.

Pelosi C, Bertrand M, Capowiez Y, Boizard H, Roger-Estrade J (2009) Earthworm collection from agricultural fields: comparisons of selected expellants in presence/absence of hand—sorting. *European Journal of Soil Biology*, **45**, 176–183.

Pop AA, Wink M, Pop VV (2003) Use of 18S, 16S rDNA and cytochrome c oxidase sequences in earthworm taxonomy (Oligochaeta, Lumbricidae). *Pedobiologia*, **47**, 428–433.

Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Richard B, Decaens T, Rougerie R *et al.* (2010) Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding. *Molecular Ecology Resources*, **10**, 606–614.

Römbke J, Jänsch S, Didden W (2005) The use of earthworms in ecological soil classification and assessment concepts. *Ecotoxicology and Environmental Safety*, **62**, 249–265.

Rougerie R, Decaens T, Deharveng L *et al.* (2009) DNA barcodes for soil animal baxonomy. *Pesquisa Agropecuaria Brasileira*, **44**, 789–802.

Schmidt O (2001) Appraisal of the electrical octet method for estimating earthworm populations in arable land. *Annals of Applied Biology*, **138**, 231–241.

Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nature Biotechnology*, **26**, 1035–1045.

Taberlet P, Griffin S, Goossens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **26**, 3189–3194.

Taberlet P, Prud'homme SM, Campione E *et al.* (2012) Soil sampling and extraction of extracellular DNA from large amount of soil for metabarcoding studies. *Molecular Ecology*, **21**, doi: 10.1111/j.1365-1294X.2011.05317.x.

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

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

Zaborski ER (2003) Allyl isothiocyanate: an alternative chemical expellant for sampling earthworms. *Applied Soil Ecology*, **22**, 87–95.

F.B. is a former master student, and the results presented here correspond to the main output of her master thesis. S.D.D and J.J.B are interested in soil fauna, as bio-indicators of ecosystem changes. C.M., C.P. and P.T. are mainly interested in developing new DNA-based methodologies for studying biodiversity. E.C. is specialized in bioinformatics applied to biodiversity research.

Data accessibility

COI DNA sequences: EMBL accessions HE611677–HE611701. 16S DNA sequences: EMBL accessions HE611648–HE611676. Data of the soil experiment: DRYAD entry doi: 10.5061/dryad.3mv8v434.