MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2010) **10**, 606–614 doi: 10.1111/j.1755-0998.2009.02822.x

DNA BARCODING

Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding

B. RICHARD,* T. DECAËNS,* R. ROUGERIE,† S. W. JAMES,‡ D. PORCO† and P. D. N. HEBERT†
Laboratoire d'Ecologie, EA 1293 ECODIV, FED SCALE, Bâtiment IRESE A, UFR Sciences et Techniques, Université de Rouen, F-76821 Mont Saint Aignan cedex, France, †Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, 50 Stone Road East, Guelph, ON, N1G 2W1 Canada, ‡Natural History Museum and Biodiversity Institute, Kansas University, 1345 Jayhawk Blvd., Lawrence, KS 66045, USA

Abstract

Species identification of earthworms is usually achieved by careful observation of morphological features, often sexual characters only present in adult specimens. Consequently, juveniles or cocoons are often impossible to identify, creating a possible bias in studies that aim to document species richness and abundance. DNA barcoding, the use of a short standardized DNA fragment for species identification, is a promising approach for species discrimination. When a reference library is available, DNA-based identification is possible for all life stages. In this study, we show that DNA barcoding is an unrivaled tool for high volume identification of juvenile earthworms. To illustrate this advance, we generated DNA barcodes for specimens of *Lumbricus* collected from three temperate grasslands in western France. The analysis of genetic distances between individuals shows that juvenile sequences unequivocally match DNA barcode clusters of previously identified adult specimens, demonstrating the potential of DNA barcoding to provide exhaustive specimen identification for soil ecological research.

Keywords: DNA barcoding, earthworms, juveniles, species identifications

Received 13 September 2009; revision received 15 November 2009; accepted 23 November 2009

Introduction

Studies on the biology and ecology of earthworms often depend on species diagnoses. Species identifications in this group require minute examinations of external and/or internal morphology of adults, necessitating the involvement of expert taxonomists. In many cases, identifications are complicated by the lack of stable, easily scored diagnostic characters or by environmentally induced variability in morphological features. In addition, many characters important in species diagnosis involve the position and structure of the clitellum and the associated tubercular pubertatis (Bouché 1972; Sims & Gerard 1999), characters, which are only observable in sexually mature specimens. As a result of their lack of diagnostic characters, the identification of juveniles of closely related species (e.g. members of the genus *Lumbricus*) is impossible in most cases. As a result, taxonomists can only provide generic identifications for juveniles, hampering soil studies for species richness evaluation. Attempts to characterize earthworm species based on electrophoresis analysis – as developed for instance by Bogh (1992) – revealed higher resolution for species identification in samples containing many immature stages or incomplete body fragments, but the method remained largely ignored by soil biologists and has never been developed as a routine, which could be used to process the high number of individuals that an ecological study often comprises.

More recently, DNA barcoding has emerged as a promising standardized approach for rapid species identifications in taxonomically complex groups. It uses a 658 bp fragment of the mitochondrial gene cytochrome-c oxidase I (COI) as a standard DNA tag for species discrimination and identification in the animal kingdom (Hebert *et al.* 2003). The effectiveness of DNA barcodes in species identification has been shown in varied taxonomic groups (Hebert *et al.* 2004a,b; Ward *et al.* 2005; Smith *et al.* 2007; Borisenko *et al.* 2008) and they are increasingly being used in species descriptions as well (Decaëns & Rougerie 2008; Martinez *et al.* 2008; Vaglia *et al.* 2008;

Correspondence: Thibaud Decaëns, Tel: 33 + (0)2 3276 9442; Fax: 33 + (0)2 3514 6655; E-mail: thibaud.decaens@univ-rouen.fr

Gibbs 2009). Their potential use in ecological studies has been discussed by Valentini *et al.* (2009). In the domain of soil biodiversity studies, DNA barcoding has been proposed as a promising approach to resolve the strong taxonomic impediment (Decaëns *et al.* 2006, 2008; Rougerie *et al.* 2009). Confirming the early results of DNA data for earthworm taxonomy (Perez-Losada *et al.* 2005), recent applications of DNA barcoding in earthworms are very promising (Huang *et al.* 2007; Chang *et al.* 2009) and the assembly of a comprehensive reference library for this group has been recently undertaken.

The genus *Lumbricus* is a temperate genus represented in France by nine species (Bouché 1972). All of them are morphologically very similar and the key distinctive features are the adult size and the exact position of the clitellum. These two characters are not applicable to juvenile individuals, which are consequently almost impossible to identify at the species level (Bouché 1972). In this study, we demonstrate and emphasize the unique potential of DNA barcoding as a mean to reliably identify juvenile worms of the genus *Lumbricus*.

Material and methods

The study was carried out in Haute-Normandie, France between February and March 2008. Earthworms were sampled from seven grassland and six forest sites located within the three main landscape units of the region: the plateau area north of the Seine River, the chalky slopes of the valley, and the floodplain of the Seine River (Table 1 and Fig. 1). Depending on the sampling site, soils were NEOLUVISOLs and LUVISOLs with acidic pH (4.0–6.0) developed on loess material; RENDOSOLs with neutral pH (7.0–8.0), aggregated structure and fast organic matter turnover; and HISTOSOLs with a neutral pH (6.0–8.0) underlain by peat layers and hydromorphic horizons whose depth depends on topography (INRA 1998) (Table 1).

In each sampling site, three points located 15 m apart from each other were sampled by combining formalin extraction and hand sorting: (i) 10 L of 0.4% formaldehyde were applied to a 1 m^2 area and (ii) a soil section of $25 \times 25 \times 25$ cm was dug out in the middle of the area and hand sorted 15 min after formaldehyde application. This method is recommended in temperate soils to allow the best estimations of earthworm populations (Baker & Lee 1993). The efficiency of formaldehyde extraction is in fact known to vary significantly depending on the species, and hand sorting is thus carried out to recover those specimens that remain in the soil. All collected earthworms were stored alive and killed in 50% ethanol in the laboratory.

