Applied Soil Ecology 124 (2018) 362–371

Contents lists available at ScienceDirect

Applied Soil Ecology

journal homepage: www.elsevier.com/locate/apsoil

# A reference library of DNA barcodes for the earthworms from Upper Normandy: Biodiversity assessment, new records, potential cases of cryptic diversity and ongoing speciation

David Porco[a,f,*], Chih-Han Chang[b,c], Lise Dupont[d], Sam James[e], Benoit Richard[f], Thibaud Decaëns[g]

[a] Musée National d'Histoire Naturelle, 25 rue Munster, 2160 Luxembourg, Luxembourg
[b] Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21210, USA
[c] Department of Environmental Science and Technology, University of Maryland, College Park, MD 20742, USA
[d] Université Paris Est Créteil (UPEC), UPMC, Paris 7, CNRS, INRA, IRD, Institut d'écologie et des sciences de l'environnement de Paris, 94010 Créteil Cedex, France
[e] Department of Sustainable Living, Maharishi University of Management, Fairfield, IA 52557 USA
[f] Normandie Univ, UNIROUEN, IRSTEA, ECODIV, 76000 Rouen, France
[g] Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, CNRS–Univ Montpellier–Univ Paul-Valéry–EPHE–SupAgro Montpellier–INRA–IRD, Montpellier, France

ARTICLE INFO

*Keywords:*
DNA barcodes
Diversity assessment
Earthworms
Cryptic diversity
MOTU delineation

ABSTRACT

This study presents the assemblage of an exhaustive reference library for one of the major groups of soil invertebrates, earthworms, focused on the long sampled French location of Upper Normandy. Previous morphological appraisal of the diversity, enumerated 20 species in the area. After an extensive campaign of DNA barcoding aiming at several typical habitats of Upper Normandy, 22 species were found and 561 sequences were produced. A total of 36 discrete Molecular Operational Taxonomic Units (MOTUs) were detected among these species. Twenty-two of these MOTUs corresponded to several complexes of MOTUs within the morphological boundaries of 8 species. Based on a previous set of investigations on the *Allolobophora chlorotica*, *Aporrectodea rosea*, *Lumbricus rubellus* and *Aporrectodea icterica* complexes and on the analysis of the distribution of pairwise comparisons, we were able to hypothesize on the specific and subspecific status of the MOTUs in the 8 complexes detected. Globally, up to 21 of the 36 MOTUs detected potentially corresponded to species level entities. The remaining 15 MOTUs were considered as putative subspecies where gene flow can still be present between the members of a complex. Discussed in the perspective of previously detected physiological and ecological discrepancies between some of these specific but also subspecific MOTUs, these results emphasize the need to take into account all those genetic entities and to annotate them consistently throughout the literature. As a consequence, the genotyping of specimens in surveys and experiments is highly recommended when complexes of MOTUs are detected. This study also illustrates the usefulness of the DNA barcoding approach as a fast and powerful exploration tool for communities and therefore as a premise for more specific and integrative approaches.

## 1. Introduction

Earthworms are considered as one of the most ecologically important groups of invertebrates in soil environments. They often represent a significant part of the soil biomass, are often dominant in this respect (Lavelle and Spain, 2001), and have been designated as ecosystem engineers for their impact on the physical, biochemical and biological properties of the soil environment and are thus directly involved in soil fertility and nutrient cycling (Jones et al., 1994; Lavelle et al., 2006). This importance in the functioning of soil allied with their current occurrence in most habitats makes earthworms a key model group for bioindication in impact surveys of land use changes, management practices in agroecosystems or forestry, habitat restoration, or ecotoxicology (Paoletti, 1999; Rutgers et al., 2009; Pulleman et al., 2012; Pérès et al., 2011).

One critical issue for these applications is that the global understanding of the impact or the response of earthworms to condition changes in relation with the diversity of their communities can directly be challenged by the limitation of the usable morphological characters (Stürzenbaum et al., 2009). This has been particularly highlighted in

* Corresponding author at: Musée National d'Histoire Naturelle, 25 rue Munster, 2160 Luxembourg, Luxembourg.
*E-mail address:* david.porco.gm@gmail.com (D. Porco).

https://doi.org/10.1016/j.apsoil.2017.11.001
Received 26 May 2017; Received in revised form 26 October 2017; Accepted 1 November 2017
Available online 13 December 2017
0929-1393/ 

Lumbricidae which have been studied through the routine use of molecular taxonomy tools: some of the most common Lumbricidae species were proved to be species and subspecific entities complexes (e.g. *Allolobophora chlorotica* in King et al., 2008; *Lumbricus terrestris* in James et al., 2010, *Aporrectodea caliginosa* in Pérez-Losada et al., 2005 and in Fernández et al., 2012). Most of these European species were described centuries ago, and, since then, their morphological diagnoses have been used in many studies as reliable hypotheses. The detection of distinct genetic entities (Molecular Operational Taxonomic Units − MOTUs) of specific-level within the morphological boundaries of nominal species has been suggested to be related with different ecological and biological properties (James et al., 2010; Andre et al., 2010a; Porco et al., 2013; Liebeke et al., 2014). As a consequence, many data in the literature might be accumulated incorrectly under species names and this situation could jeopardize any attempts to use these organisms as bioindicators or to produce reliable generalisations. This emphasizes the importance to flag and document those cryptic MOTUs as a prior to any physiological or ecological study in order to make sure that the variations measured are actual intraspecifc ones.

DNA barcoding is currently the most used molecular taxonomy tool (Hebert et al., 2003) and could help solving these issues. It consists in the sequencing of a fragment of the 5′ end of the mitochondrial gene COI (cytochrome c oxidase subunit 1) used as a flag for species boundaries. In addition to the detection and flagging of cryptic or overlooked diversity cases (e.g. King et al., 2008; James et al., 2010; Novo et al., 2009; Fernández et al., 2012; Porco et al., 2013), the use of this approach holds many other advantages concerning the study of Lumbricidae. It allowed recovering specific level data for juvenile specimens that may represent a significant fraction of the samples, thus introducing a bias in communities surveys (Richard et al., 2010). Moreover, it allows for the processing of numerous specimens without the intervention of a taxonomist. Some topics, such as bioindication, biological invasions or global change impact surveys, require broad geographic sampling scales which produce a considerable amount of specimens to identify. If this amount exceeds the time and effort that available taxonomists can put into such an attempt, this could hamper broad scale studies and lead authors to use bad or degraded taxonomy (Bortolus, 2008). Thus DNA barcoding can bring much in soil ecology studies. However, to make these advantages available to the community, thoroughly curated and annotated DNA barcodes reference libraries will have to be produced and made publically available.