Specimens of the genus *Lumbricus* were separated from others and adults were morphologically identified to species using Sims & Gerard (1999). Barcoding was

Table 1 Details of collecting sites. Soil classification according to INRA (1998), landscape units after Decaëns *et al.* (2008)

Exact site	GPS coordinates		Sampling date	Soil	Landscape unit	Habitat	Labels
	Latitude	Longitude					
Mont Saint Aignan, campus of the University of Rouen	N49°27′28.1″	E001°04′14.5″	5 April 2007	RENDOSOL	Plateau	Grassland	PGr1
Mont Saint Aignan, campus of the University of Rouen	N49°27′32.2″	E001°04′37.0″	11 & 14 February 2008	RENDOSOL	Plateau	Grassland	PGr2
Yvetot, 'Lycée Agricole'	N49°36′33.4″	E000°44′29.6″	31 March 2008	NEOLUVISOL	Plateau	Grassland	PGr3
Mont Saint Aignan, campus of the University of Rouen	N49°27′26.8″	E001°04′10.0″	6 June 2007	RENDOSOL	Plateau	Forest	PFo1
Saint Adrien	N49°22′26.0″	E001°07′46.5″	21 February 2008	RENDOSOL	Plateau	Forest	PFo2
Eawy forest	N49°41′25.8″	E001°16′35.4″	26 February 2008	LUVISOL	Plateau	Forest	PFo3
Eawy forest	N49°41′43.5″	E001°17′43.2″	26 March 2008	LUVISOL	Plateau	Forest	PFo4
Saint Adrien	N49°22′17.4″	E001°07′51.3″	20 February 2008	RENDOSOL	Chalky slope	Grassland	CGr1
Henouville	N49°29′01.3″	E000°55′49.8″	17 March 2008	RENDOSOL	Chalky slope	Grassland	CGr2
Mont Saint Aignan, campus of the University of Rouen	N49°27′28.1″	E001°04′14.5″	8 June 2007	RENDOSOL	Chalky slope	Forest	CFo1
Marais Vernier, 'Réserve des Mannevilles'	N49°26′01.4″	E000°30′36.6″	5 March 2008	HISTOSOL	Floodplain	Grassland	FGr1
Marais Vernier, 'Réserve des Mannevilles'	N49°25′42.9″	E000°31′09.8″	2 April 2008	HISTOSOL	Floodplain	Grassland	FGr2
Marais Vernier, 'Réserve des Mannevilles'	N49°25′29.5″	E000°31′05.2″	2 April 2008	HISTOSOL	Floodplain	Forest	FFo1

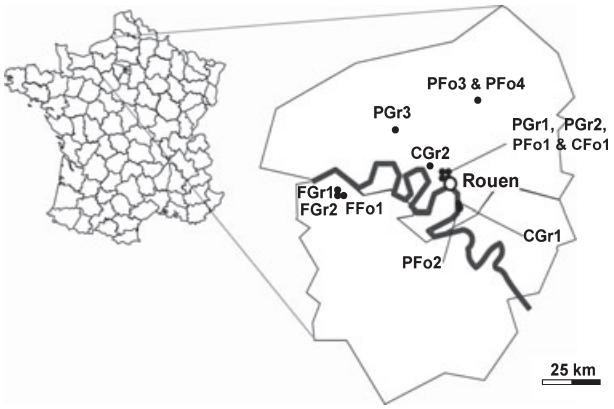

Fig. 1 Location map of the sampling sites in Haute-Normandie, France. See Table 1 for the detailed site data.

performed on a subset of these specimens. In each sample, up to three adult specimens per identified species and three juveniles were taken randomly. For each of them, a small (about 1 mm^2) sample of tissue was cut from the caudal segments and stored in a small volume of 100% ethanol. The remainder of each specimen was fixed in 4% formaldehyde for 48 h and stored individually in 100% ethanol. Vouchers were deposited in the ECODIV laboratory with a unique identifier (sampleID) linking it to the tissue sample.

Lysis of the tissues was carried out in 50 μL volume of lysis buffer and proteinase K incubated at 56 °C overnight. DNA extraction followed a standard automated protocol on 96-well glass fibre plates (Ivanova *et al.* 2006). The 5′ region of COI used as a standard DNA barcode was amplified using M13 tailed primers LCO1490 and HCO2198 (Folmer *et al.* 1994). A standard PCR reaction protocol (Hajibabaei *et al.* 2005) was used for PCR amplifications and products were checked on a 2% E-gel® 96 Agarose (Invitrogen). Unpurified PCR amplicons were sequenced in both directions using M13 tails as primers. The sequencing reactions followed standard protocols of the Canadian Center for DNA Barcoding (Hajibabaei *et al.* 2005), with products subsequently purified using Agencourt® CleanSEQ protocol (Agencourt) and processed using BigDye version 3.1 on an ABI 3730 DNA Analyzer (Applied Biosystems).

Sequences were assembled with Sequencer 4.5 (Gene Code Corporation, Ann Arbor, MI, USA) and aligned by eye using BIOEDIT version 7.0.5.3 (Hall 1999); we observed no indels in this coding region of the mitochondrial genome and therefore all base positions were aligned with confidence in positional homology. Sequences are publicly available on BOLD (Ratnasingham & Hebert 2007; http://www.barcodinglife.org) within the project EWNOR (Earthworms of Normandie) and in GenBank (accession numbers in Appendix). Distance analyses were conducted with MEGA4 (Tamura *et al.* 2007) using a Neighbor-Joining (Saitou & Nei 1987) algorithm and distances corrected with the Kimura-2 parameter (Kimura 1980). The robustness of nodes was evaluated through bootstrap re-analysis of 1000 pseudoreplicates.