In this study, we assembled and analyzed such a DNA barcode reference library for the Lumbricidae from a region of northern France: Upper Normandy. This region is among the most sampled area in France where many species have been thoroughly scrutinized on the morphological ground (Bouché, 1972), and many populations were extensively studied for community assemblages, species distribution, feeding habits, resilience to climate change and landscape genetics (Decaëns et al., 1997, 2003, 2008, 2011; Dutoit et al., 1997; Margerie et al., 2001; Aubert et al., 2003; Hedde et al., 2007; Richard et al., 2012; Dupont et al., 2015, 2017; Clause et al., 2016).

Furthermore, we annotated this DNA barcode reference library for the specific and the subspecific status of the cryptic or overlooked MOTUs found within the morphological boundaries of the sampled species. Based on the methodology used in the well-studied MOTUs complex *A. chlorotica*, these specific/subspecific status hypotheses will be produced through modes delineation within the distribution of pairwise comparisons of DNA barcode sequences (Dupont et al., 2016). Such an annotation is critical as specific diversity assessments can directly be impacted and can, in turn, introduce a significant bias in ecological surveys based on this metric.

This study aimed at (1) assembling a reference library of DNA barcodes for Upper Normandy where 20 Lumbricidae species are expected to be present (Bouché, 1972, Decaëns et al., 2008), (2) evaluating the extent of the cryptic/overlooked diversity in this well-studied fauna (3) assessing the specific status of this cryptic/overlooked diversity through mode delineation in the pairwise comparison distribution and (4) discussing the consequences of our findings for soil ecologists.

## 2. Materials and methods

### 2.1. Study sites and sampling periods

Upper Normandy is a region of northern France with a temperate oceanic climate (mean annual temperature from 8 °C to 12 °C and mean annual precipitation between 900 and 1000 mm). Eight sites were sampled in three main landscape units: 1) the flood plains of the Seine River (Marais Vernier, Grand Mare in the reservation area of the Mannevilles and the lower area of Hénouville) and the Andelle river (Radepont); 2) the Chalky slopes of the Seine valley (St Adrien, slopes of Hénouville, meadow at Mont Saint Aignan); 3) the Plateau, with mainly meadows at Yvetot, and beech forests at the Eawy forest and Mont Saint Aignan (Supplementary material Table 1).

The sampling mainly took place from 2007 to 2010 (spring 2007, spring 2008, autumn 2008, winter 2008, spring 2009, summer 2009, winter 2009) with complementary sampling in spring 2012. The precise collection dates and localities are available in the public dataset DS-EWNRL on BOLD (accessible through the DOI dx.doi.org/10.5883/DS-EWNRL). This sampling effort was specifically undertaken to get fresh material for barcoding.

### 2.2. Earthworms sampling

For this study, earthworms were sampled by a combination of formalin extraction (i.e. application of a 0.4% formalin solution on soil surface), hand sorting of soil blocks of 25 × 25 × 25 cm, and qualitative search of individuals in several types of microhabitats (dung, decayed wood). Collected specimens were washed in water when collected by formalin extraction, and then killed and preserved in absolute ethanol. When necessary, ethanol was changed after a few hours in order to allow a satisfying fixation of the tissues.

### 2.3. Morphological analysis

Species were determined after the identification key available (Bouché, 1972) with a nomenclatural update after Sims and Gerard (1985) and the Global Biodiversity Information Facility (GBIF − http://www.gbif.org/). In particular, this led us to use the valid name of *Aporrectodea terrestris* (Savigny, 1826) instead of the currently employed but invalid name *Aporrectodea giardi* (Ribaucourt, 1901). All the specimens morphologically identified (561) were subsequently sequenced for the barcode region. A number of the species have been defined in Bouché (1972) as 'polytypic' species (i.e. having "varieties" and subspecies). Morphological examinations were undertaken for all available specimens to test if the morphotypes detected by Bouché (1972) were actually corresponding to any genetic variations that could be found in this survey within the morphological boundaries of nominal species. All the vouchers of this study were deposited in the collection of the laboratory ECODIV in Rouen (France).

### 2.4. Molecular analysis

DNA was extracted from a 1 $mm^2$ piece of muscle from the posterior part of the animal in 40 µl of lysis buffer and proteinase K incubated at 56 °C overnight. DNA extraction followed a standard automated protocol using 96-well glass fiber plates (Ivanova et al., 2006 − CCDB standard protocol www.ibolproject.org/docs/CCDB_DNA_Extraction.pdf). The 5′ region of COI used as a standard DNA barcode was amplified using M13 tailed primers LCO1490 and HCO2198 (Folmer et al., 1994). Samples that failed to generate an amplicon were subsequently amplified with a pair of internal primers combined with full length ones

**Table 1**
Intraspecific/MOTU and interspecific/MOTU genetic divergences and mode affiliations of Molecular Operational Taxonomic Units (MOTUs) of earthworm from Upper Normandy, France. For each MOTU of earthworm information is given on: the mean and maximum Intraspecific/MOTU divergence, the maximum genetic divergence measured among MOTUs within a complex, the nearest neighbor i.e. the species/MOTU exhibiting the minimal genetic distance to the species/MOTU named in the first column (this minimal distance is displayed in the column 'Distance to NN') and the mode assignation. In the 'Mode' column, the mode numbers in brackets corresponds to undersampled species of earthworm putatively assigned a posteriori to the different modes.