Results and discussion

A total of 141 specimens were processed, including 81 adults and 60 juveniles. DNA barcodes were obtained from all specimens, but the analysed data set was reduced to 131 specimens after exclusion of 10 sequences whose length was lower than 400 bp or having more than 10 ambiguous base-calls. Morphological examination of the adults revealed four species: *Lumbricus terrestris* (Linnée, 1758), *Lumbricus festivus* (Savigny, 1826), *Lumbricus rubellus* (Hoffmeister, 1843) and *Lumbricus castaneus* (Savigny, 1826). Table 2 lists the species morphologically identified at each of the thirteen sampling sites. The percentage of juveniles in samples ranged from 0 to 100% with an average value of 41.5%. Although our sampling procedure was not specifically designed to allow a quantitative estimation of adult to immature ratio, these values are assumed to fall within the range of what is usually found in natural earthworm populations (i.e. from 30 to 70% of juveniles, Decaëns, personal observation).

The distribution of intra- and interspecific distances calculated on adult specimens is clearly bimodal (Fig. 2), much like the pattern already documented in Taiwanese earthworms (Chang *et al.* 2009) and thus ensuring the efficiency of DNA barcodes as a tag for species discrimination (Rougerie *et al.* 2009). The Neighbor-Joining tree (Fig. 3) clearly illustrates the genetic distinctness of the different morphologically identified species, but it also highlights, unexpectedly, that *L. terrestris* includes two strongly divergent genetic clusters. This last result is currently being further investigated, but preliminary evidence supports the presence of two cryptic species (Decaëns *et al.* unpublished). As a result, we consider these two clusters as provisional species in this study. A high intra-specific variability of COI was also observed in *L. castaneus* and *L. rubellus*, and will require further investigation and sampling to assess the possible occurrence of cryptic species in these taxa as well.

All pairwise distances involving at least one unidentified juvenile specimen show the same distribution within the bimodal pattern as those involving only adult specimens (Fig. 2), and the analysis of genetic similarity combining adults and juveniles clearly associates both age classes within the same genetic clusters (Fig. 3). Thus, the barcoding approach permits reliable and straightforward species-level identifications for all juveniles that would otherwise be impossible to identify. This results in the identification of morphologically undetected species in two sampling sites out of thirteen (Table 2), and in three

Table 2 Number of specimens per species estimated with and without identification of juveniles. The numbers between brackets refer to the relative proportion (%) of each species in the total of specimens that were identified to species level. The numbers between brackets in the juvenile column refer to the proportion of juveniles (%) in the total number of specimens of each sampling site. Label sites refer to individual codes for each collecting site as given in Table 1; Species abbreviations as follow: Lter, *L. terrestris*; Lter1 & Lter2, first and second clusters of *L. terrestris* as identified in Fig. 3; Lfes, *L. festivus*; Lcas, *L. castaneus*; Lrub, *L. rubellus*; Juv, juvenile specimens; bold italic numbers highlight those species, which were only detected through the DNA-barcoding approach

Label Sites/ Species	Without DNA barcode identification					With DNA barcode identification				
	Lter	Lfes	Lcas	Lrub	Juv	Lter1	Lter2	Lfes	Lcas	Lrub
PGr1	5 (62.5)	2 (25)	1 (12.5)	0 (0.0)	9 (52.9)	***8 (50.0)***	***3 (18.7)***	4 (25.0)	1 (6.2)	0 (0.0)
PGr2	9 (69.2)	3 (23.1)	1 (7.7)	0 (0.0)	8 (38.1)	***4 (19.1)***	***12 (57.1)***	3 (14.3)	2 (9.5)	0 (0.0)
PGr3	2 (50.0)	1 (25)	0 (0.0)	1 (25)	4 (50.0)	0 (0.0)	2 (25.0)	5 (62.5)	0 (0.0)	1 (12.5)
PFo1	1 (100)	0 (0.0)	0 (0.0)	0 (0.0)	2 (66.7)	0 (0.0)	3 (100)	0 (0.0)	0 (0.0)	0 (0.0)
PFo2	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (100)	0 (0.0)
PFo3	0 (0.0)	0 (0.0)	3 (100)	0 (0.0)	1 (25.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (75.0)	***1 (25.0)***
PFo4	0 (0.0)	0 (0.0)	8 (66.7)	4 (33.3)	3 (20)	0 (0.0)	0 (0.0)	0 (0.0)	8 (53.3)	7 (46.7)
CGr1	8 (100)	0 (0.0)	0 (0.0)	0 (0.0)	9 (52.9)	***2 (11.8)***	***15 (88.2)***	0 (0.0)	0 (0.0)	0 (0.0)
CGr2	9 (100)	0 (0.0)	0 (0.0)	0 (0.0)	9 (50.0)	18 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
CFo1	7 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	7 (100)	0 (0.0)	0 (0.0)	0 (0.0)
FGr1	0 (0.0)	5 (50)	0 (0.0)	5 (50)	5 (33.3)	0 (0.0)	0 (0.0)	5 (33.3)	0 (0.0)	10 (66.7)
FGr2	0 (0.0)	0 (0.0)	0 (0.0)	2 (100)	2 (50.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (100.0)
VFo1	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (100)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	***2 (100)***

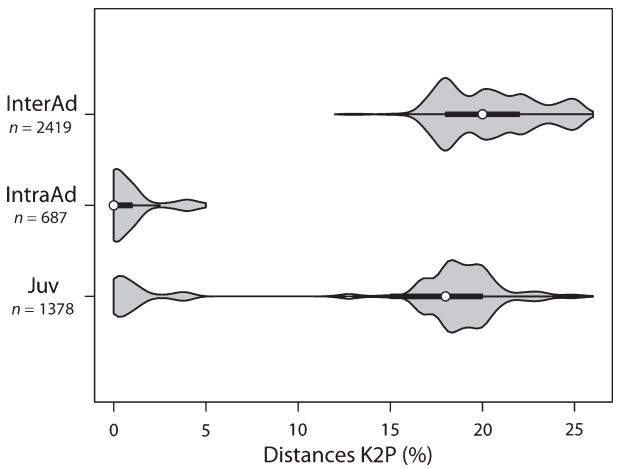

Fig. 2 Violin plots of intraspecific (IntraAd), interspecific (InterAd) pairwise distances for adult specimens and for juvenile specimens (Juv). The number of pairwise comparisons is indicated. The medians of data (white points), black bars indicating the interquartile range, and spikes extending to the upper- and lower-adjacent values are included.

other cases, the barcoding approach permits detection of both clusters of *L. terrestris*, which also increases the estimated species richness. Finally, molecular identifications also affected species relative abundance in almost all cases where more than one species was present. As a consequence, the global assessment of earthworm assemblages is significantly improved when extended to juvenile specimens through the use of DNA barcodes for species identification.

We thus emphasize that DNA barcoding, by enabling such identifications and by revealing cryptic species, represents a very important breakthrough for the integration of earthworm diversity data into soil studies. Indeed, earthworm biological and ecological research depends much on reliable species lists and has long been impeded by the impossibility to properly consider juveniles. So far, authors have either considered juveniles specimens as a distinct category (Margerie *et al.* 2001; Blackshaw *et al.* 2007; Pelosi *et al.* 2008) or have simply excluded them from analysis (Smetak *et al.* 2007). The collecting locality and/or the composition of local earthworm community is also occasionally used to infer species identity of juvenile specimens, but the method is error-prone, especially when considering the prevalence of sympatric cryptic species within common earthworm taxa (King *et al.* 2008). As exemplified by our results, these strategies unquestionably affect density/abundance measurements and can potentially lead to underestimates of species richness.

The value of DNA barcodes in extending species identification to juvenile specimens is not unique to the genus *Lumbricus* nor to earthworms; it can be profitable to the study of any group for which DNA barcodes are diagnostic and identification of some development stages is challenging. The reduction of sequencing cost, the multiplication of sequencing facilities, and more specifically the development of an international DNA barcoding initiative (iBOL; http://www.ibolproject.org) are rapidly facilitating access to this approach, making it a potential standard procedure in soil biodiversity assessment. Because earthworms are key organisms of soil

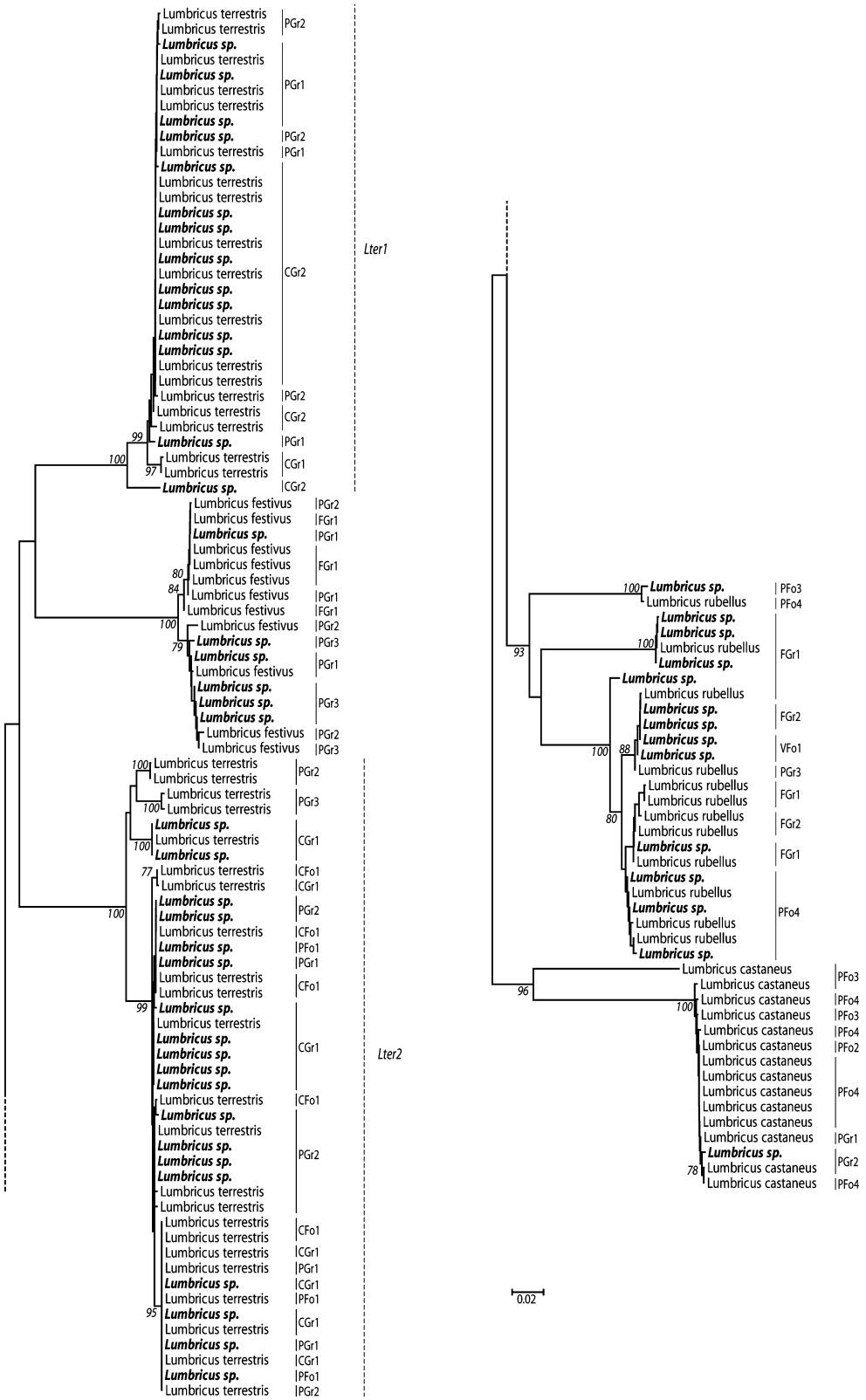

Fig. 3 Neighbor-Joining tree (K2P distances) of DNA barcodes for 131 *Lumbricus* specimens. Juvenile specimens are in bold italics. Bootstrap values greater than 75% are shown above branches. Lter1 and Lter2 refer to the two distinct genetic clusters of *L. terrestris*. Locality codes are figured on the tree and refer to codes used in Fig. 1 and in Table 1.

ecosystems and are important indicators of soil health and quality (Paoletti 1999; Lavelle *et al.* 2006; Suthar 2009), it is crucial to improve the resolution and the reliability of species identification (Nahmani *et al.* 2006) and to enable identifications for all life stages. DNA barcoding can meet both these needs and should rapidly be integrated into soil studies addressing earthworm diversity or based on diversity data.