| | Mean Intraspecific/MOTU divergence | Maximum intraspecific/MOTU distances | Maximum intracomplex divergence | Nearest Neighbor | Distance to NN | Mode |
|---|---|---|---|---|---|---|
| *Allolobophora chlorotica L1* | 2.81 | 5.70 | 20.52 | *Allolobophora chlorotica L2* | 11.73 | M3 |
| *Allolobophora chlorotica L2* | 2.95 | 7.70 | | *Allolobophora chlorotica L1* | 11.73 | M3 |
| *Allolobophora chlorotica L3* | 0.55 | 1.11 | | *Allolobophora chlorotica L1* | 11.82 | M3 |
| *Allolobophora chlorotica L4* | 0.71 | 1.87 | | *Allolobophora chlorotica L3* | 14.64 | M4 |
| *Aporrectodea cupulifera* | 0.13 | 0.31 | | *Octolasion cyaneum* | 18.41 | M5 |
| *Aporrectodea caliginosa L1* | 0.23 | 0.36 | 17.98 | *Aporrectodea caliginosa L2* | 13.88 | M4 |
| *Aporrectodea caliginosa L2* | 1.11 | 2.82 | | *Aporrectodea caliginosa L3* | 10.7 | M3 |
| *Aporrectodea caliginosa L3* | 0.62 | 1.11 | | *Aporrectodea caliginosa L2* | 10.7 | M3 |
| *Aporrectodea terrestris* | 0.07 | 0.62 | 14.50 | *Aporrectodea longa* | 9.22 | M2 |
| *Aporrectodea icterica L1* | 1.29 | 3.14 | | *Aporrectodea icterica L2* | 12.67 | (M3) |
| *Aporrectodea icterica L2* | NA | N/A | | *Aporrectodea icterica L1* | 12.67 | (M3) |
| *Aporrectodea longa* | 1.16 | 2.83 | | *Aporrectodea terrestris* | 9.22 | M2 |
| *Aporrectodea rosea L1* | 0.00 | 0.00 | 15.88 | *Aporrectodea rosea L4* | 13.97 | M3 |
| *Aporrectodea rosea L2* | 0.00 | 0.00 | | *Aporrectodea rosea L1* | 14.64 | M3 |
| *Aporrectodea rosea L4* | 0.31 | 0.74 | | *Aporrectodea rosea L1* | 13.97 | M3 |
| *Dendrobaena attemsi* | 0.72 | 1.44 | 22.26 | *Allolobophora chlorotica L1* | 19.84 | M5 |
| *Dendrobaena octaedra L1* | 3.51 | 7.04 | | *Lumbricus rubellus L1* | 18.44 | (M5) |
| *Dendrobaena octaedra L2* | 0.30 | 0.31 | | *Allolobophora chlorotica L1* | 20.19 | (M5) |
| *Dendrodrilus rubidus* | 1.82 | 3.96 | | *Allolobophoridella eiseni* | 17.77 | M5 |
| *Allolobophoridella eiseni* | 1.08 | 1.08 | | *Dendrodrilus rubidus* | 17.77 | M5 |
| *Eisenia fetida* | 0.06 | 0.25 | | *Aporrectodea caliginosa L1* | 15.61 | M5 |
| *Eiseniella tetraedra* | 3.00 | 1.39 | | *Aporrectodea caliginosa L2* | 19.00 | M5 |
| *Lumbricus castaneus L1* | 0.17 | 0.48 | 20.66 | *Lumbricus castaneus L3* | 16.62 | M4 |
| *Lumbricus castaneus L2* | 0.15 | 0.47 | | *Lumbricus castaneus L3* | 17.02 | M4 |
| *Lumbricus castaneus L3* | 0.19 | 0.49 | | *Lumbricus castaneus L1* | 16.62 | M4 |
| *Lumbricus festivus* | 0.77 | 1.99 | | *Lumbricus herculeus* | 17.19 | M5 |
| *Lumbricus herculeus* | 0.46 | 4.17 | | *Lumbricus terrestris* | 15.99 | M4 |
| *Lumbricus rubellus L1* | 0.93 | 3.63 | 16.24 | *Lumbricus rubellus L3* | 12.32 | M3 |
| *Lumbricus rubellus L2* | 0.22 | 0.77 | | *Lumbricus rubellus L1* | 13.57 | M3 |
| *Lumbricus rubellus L3* | 0.00 | 0.00 | | *Lumbricus rubellus L1* | 12.32 | M3 |
| *Lumbricus terrestris* | 1.41 | 4.63 | | *Lumbricus rubellus L1* | 14.9 | M4 |
| *Murchieonia minuscula* | 0.00 | 0.00 | | *Aporrectodea caliginosa L2* | 19.58 | M5 |
| *Octolasion cyaneum* | 1.45 | 2.65 | 23.50 | *Octolasion lacteum L1* | 15.13 | M5 |
| *Octolasion lacteum L1* | NA | N/A | | *Octolasion cyaneum* | 15.13 | (M5) |
| *Octolasion lacteum L2* | NA | N/A | | *Lumbricus terrestris* | 16.88 | (M5) |
| *Satchellius mammalis* | NA | N/A | | *Lumbricus rubellus L1* | 18.79 | M5 |

(LepF1-MLepR1 and MLepF1-LepR1) (Hajibabaei et al., 2006). A standard PCR reaction protocol was used for amplifications, and products were checked on a 2% E-gel 96 Agarose (Invitrogen, Thermo Fisher Scientific, Waltham, MA, USA). Unpurified PCR amplicons were sequenced in both directions using M13 tailed primers, with products subsequently purified using Agencourt CleanSEQ protocol (Beckman Coulter, Mississauga, Ont, Canada) and processed using BigDye version 3.1 (Thermo Fisher Scientific, Waltham, MA, USA) on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, CA, USA). Sequences were assembled with Sequencher 4.5 (GeneCode Corporation, Ann Arbor, MI, USA) and aligned by eye using BIOEDIT version 7.0.5.3 (Hall, 1999). Sequences are publicly available on BOLD (dataset DS-EWNRL accessible through the DOI dx.doi.org/10.5883/DS-EWNRL) and on Genbank (FJ937284-FJ937325, GU013794-GU013814, GU013819-GU013825, GU013830, GU013831, GU013838, GU013839, GU013888, GU013889, GU013915, GU013987, GU013991-GU014000, GU014029, GU014071-GU014080, GU014129-GU014131, GU014223-GU014232, GU206151-GU206173, GU206175, GU206178-GU206182, GU206184-GU206187, GU206190, GU206191, GU206195, GU206197-GU206211, GU206213-GU206239, HM417894, HM417897-HM417921, HM417933-HM417977, HM879972-HM880030, HQ024538, HQ024539, HQ024639, HQ024640, HQ682476-HQ682478, HQ682496, HQ682497, JQ908635-JQ908651, JQ908672-JQ908677, JQ908686-JQ908692, JQ908733, JQ908766, JQ908771-JQ908773, JQ908775, JQ908776, JQ908779, JQ908780, JQ908782-JQ908784, JQ908800-JQ908802, JQ908806-JQ908808, JQ908825, JQ908827-JQ908829, JQ908833-JQ908836, JQ908847, JQ908890-JQ908897, JQ908902, JQ908903, JQ908906-JQ908908, JQ908912, JQ908924-JQ908941, JQ908943, JQ908948, JQ908951, JQ908953, JQ908955-JQ908958, JQ908961, JQ908962, JQ908995-JQ909009, JQ909014-JQ909018, JQ909025, JQ909027, JQ909031, JQ909033-JQ909035, JQ909037, JQ909038, JQ909041, JQ909044, JQ909060-JQ909068, JQ909071-JQ909075, JQ909085, JQ909135, JQ909140, JQ909144-JQ909148, JQ909153-JQ909162, MF121684-MF121784).

### 2.5. Molecular distance analysis and MOTUs delineation

Distance analyses were performed with MEGA7 (Kumar et al., 2016), using a Neighbor-Joining (Saitou and Nei, 1987) algorithm with the Kimura-2 parameter model (K2P − Kimura, 1980) to estimate genetic distances. The robustness of nodes was evaluated through bootstrap re-analysis of 1000 pseudoreplicates. The tree was replotted using the online utility iTOL (Letunic and Bork, 2007). Molecular Operational Taxonomic Units (MOTUs) were defined with the software 'mothur' using Hcluster command with the option 'Furthest neighbor' (Schloss et al., 2009).

### 2.6. Rarefaction curves and diversity estimators

In order to measure and compare the completion of the diversity survey for the molecular and the morphological approaches, rarefaction curves for specific diversity were generated with EcoSim 7.71 (Gotelli and Entsminger, 2006) with a 95% confidence level and plotted with R

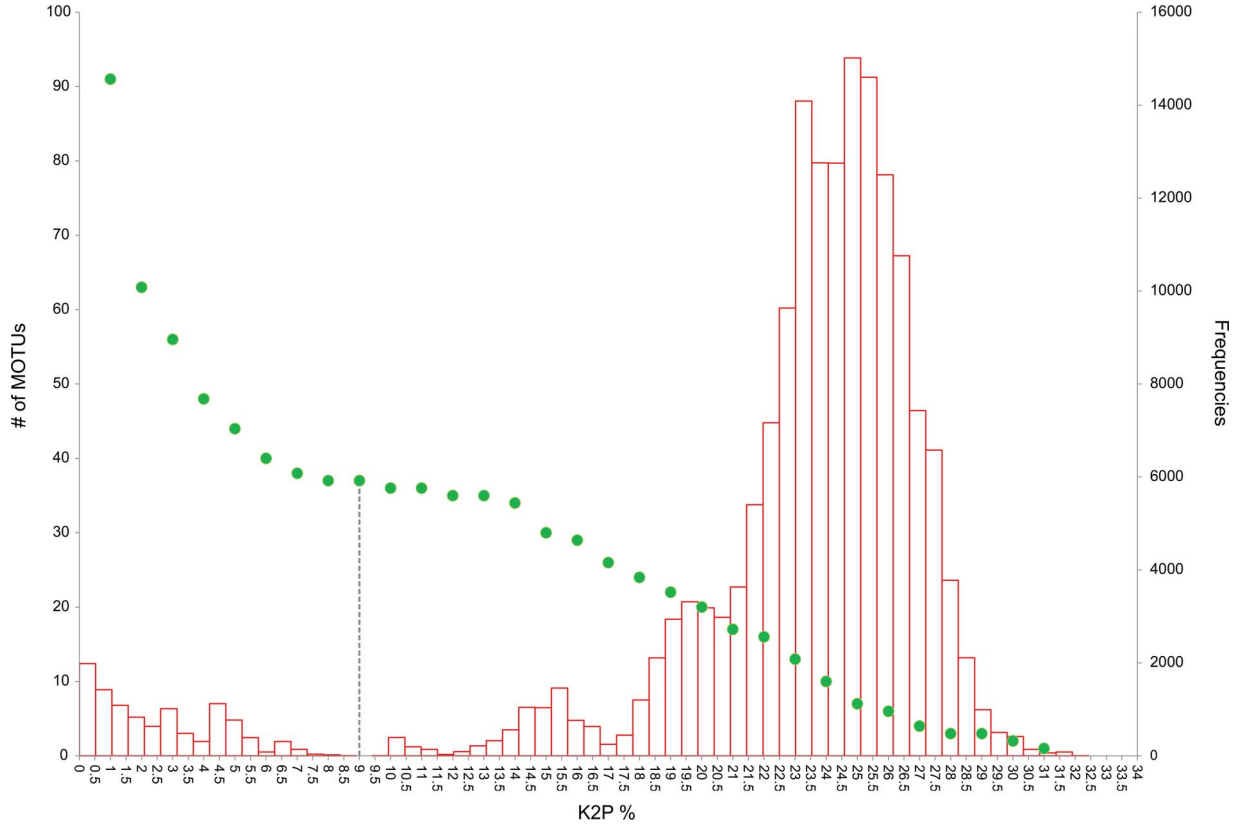


**Fig. 1.** Barcode gap graphic representation of the distribution of Molecular Operational Taxonomic Units (MOTUs) of earthworm species from Upper Normandy, France. Red histogram: distribution of frequencies of the pairwise comparisons of K2P distances (left ordinate). Green plot line: plot of the number of MOTUs against the different threshold values (right ordinate), low threshold values lead to sequences over-agglomeration and thus low MOTU number when high ones lead to oversplitting, producing high MOTU count. Neither of these extreme threshold positions could produce an actual representation of the specific diversity. Grey dash line represents the projection of the threshold chosen on the K2P% axis (9%) separating low intraspecific/MOTU distances from higher interspecific/MOTU ones. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.1.3 (R Development Core Team, 2012) using the package 'Plotrix' (Lemon, 2006). They were calculated separately for MOTUs and morphologically identified species as a function of the sampling effort (number of collected individuals). Additionally, theoretical species richness was calculated globally using the Chao1 and ACE diversity estimators calculated with the package 'Vegan' (Oksanen et al., 2012).

### 2.7. Distribution of pairwise comparisons and modes analysis

The multimodality of the pairwise distribution was analysed in order to classify MOTUs into different modes that could be related to a putative reproductive isolation status through the comparison with previously published datasets (Dupont et al., 2016). The distance matrix generated in the preceding step was used to analyze the distribution of all the pairwise comparisons issued both from the global dataset and a partial dataset. The partial dataset only took into account the intraspecific pairwise comparisons. This allowed filtering the pairwise comparisons external to species and complexes that could have degraded the distribution pattern; the species insufficiently sampled (*Octolasion lacteum*, *Dendrobaena octaedra*, *Eiseniella tetraedra* and *Aporrectodea icterica*) were excluded from this analysis. In order to delineate the different modes composing the distributions of the pairwise comparisons produced from each of the two datasets, a mixture of normal distributions was fitted through the Expectation-Maximization algorithm implemented in the R package 'mixtools' (Benaglia et al., 2009). The fitting, based on the distribution of the density of probabilities, was cross-validated by comparing the Bayesian Information Criterion (BIC) values for mixtures including a lesser number of modes than the one defined a priori as the optimal expectation.

## 3. Results

Twenty-two species were morphologically identified among the 561 specimens sequenced (Supplementary material Table 1). The plotting of the number of MOTUs against threshold values showed the characteristic plateau representing a stabilization of the number of MOTUs around an optimal value theoretically corresponding to biological species (Plaisance et al., 2011) (Fig. 1). Here, this plateau was found between the values 8% to 14% respectively yielding 37 to 33 MOTUs. The calculation of the pairwise K2P distance frequency for the dataset allowed determining the position of the barcode gap corresponding to a portion of the plateau. This area is situated from 8% to 9% (Fig. 1). In order to propose the most conservative hypothesis, we selected the higher boundary for divergence i.e. 9%.