Acknowledgements

We thank NSERC and Genome Canada through the Ontario Genomics Institute for supporting barcode analysis, and the Lycée Agricole d'Yvetot, the Conservatoire de Sites de Haute Normandie, the Office National des Forêts and the Réserve Naturelle des Mannevilles for allowing earthworm sampling in their soils.

References

Baker G, Lee KE (1993) Earthworms. In: *Field Samplings and Methods of Analysis* (ed. Carter MR), pp. 359–371. Lewis Publishers, Boca Raton.

Blackshaw RP, Donovan SE, Hazarika S, Bol R, Dixon ER (2007) Earthworm responses to long term agricultural management practices: spatial relationships with soil properties. *European Journal of Soil Biology*, **43**, S171–S175.

Bogh PS (1992) Identification of earthworms (Lumbricidae): choice of methods and distinction criteria. *Megadrilogica*, **4**, 163–174.

Borisenko AV, Lim BK, Ivanova NV, Hanner RH, Hebert PDN (2008) DNA barcoding in surveys of small mammal communities: a field study in Suriname. *Molecular Ecology Resources*, **8**, 471–479.

Bouché M (1972) *Lombriciens de France*. Ecologie et Systématique. Institut National de Recherches Agronomiques, Paris.

Chang C-H, Rougerie R, Chen J-H (2009) Identifying earthworms through DNA barcodes: Pitfalls and promise. *Pedobiologia*, **52**, 171–180.

Decaëns T, Rougerie R (2008) Descriptions of two new species of Hemileucinae (Lepidoptera: Saturniidae) from the region of Muzo in Colombia- Evidence from morphology and DNA barcodes. *Zootaxa*, **1944**, 34–52.

Decaëns T, Jiménez JJ, Gioia C, Measey J, Lavelle P (2006) The values of soil animals for conservation biology. *European Journal of Soil Biology*, **42**, S23–S38.

Decaëns T, Lavelle P, Jimenez JJ (2008) Priorities for conservation of soil animals. *CAB reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, **3**, 14.

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

Gibbs J (2009) Integrative taxonomy identifies new (and old) species in the *Lasioglossum (Dialictus) tegulare* (Robertson) species group (Hymenoptera,Halictidae). *Zootaxa*, **2032**, 1–38.

Hajibabaei M, DeWaard JR, Ivanova NV *et al.* (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1959–1967.

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, Alina C, Shelley LB, Jeremy RD (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Plos Biology*, **2**, 1657–1663.

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

INRA (1998) *A Sound Reference Base for Soils - ''Référentiel Pédologique''*. INRA, Paris.

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution*, **16**, 111–120.

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.

Lavelle P, Decaëns T, Aubert A *et al.* (2006) Soil invertebrates and ecosystem services. *European Journal of Soil Biology*, **42**, S3–S15.

Margerie P, Decaens T, Bureau F, Alard D (2001) Spatial distribution of earthworm species assemblages in a chalky slope of the Seine Valley (Normandy, France). *European Journal of Soil Biology*, **37**, 291–296.

Martinez JJ, Zaldivar-Riveron A, Saez AG (2008) Reclassification of *Bracon mendocinus*, a gall-associated doryctine wasp, and description of a new closely related species of *Allorhogas* (Hymenoptera: Braconidae). *Journal of Natural History*, **42**, 2689–2701.

Nahmani J, Lavelle P, Rossi JP (2006) Does changing the taxonomical resolution alter the value of soil macroinvertebrates as bioindicators of metal pollution? *Soil Biology and Biochemistry*, **38**, 385–396.

Paoletti MG (1999) The role of earthworms for assessment of sustainability and as bioindicators. *Agriculture Ecosystems & Environment*, **74**, 137–155.

Pelosi C, Bertrand M, Capowiez Y, Boizard H, Roger-Estrade J (2008) Earthworm collection from agricultural fields: comparisons of selected expellants in presence/absence of handsorting. *European Journal of Soil Biology*, **45**, 176–183.

Perez-Losada M, Eiroa J, Mato S, Dominguez J (2005) Phylogenetic species delimitation of the earthworms *Eisenia fetida* (Savigny, 1826) and *Eisenia andrei* Bouche, 1972 (Oligochaeta, Lumbricidae) based on mitochondrial and nuclear DNA sequences. *Pedobiologia*, **49**, 317–324.

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

Rougerie R, Decaëns T, Deharveng L *et al*. (2009) DNA barcodes for soil animal taxonomy: transcending the final frontier. *Pesquisa Agropecuária Brasileira*, **44**, 789–801.

Saitou N, Nei M (1987) The Neighbor-Joining Method - a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sims RW, Gerard BM (1999) *Earthworms*. FSC Publications, London.

Smetak KM, Johnsn-Maynard JL, Lloyd JE (2007) Earthworm population density and diversity in different-aged urban systems. *Applied Soil Ecology*, **37**, 161–168.

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4967–4972.

Suthar S (2009) Earthworm communities a bioindicator of arable land management practices: a case study in semiarid region of India. *Ecological Indicators*, **9**, 588–594.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Vaglia T, Haxaire J, Kitching IJ, Meusnier I, Rougerie R (2008) Morphology and DNA barcoding reveal three cryptic species within the *Xylophanes neoptolemus* and *loelia* species-groups (Lepidoptera: Sphingidae). *Zootaxa*, **1923**, 18–36.

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

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **360**, 1847–1857.