Applied to the whole dataset, this 9% threshold value allowed the recovery of 36 MOTUs (Fig. 2). Fourteen of these MOTUs corresponded to the strict morphological delineation of nominal species. For these, maximal intraspecific distances ranged from 0% to 4.6% and minimal interspecific (Nearest Neighbor) from 9.2% to 19.7% (Table 1). The 22 other MOTUs were discrete lineages detected within the morphological boundaries of nominal species. Nine species were concerned: *Dendrobaena octaedra*, *Eiseniella tetraedra*, *Aporrectodea icterica*, *Allolobophora chlorotica*, *Aporrectodea rosea*, *Aporrectodea caliginosa*, *Lumbricus rubellus*, *Lumbricus castaneus* and *Octolasion lacteum*. Therefore, high maximum intraspecific divergence values were observed in these species (ranging from 9.2% up to 23.5% – Table 1). In these complexes, the intra-MOTU distance values ranged between 0% and 7.65% while interMOTU distances ranged from 7.7% up to 20.2%. The shortest distance found between MOTUs was within *E. tetraedra* with 7.7% divergence. This interMOTU divergence corresponded to the

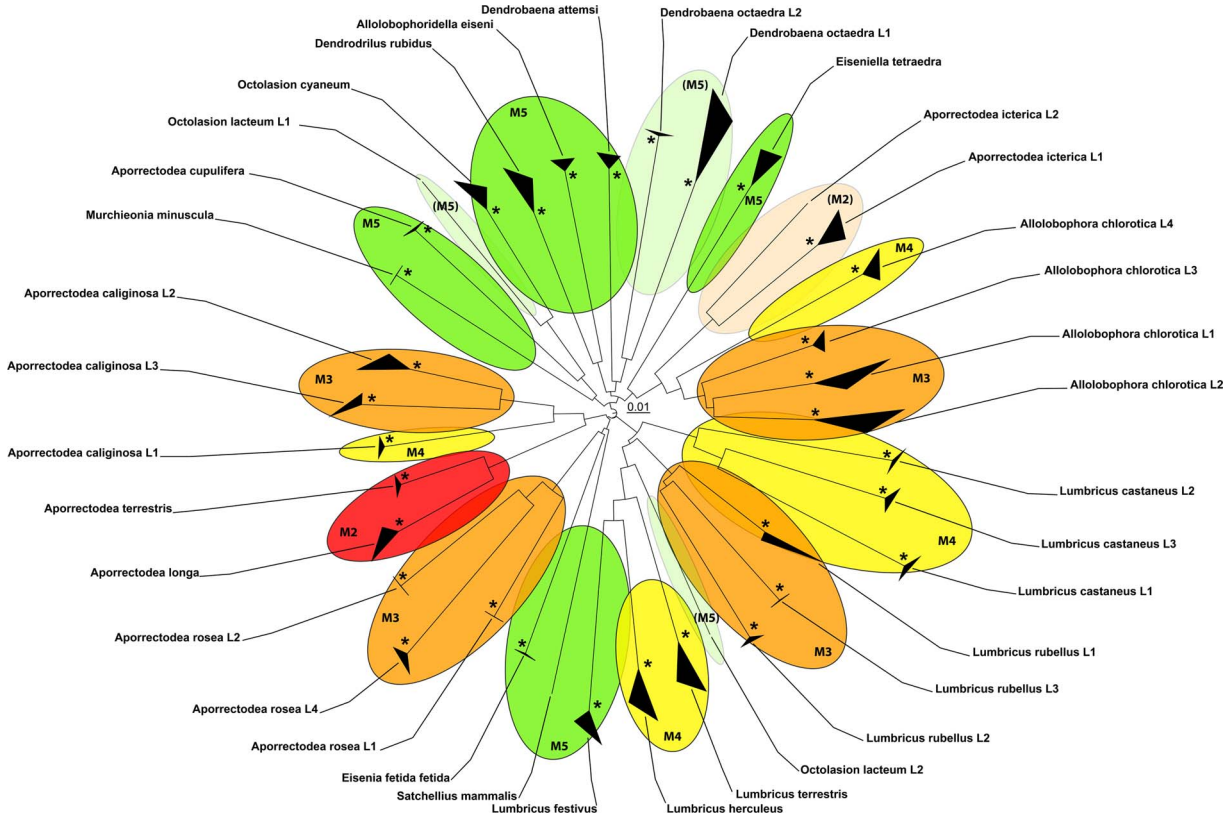


**Fig. 2.** Neighbor-joining tree for the 36 Molecular Operational Taxonomic Units (MOTUs) of earthworm species from Upper Normandy, France, delineated in the dataset. Each MOTU represented by more than a specimen was represented by a triangle: the Upper and lower sides of the triangle represent respectively the maximal and the minimal genetic distance within the MOTU (resulting in a flat line when no intraspecific variation was measured). Ellipses are colored and annotated by modes (red = M2, orange = M3, yellow = M4, green = M5). The mode numbers in brackets and pastel colors correspond to undersampled species putatively assigned a posteriori to the different modes. Nodes with bootstrap values ≥99% are marked with an asterisk.

intraspecific distances range, moreover, one of the MOTUs for this species was only represented by a single sequence, so the situation for *E. tetraedra* could possibly result from undersampling. This led us to reject the splitting of *E. tetraedra* into two MOTUs. Thus we retained a total of 36 MOTUS. The diversity assessment using the ACE and Chao1 richness estimators showed that the diversity was almost completely sampled with respectively 38 and 37 MOTUs predicted in the community (Fig. 3).

Concerning the partial dataset, the fitting of mixtures of normal distributions allowed to retrieve 5 modes (Fig. 4b) and to assign accordingly the different MOTUs delineated in the previous step of the analysis to each of them. Only 4 modes were recovered for the global distribution of pairwise comparisons as one of the modes remained underscaled (M2) (Fig. 4a). The cross-validation through BIC values confirmed that the mixture models comprising the amount of modes retrieved were more consistent than those comprising a lesser number of modes (which exhibited higher BIC values). Three species comprising six MOTUs (*Aporrectodea icterica*, *Dendrobaena octaedra*, *Octolasion lacteum*), were not included in these analyses because their sampling size was too low, but from the positions of the pairwise comparisons obtained within these 3 complexes of MOTUs, potential affiliation to the different modes was hypothesized (Table 1).

The morphological examination of the specimens sequenced for this study did not allow us to find any match between the MOTUs in the detected complexes and the subspecific categories described in Bouché (1972).

## 4. Discussion

### 4.1. Barcode as an identification tool in Lumbricidae

As demonstrated in previous studies, DNA barcoding enables an accurate identification at the species level in Lumbricidae (James et al., 2010; Porco et al., 2013; Richard et al., 2010). Moreover, this tool allowed reaching an even finer scale of identification with the detection of cryptic MOTUs exhibiting the same morphological diagnosis and thus bearing a single species name. Indeed, 8 species out of the 22 collected in Upper Normandy for this study were found to be complexes of several cryptic MOTUs. Some of these cases were already documented (*L. terrestris/L. herculeus* in James et al., 2010, *A. caliginosa* in Pérez-Losada et al., 2009, *A. rosea*, *L. rubellus* in Porco et al., 2013, confirmed for *L. rubellus* in Martinsson and Erseus, 2017, *A. icterica* in Torres-Leguizamon et al., 2014), and some were newly detected in this survey: *D. octaedra*, *L. castaneus* and *O. lacteum*. It is worth noting that one of the *O. lacteum* MOTU (L1) branched closer to *O. cyaneum* than the other *O. lacteum* MOTU (L2).

In the present study, the genetic distances between these cryptic MOTUs ranged from 10.7% to 20.2%. While higher distances (> 17%) that are comparable to those found between well-defined species suggest that these MOTUs are potential specific-level entities, the status for the MOTUs exhibiting the lower distances needs additional evaluation. The distribution of the pairwise comparisons was analyzed in order to get further insight in this respect.

### 4.2. MOTUs specific status inferred from pairwise distribution modes – insights from the A. chlorotica case

In Lumbricidae, the detailed investigation of the *A. chlorotica* complex case permitted a tentative insight into the status of the cryptic MOTUs delineated in this study as biological species. Three studies on this complex showed that some of the MOTUs found with COI were not recovered with nuclear markers (L2, L3) (King et al., 2008), that some

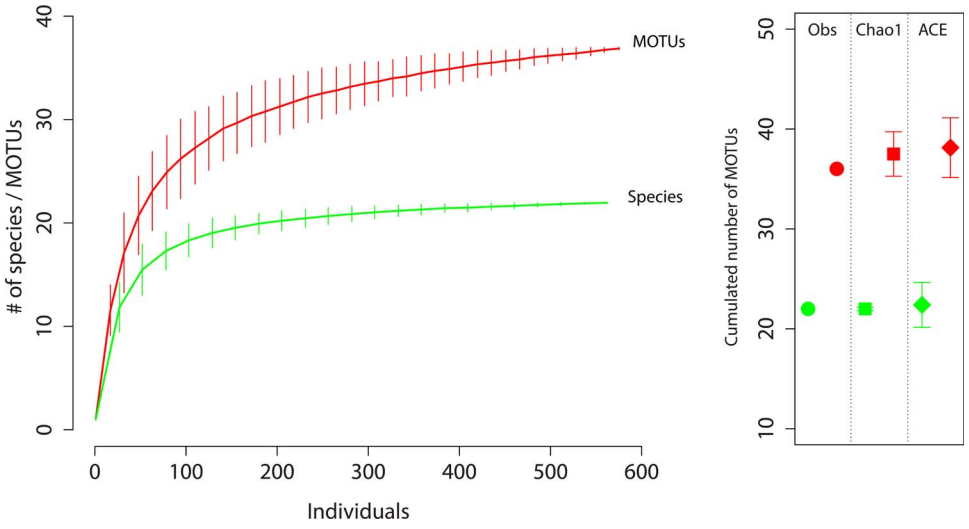


**Fig. 3.** Rarefaction curves of the number of species and Molecular Operational Taxonomic Units (MOTUs) of earthworm against sampling effort (expressed as a function of the number of specimens collected) both for MOTUs and species. Observed (Obs) and estimated (Chao1 and ACE) species richness are plotted.

were likely able to interbreed (L1, L2, and L3) (Dupont et al., 2011) and that cross-lineage hybridization was a low frequency phenomenon and might exhibit an asymmetric reproductive isolation (between MOTUs L1 and L2/L3) (Dupont et al., 2016). The subspecies status might be assigned to some of these MOTUs. Even if other markers or characters are needed to explore this kind of MOTU complexes, it has been shown that such situations could be detected from data accumulation and specifically frequencies of pairwise comparisons even from a single gene (Dupont et al., 2016). Indeed, with the inclusion of MOTU complexes into the analyses, the classic bimodal distribution (an intraspecific mode and an interspecific mode) was modified into a multimodal one (an intraMOTU mode and several interMOTU modes) where interMOTU modes of lower divergence were constituted by the distances between interbreeding entities (Dupont et al., 2016).

In the present study, the global distribution of the pairwise comparison analyzed along with the distributions from the main MOTU complexes detected, revealed the presence of five discrete modes (Fig. 4). These five modes corresponded to several classes of pairwise comparison among MOTUs from the different complexes (Table 1). Apart from M1 standing for intraspecific and intraMOTU comparisons, the M3 and M4 modes, detected previously (Dupont et al., 2016), were recovered here. M3 was flagged in this former study as containing pairwise comparisons for entities experiencing gene flow. In this mode M3, we retrieved all pairwise comparisons for the MOTUs of *L. rubellus*, *A. rosea* but also *A. caliginosa* (concerning only L1/L2 and L2/L3 for this latter). Such a pattern, combined with the elucidation of the gene flow network in the *A. chlorotica* complex, lead us to consider the hypothesis of a possible gene flow among the MOTUs from other complexes encompassed in this mode, thus considering either potential gene flow or incomplete lineage sorting events to be accountable for this portion of cryptic diversity. This hypothesis of variable levels of gene flow in M3 MOTUs is further supported by previous findings concerning (1) the *A. rosea* complex, in which the gene flow among lineages and populations was confirmed, even if limited, by a multiple markers study (Fernández et al., 2016), (2) two lineages of *L. rubellus*, corresponding to two of the species MOTUs detected in this study, which were found hybridizing (Andre et al., 2010a; Giska et al., 2015) and (3) the *A. icterica* complex, in which clear interbreeding was also uncovered between populations belonging to the two MOTUs detected in this study (Torres-Leguizamon et al., 2014). Thus, the species-level status can be ruled out for the MOTUs belonging to the mode M3.

In the mode M4 pairwise comparisons were found that involved respectively *L. terrestris/L. herculeus*, *A. caliginosa* L1/L3, *L. castaneus* L1/L2/L3, and *A. chlorotica* L4 against all the other MOTUs in the *A. chlorotica* complex. This mode overlaps with the last clearly discriminated one, M5, which contains all comparisons for well discriminated species (i.e. corresponding to a single MOTU) with the notable exception of the tandem *A. longa/A. terrestris* (within the range of the mode M2). In the light of the few genetic exchanges found between the *A. chlorotica* L4 and L1 MOTUs, this overlap was interpreted in Dupont et al. (2016) as another level of reproductive isolation corresponding to a near complete or complete speciation. The MOTUs belonging to this mode could therefore be considered as specific level entities.