Appendix

Accession numbers of the COI sequences in BOLD and GenBank

BOLD Sample ID	BOLD Process ID	Species names	GenBank accession numbers	COI-5P length (Ambiguous base-calls)	Site labels
EW-ECO-0012	EWNOR012-07	*Lumbricus terrestris*	FJ937319	658 (0n)	PGr1
EW-ECO-0013	EWNOR013-07	*Lumbricus terrestris*	FJ937315	658 (0n)	PGr1
EW-ECO-0014	EWNOR014-07	*Lumbricus terrestris*	FJ937316	658 (0n)	PGr1
EW-ECO-0015	EWNOR015-07	*Lumbricus terrestris*	FJ937317	658 (0n)	PGr1
EW-ECO-0016	EWNOR016-07	*Lumbricus terrestris*	FJ937318	658 (0n)	PGr1
EW-ECO-0017	EWNOR017-07	*Lumbricus festivus*	FJ937287	618 (0n)	PGr1
EW-ECO-0018	EWNOR018-07	*Lumbricus festivus*	FJ937289	658 (0n)	PGr1
EW-ECO-0019	EWNOR019-07	*Lumbricus* sp.	FJ937295	658 (0n)	PGr1
EW-ECO-0020	*EWNOR020-07*	*Lumbricus* sp.	*GU206212*	*248 (5n)*	*PGr1*
EW-ECO-0021	EWNOR021-07	*Lumbricus* sp.	FJ937296	658 (0n)	PGr1
EW-ECO-0022	EWNOR022-07	*Lumbricus* sp.	FJ937297	658 (0n)	PGr1
EW-ECO-0023	EWNOR023-07	*Lumbricus* sp.	FJ937298	658 (0n)	PGr1
EW-ECO-0024	EWNOR024-07	*Lumbricus* sp.	FJ937299	658 (0n)	PGr1
EW-ECO-0025	EWNOR025-07	*Lumbricus* sp.	FJ937300	658 (0n)	PGr1
EW-ECO-0026	EWNOR026-07	*Lumbricus* sp.	FJ937301	465 (0n)	PGr1
EW-ECO-0027	EWNOR027-07	*Lumbricus castaneus*	FJ937285	658 (0n)	PGr1
EW-ECO-0029	EWNOR029-07	*Lumbricus* sp.	FJ937308	658 (0n)	PGr1
EW-ECO-0059	EWNOR059-07	*Lumbricus terrestris*	GU014224	669 (0n)	PFo1
EW-ECO-0060	EWNOR060-07	*Lumbricus* sp.	GU014230	669 (0n)	PFo1
EW-ECO-0065	EWNOR065-07	*Lumbricus* sp.	GU014231	669 (0n)	PFo1
EW-ECO-0077	EWNOR077-07	*Lumbricus terrestris*	GU014223	669 (0n)	CFo1
EW-ECO-0089	EWNOR089-07	*Lumbricus terrestris*	GU014225	669 (0n)	CFo1
EW-ECO-0090	EWNOR090-07	*Lumbricus terrestris*	GU014226	669 (0n)	CFo1
EW-ECO-0091	EWNOR091-07	*Lumbricus terrestris*	GU014227	669 (0n)	CFo1
EW-ECO-0092	EWNOR092-07	*Lumbricus terrestris*	GU014228	669 (0n)	CFo1
EW-ECO-0093	EWNOR093-07	*Lumbricus terrestris*	GU014229	669 (0n)	CFo1
EW-ECO-0094	EWNOR094-07	*Lumbricus terrestris*	GU206226	589 (0n)	CFo1
EW-ECO-0096	EWNOR095-08	*Lumbricus* sp.	FJ937304	658 (0n)	PGr2
EW-ECO-0098	EWNOR097-08	*Lumbricus* sp.	GU206211	633 (0n)	PGr2
EW-ECO-0099	EWNOR098-08	*Lumbricus terrestris*	FJ937312	644 (0n)	PGr2
EW-ECO-0100	EWNOR099-08	*Lumbricus terrestris*	FJ937313	632 (0n)	PGr2
EW-ECO-0101	EWNOR100-08	*Lumbricus terrestris*	FJ937314	658 (0n)	PGr2
EW-ECO-0108	EWNOR107-08	*Lumbricus* sp.	FJ937305	644 (0n)	PGr2
EW-ECO-0109	EWNOR108-08	*Lumbricus* sp.	FJ937306	634 (0n)	PGr2
EW-ECO-0110	EWNOR109-08	*Lumbricus* sp.	FJ937307	644 (0n)	PGr2

Appendix Continued

BOLD Sample ID	BOLD Process ID	Species names	GenBank accession numbers	COI-5P length (Ambiguous base-calls)	Site labels
EW-ECO-0111	EWNOR110-08	*Lumbricus terrestris*	FJ937321	634 (0n)	PGr2
EW-ECO-0112	EWNOR111-08	*Lumbricus terrestris*	FJ937322	644 (0n)	PGr2
EW-ECO-0113	EWNOR112-08	*Lumbricus terrestris*	FJ937323	644 (0n)	PGr2
EW-ECO-0120	EWNOR119-08	*Lumbricus terrestris*	FJ937324	645 (0n)	PGr2
EW-ECO-0121	EWNOR120-08	*Lumbricus terrestris*	FJ937327	644 (0n)	PGr2
EW-ECO-0122	EWNOR121-08	*Lumbricus terrestris*	FJ937320	657 (0n)	PGr2
EW-ECO-0129	EWNOR128-08	*Lumbricus* sp.	FJ937309	634 (0n)	PGr2
EW-ECO-0130	EWNOR129-08	*Lumbricus* sp.	FJ937310	519 (0n)	PGr2
EW-ECO-0131	EWNOR130-08	*Lumbricus* sp.	FJ937311	572 (1n)	PGr2
EW-ECO-0132	EWNOR131-08	*Lumbricus festivus*	FJ937290	655 (0n)	PGr2
EW-ECO-0133	EWNOR132-08	*Lumbricus festivus*	FJ937291	644 (0n)	PGr2
EW-ECO-0134	EWNOR133-08	*Lumbricus festivus*	FJ937286	570 (1n)	PGr2
EW-ECO-0135	EWNOR134-08	*Lumbricus castaneus*	FJ937284	655 (0n)	PGr2
EW-ECO-0156	EWNOR155-08	*Lumbricus terrestris*	GU206239	643 (1n)	CGr1
EW-ECO-0157	EWNOR156-08	*Lumbricus terrestris*	GU206225	658 (0n)	CGr1
EW-ECO-0158	EWNOR157-08	*Lumbricus terrestris*	GU206238	508 (0n)	CGr1
EW-ECO-0166	EWNOR165-08	*Lumbricus* sp.	GU206224	650 (1n)	CGr1
EW-ECO-0167	EWNOR166-08	*Lumbricus* sp.	GU206223	658 (0n)	CGr1
EW-ECO-0168	EWNOR167-08	*Lumbricus* sp.	GU206222	632 (0n)	CGr1
EW-ECO-0175	EWNOR174-08	*Lumbricus terrestris*	GU206237	580 (0n)	CGr1
EW-ECO-0176	EWNOR175-08	*Lumbricus terrestris*	GU206221	658 (0n)	CGr1
EW-ECO-0177	EWNOR176-08	*Lumbricus* sp.	GU206220	658 (0n)	CGr1
EW-ECO-0178	EWNOR177-08	*Lumbricus* sp.	GU206219	658 (0n)	CGr1
EW-ECO-0179	EWNOR178-08	*Lumbricus* sp.	GU206218	658 (0n)	CGr1
EW-ECO-0194	EWNOR193-08	*Lumbricus terrestris*	GU206217	658 (0n)	CGr1
EW-ECO-0195	EWNOR194-08	*Lumbricus terrestris*	GU206216	658 (0n)	CGr1
EW-ECO-0196	EWNOR195-08	*Lumbricus terrestris*	GU206215	658 (0n)	CGr1
EW-ECO-0197	EWNOR196-08	*Lumbricus* sp.	GU206214	658 (0n)	CGr1
EW-ECO-0198	EWNOR197-08	*Lumbricus* sp.	GU206213	658 (0n)	CGr1
EW-ECO-0199	EWNOR198-08	*Lumbricus* sp.	GU206210	658 (0n)	CGr1
EW-ECO-0215	EWNOR214-08	*Lumbricus castaneus*	GU206161	658 (0n)	PFo2
EW-ECO-0227	EWNOR226-08	*Lumbricus castaneus*	GU206160	658 (0n)	PFo3
EW-ECO-0228	EWNOR227-08	*Lumbricus castaneus*	GU206159	658 (0n)	PFo3
EW-ECO-0230	EWNOR229-08	*Lumbricus* sp.	GU206170	658 (0n)	PFo3
EW-ECO-0254	EWNOR253-08	*Lumbricus castaneus*	GU206163	658 (0n)	PFo3
EW-ECO-0264	EWNOR263-08	*Lumbricus festivus*	GU206168	658 (0n)	FGr1
EW-ECO-0265	EWNOR264-08	*Lumbricus festivus*	GU206167	658 (0n)	FGr1
EW-ECO-0266	EWNOR265-08	*Lumbricus festivus*	GU206166	658 (0n)	FGr1
EW-ECO-0267	EWNOR266-08	*Lumbricus rubellus*	GU206191	603 (0n)	FGr1
EW-ECO-0268	EWNOR267-08	*Lumbricus rubellus*	GU206190	636 (0n)	FGr1
EW-ECO-0269	*EWNOR268-08*	*Lumbricus rubellus*	*GU206189*	*593 (17n)*	*FGr1*
EW-ECO-0278	EWNOR277-08	*Lumbricus* sp.	GU206209	630 (0n)	FGr1
EW-ECO-0279	EWNOR278-08	*Lumbricus* sp.	GU206208	646 (0n)	FGr1
EW-ECO-0280	*EWNOR279-08*	*Lumbricus* sp.	*GU206183*	*292 (1n)*	*FGr1*
EW-ECO-0284	EWNOR283-08	*Lumbricus festivus*	GU206165	618 (1n)	FGr1
EW-ECO-0285	EWNOR284-08	*Lumbricus* sp.	GU206207	658 (0n)	FGr1
EW-ECO-0293	EWNOR292-08	*Lumbricus rubellus*	GU206187	458 (4n)	FGr1
EW-ECO-0294	EWNOR293-08	*Lumbricus rubellus*	GU206186	653 (0n)	FGr1
EW-ECO-0295	EWNOR294-08	*Lumbricus rubellus*	GU206185	554 (11n)	FGr1
EW-ECO-0296	EWNOR295-08	*Lumbricus festivus*	GU206164	648 (0n)	FGr1
EW-ECO-0297	EWNOR296-08	*Lumbricus* sp.	GU206184	511 (0n)	FGr1
EW-ECO-0298	EWNOR297-08	*Lumbricus* sp.	GU206206	518 (9n)	FGr1
EW-ECO-0306	EWNOR305-08	*Lumbricus terrestris*	GU206236	648 (0n)	CGr2
EW-ECO-0307	EWNOR306-08	*Lumbricus terrestris*	GU206235	648 (0n)	CGr2
EW-ECO-0308	EWNOR307-08	*Lumbricus terrestris*	GU206234	640 (0n)	CGr2