One peculiar case is the mode M2, which is mainly represented by the pairwise comparisons between the sequences of *A. longa* and *A. terrestris*. This is an unexpected finding as they have been considered so far as well-defined species with clear morphological diagnostic characters (body size and sexual markings on the clitellum). This situation may result from a recent divergence and an asymmetric hybridization among MOTUs similar to the phenomena described in *A. chlorotica* (Dupont et al., 2016). This also stresses the inherent limitation of applying a unique criterion for specific delineation, even if conservative, drawn from a single mitochondrial marker. Further investigations, employing nuclear markers, will have to be undertaken in order to confirm the mode based hypotheses produced in this study. Several nuclear markers have been previously used to check on the specific status of COI MOTUs in Lumbricidae: microsatellites (Dupont et al., 2016), AFLP (King et al., 2008), H3 (Dupont et al., 2016) and 28S when sequenced extensively (i.e. sequence length superior to 2000pb – Pérez-Losada et al., 2005). Most of the previous studies cited used only one or only a few of the available nuclear markers for specific delimitation. A recent RADseq study, targeting the *L. rubellus* complex, determined that a massive multilocus nuclear approach would be necessary for an unbiased assessment of the specific status of MOTUs within a complex (Giska et al., 2015). It is worth noting that the classification of MOTUs in pairwise distribution modes produced congruent results with the conclusion of this RADseq study on *L. rubellus* (see paragraph 4.4).

Under the conservative hypothesis derived from our methodology (i.e. that all the MOTUs having their pairwise comparisons encompassed in M2 and M3 could, at least, exhibit a partial lack of reproductive barrier and thus cannot be accounted as species level MOTUs, including *A. longa/A. terrestris*), the total number of specific level MOTUs is lowered to 27. Even so, this molecular estimate of specific diversity is representing an important gain for the Lumbricidae fauna of Upper Normandy by comparison with the 22 species morphologically identified with an increase of 28.5%. Further studies will support or reject the hypothesis of gene flow events between *A. longa* and *A. terrestris*, but also in all the other MOTUs yielding pairwise comparisons within the boundaries of the mode M3.

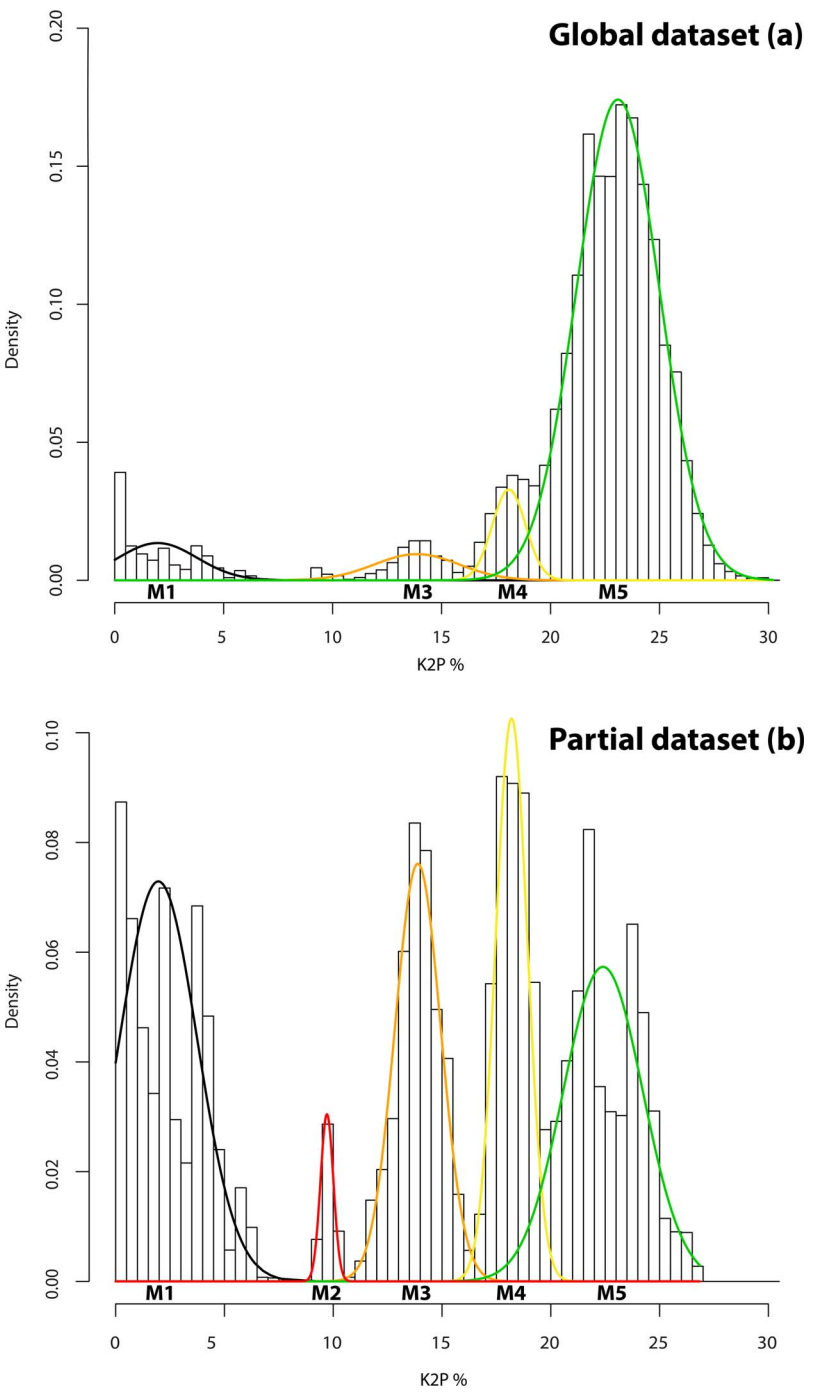


**Fig. 4.** Normal distributions fitted on the distribution of the pairwise comparisons of K2P distances Molecular Operational Taxonomic Units (MOTUs) of earthworm species from Upper Normandy, France, delineating the different modes (same color code as in Fig. 2). (a) Four normal distributions were fitted, delineating 4 modes in the global distribution of pairwise comparisons − scale issue prevents the fifth mode to be recognized. (b) Five normal distributions were fitted, delineating 5 modes in the distributions of pairwise comparisons from the different MOTU complexes detected and those calculated from single MOTU species. Line colors of the normal distributions feature the different modes which are also annotated under the x-axis. On y-axis the density of frequencies was used. This value is proportional to frequencies values rapported to the amplitude of K2P distance bins in which pairwise comparisons were classified (i.e. 0.5%). The shape of distribution is unchanged but this transformation allows the use of a probability density function in order to fit normal distributions.

Overall, these results confirmed the high level of cryptic diversity detected previously in Lumbricidae and even bring it further for some of the species examined (*L. castaneus*). Additional sampling may enable the detection of even more cryptic MOTUs within the boundaries of the species that could not be extensively sampled for this study (e.g. *Aporrectodea cupulifera*, *Allolobophoridella eiseni*, *Eisenia fetida*, *Eiseniella tetraedra*, *Murchieona minuscula*, *Satchellius mammalis*). However, in the light of the recent investigations regarding *A. chlorotica, A. rosea* and *A. icterica* (Torres-Leguizamon et al., 2014; Dupont et al., 2016; Fernández et al., 2016), this cryptic diversity should not be interpreted as a burst of specific diversity as it is likely to be related to ongoing speciation processes or even to ancient polymorphism. This result also showed that an exploratory approach with a single gene can shed light on the whole family, helping to pinpoint issues requiring further investigations on the different complexes detected, particularly concerning the current speciation level among MOTUs, and their potentially divergent biological and ecological characteristics.