Appendix Continued

BOLD Sample ID	BOLD Process ID	Species names	GenBank accession numbers	COI-5P length (Ambiguous base-calls)	Site labels
EW-ECO-0309	EWNOR308-08	*Lumbricus* sp.	GU206233	648 (0n)	CGr2
EW-ECO-0310	EWNOR309-08	*Lumbricus* sp.	GU206205	648 (0n)	CGr2
EW-ECO-0311	EWNOR310-08	*Lumbricus* sp.	GU206204	648 (0n)	CGr2
EW-ECO-0320	EWNOR319-08	*Lumbricus terrestris*	GU206232	648 (0n)	CGr2
EW-ECO-0321	EWNOR320-08	*Lumbricus terrestris*	GU206231	648 (0n)	CGr2
EW-ECO-0322	EWNOR321-08	*Lumbricus terrestris*	GU206230	648 (0n)	CGr2
EW-ECO-0325	EWNOR324-08	*Lumbricus* sp.	GU206203	648 (0n)	CGr2
EW-ECO-0326	EWNOR325-08	*Lumbricus* sp.	GU206202	648 (0n)	CGr2
EW-ECO-0327	EWNOR326-08	*Lumbricus* sp.	GU206201	648 (0n)	CGr2
EW-ECO-0334	EWNOR333-08	*Lumbricus terrestris*	GU206229	648 (0n)	CGr2
EW-ECO-0335	EWNOR334-08	*Lumbricus terrestris*	GU206228	648 (0n)	CGr2
EW-ECO-0336	EWNOR335-08	*Lumbricus terrestris*	GU206227	648 (0n)	CGr2
EW-ECO-0337	EWNOR336-08	*Lumbricus* sp.	GU206200	648 (0n)	CGr2
EW-ECO-0338	EWNOR337-08	*Lumbricus* sp.	GU206199	648 (0n)	CGr2
EW-ECO-0339	EWNOR338-08	*Lumbricus* sp.	GU206198	648 (0n)	CGr2
EW-ECO-0346	*EWNOR345-08*	*Lumbricus rubellus*	*GU206183*	*292 (0n)*	*PFo4*
EW-ECO-0347	EWNOR346-08	*Lumbricus rubellus*	GU206182	511 (0n)	PFo4
EW-ECO-0348	EWNOR347-08	*Lumbricus rubellus*	GU206181	635 (0n)	PFo4
EW-ECO-0349	EWNOR348-08	*Lumbricus* sp.	GU206180	648 (0n)	PFo4
EW-ECO-0350	EWNOR349-08	*Lumbricus castaneus*	GU206158	658 (0n)	PFo4
EW-ECO-0351	EWNOR350-08	*Lumbricus castaneus*	GU206157	658 (0n)	PFo4
EW-ECO-0352	EWNOR351-08	*Lumbricus castaneus*	GU206156	658 (0n)	PFo4
EW-ECO-0370	EWNOR369-08	*Lumbricus castaneus*	GU206162	648 (0n)	PFo4
EW-ECO-0371	EWNOR370-08	*Lumbricus* sp.	GU206197	635 (2n)	PFo4
EW-ECO-0372	*EWNOR371-08*	*Lumbricus* sp.	*GU206196*	*565 (34n)*	*PFo4*
EW-ECO-0373	EWNOR372-08	*Lumbricus* sp.	GU206179	635 (0n)	PFo4
EW-ECO-0376	EWNOR375-08	*Lumbricus castaneus*	GU206155	658 (0n)	PFo4
EW-ECO-0377	EWNOR376-08	*Lumbricus castaneus*	GU206154	658 (0n)	PFo4
EW-ECO-0378	EWNOR377-08	*Lumbricus castaneus*	GU206153	658 (0n)	PFo4
EW-ECO-0379	EWNOR378-08	*Lumbricus castaneus*	GU206152	658 (0n)	PFo4
EW-ECO-0380	EWNOR379-08	*Lumbricus rubellus*	GU206178	644 (0n)	PFo4
EW-ECO-0381	EWNOR380-08	*Lumbricus rubellus*	GU206169	658 (0n)	PFo4
EW-ECO-0386	EWNOR385-08	*Lumbricus* sp.	FJ937293	658 (0n)	PGr3
EW-ECO-0387	EWNOR386-08	*Lumbricus* sp.	FJ937294	658 (0n)	PGr3
EW-ECO-0388	EWNOR387-08	*Lumbricus festivus*	FJ937288	584 (0n)	PGr3
EW-ECO-0398	EWNOR397-08	*Lumbricus* sp.	FJ937302	644 (0n)	PGr3
EW-ECO-0399	EWNOR398-08	*Lumbricus* sp.	FJ937303	658 (0n)	PGr3
EW-ECO-0408	EWNOR407-08	*Lumbricus terrestris*	FJ937325	658 (0n)	PGr3
EW-ECO-0409	EWNOR408-08	*Lumbricus terrestris*	FJ937326	658 (0n)	PGr3
EW-ECO-0410	EWNOR409-08	*Lumbricus rubellus*	FJ937292	545 (0n)	PGr3
EW-ECO-0416	*EWNOR415-08*	*Lumbricus rubellus*	*GU206177*	*470 (51n)*	*FGr2*
EW-ECO-0418	EWNOR417-08	*Lumbricus rubellus*	GU206175	613 (0n)	FGr2
EW-ECO-0421	EWNOR420-08	*Lumbricus rubellus*	GU206173	579 (0n)	FGr2
EW-ECO-0422	EWNOR421-08	*Lumbricus* sp.	GU206195	658 (0n)	FGr2
EW-ECO-0423	*EWNOR422-08*	*Lumbricus* sp.	*GU206194*	*541 (59n)*	*FGr2*
EW-ECO-0424	EWNOR423-08	*Lumbricus* sp.	GU206172	658 (0n)	FGr2
EW-ECO-0425	*EWNOR424-08*	*Lumbricus* sp.	*GU206193*	*426 (38n)*	*FGr2*
EW-ECO-0426	*EWNOR425-08*	*Lumbricus* sp.	*GU206192*	*380 (39n)*	*FGr2*
EW-ECO-0430	EWNOR429-08	*Lumbricus* sp.	GU206171	620 (0n)	VFo1
EW-ECO-0432	EWNOR431-08	*Lumbricus* sp.	GU206151	658 (0n)	VFo1
EW-ECO-0434	*EWNOR433-08*	*Lumbricus* sp.	*GU206150*	*269 (2n)*	*VFo1*

For each sequence, the length and the number of ambiguous base-calls are specified. Site labels refer to Table 1. Lines in italics signify that the sequence was not used in the data analyses (size < 400 bp and/or number of ambiguous base-calls >10)