Another important point is that whether these cryptic MOTUs qualify for the specific status or not, their high level of genetic divergence (i.e. way beyond the population level) advocates for their systematic annotation consistently throughout the literature. This will allow the accumulation of data on their respective intrinsic properties which would be otherwise agglomerated under a single species name. This might be a sound way to proceed, even if some of these MOTUs could be considered as belonging to a single species after further investigation (e.g. *A. icterica*), as the untangling of the data a posteriori might prove difficult or impossible. Also, documenting precisely this diversity may prove critical to avoid underestimation or misleading

interpretations of distribution patterns. This is especially important as continental-scale analyses are currently undertaken concerning the distribution of Lumbricidae diversity through databases merging (Rutgers et al., 2016). Furthermore, reference libraries are bound to play a major role as an increasing number of metabarcoding surveys on Lumbricidae are attempted and will need comprehensive references to produce optimal results (Bienert et al., 2012; Pansu et al., 2015). In these reference libraries, which will likely encompass different markers, COI could be a crucial element to connect the results of these metabarcoding surveys, using different markers, with the previous data accumulated so far in Lumbricidae DNA barcoding.

### 4.3. Morphological characterization of the MOTUs

Here, we tried to find correspondences between the deep genetic partitions found in this study and the subspecific categories defined in the most documented morphological study concerning earthworms from France but also specifically from Upper Normandy (Bouché, 1972).

None of the populations from species considered polytypic by Bouché (1972), and examined here, showed a consistent relationship between MOTU membership and the morphologically-defined subspecies or other infraspecific categories he defined. Bouché's infraspecific diagnostic characters were found polymorphic within the MOTUs of a single complex, preventing the establishment of any correspondence. Thus, this diversity can still be considered as cryptic and none of these characters can be rehabilitated as a species or subspecies diagnosis criterion. Furthermore, Bouché's monotypic species exhibiting a diversity of MOTUs (*O. lacteum*), and reexamined with morphology can also soundly be considered as cryptic diversity cases as no morphological differentiation could be found. Another situation relates to species described as morphologically polytypic in Upper Normandy by Bouché (1972) and in which only one MOTU was retrieved in the present study (*A. longa* and *A. terrestris*). In this case we can hypothesize that our geographic sampling only focusing on Normandy might not have included the MOTUs which could have potentially corresponded with Bouché's morphotypes.

Other species represented by a unique MOTU (*D. rubidus*, *M. minuscula*, and *O. cyaneum*) were found polytypic in France by Bouché (1972), but the status of these species for Upper Normandy was not documented. However, the morphological examination of the barcoded specimens showed that only one morphotype was sampled in this survey.

After the close morphological examination of the MOTUs found in this study, the cryptic nature of this diversity was confirmed and none of the previous infraspecific nomenclature from Bouché could be applied to annotate it.

### 4.4. Interest and advances for soil ecologists: the Lumbricus rubellus case

*Lumbricus rubellus* is a sentinel organism (Bundy et al., 2008) in which two distinct COI lineages were detected (King et al., 2008) and correspond to two of the *L. rubellus* MOTUs found in this study. Functional implications of these deep genetic divergences were shown concerning physiological mitigation to Pb exposure (Andre et al., 2010a), tolerance level to Pb and Zn (Andre et al., 2010b), metabolic profiles (Liebeke et al., 2014), strategies to respond to long-term toxic exposure (Kille et al., 2013), and differential preferences in soil pH and organic matter content (Spurgeon et al., 1994). Moreover the two *L. rubellus* lineages were found to have different evolutionary histories (Kille et al., 2013). Similarly, other studies have shown in other taxonomic groups that cryptic species could exhibit different levels of tolerance and response to various pollutants (Sturmbauer et al., 1999; Rocha-Olivares et al., 2004). In Lumbricidae, the *L. terrestris/L. herculeus* tandem showed divergence in ecological requirement (James et al., 2010). These elements suggest that deep mitochondrial divergences could be related to significant physiological or ecological discrepancies among MOTUs of a complex. Thus cryptic complexes should be thoroughly monitored to avoid any misleading accumulation of results and knowledge under a single deceiving species name.

Furthermore, the *L. rubellus* case is particularly interesting as hybrids were found between the two lineages (Andre et al., 2010a) and a broad genome screening through RADseq analysis confirmed that they were not reproductively isolated (Giska et al., 2015), but nevertheless exhibited pre-reproductive isolation mediated by water-soluble pheromones (Jones et al., 2016). This exemplifies that a cryptic diversity, which is not of specific level, can dramatically impact the biological and ecological properties of COI highly divergent entities. Moreover, these results in *L. rubellus* support the assessment of the specific status developed in the present study i.e. the modal classification of MOTUs through the analysis of pairwise distances distribution: the three cryptic MOTUs detected for *L. rubellus* were classified in the mode M3 in which the occurrence of variable level of gene flow among MOTUs was hypothesized.

The direct consequence of these findings is that, no matter the status of the COI MOTUs uncovered, records should be kept of which ones of these MOTUs are actually employed in various studies. Low genetic variability was previously pointed as a critical prerequisite for species of earthworms employed in biomonitoring (Kautenburger, 2006). The *L. rubellus* case showed clearly that the genetic variability within model species, past a certain level, such as detected in the present study, could allow predicting high risk of confounding factors thus invalidating the use of such species for biomonitoring. The only way to ensure the genetic background equivalence among all the specimens used in surveys or experiments is to go through their systematic genotyping (Andre et al., 2010b; Rocha-Olivares et al., 2004).

Conversely, it is worth noting that hints of such cryptic diversity (either of specific or subspecific level) with potential ecological consequences could also be gathered through the apparent versatility of nominal species regarding wide ecological ranges in soil properties (Sims and Gerard, 1985) or contamination levels (Spurgeon et al., 1994). Several specimens from the range limit of these various conditions could then be sequenced in order to detect any potential cryptic MOTUs. Such a reverse detection approach, from ecological requirements to cryptic diversity assessment, could be especially facilitated as massive databases are assembled on distributions, life history and functional traits of soil invertebrates (COLTRAIT – http://www.bdd-inee.cnrs.fr/spip.php?article51, BETSI – Pey et al., 2014).

### 4.5. Consistent annotation of Lumbricidae MOTUs

In this context, assembling reference libraries is essential in order to allow a consistent cross-comparison of the data obtained in the previous studies. Here we propose to connect the MOTUs found in this paper with the findings in some of the main previous surveys targeting Lumbricidae (King et al., 2008; Knott and Haimi, 2010; Klarica et al., 2012; Fernández et al., 2012; Pérez-Losada et al., 2012; Shekhovtsov et al., 2013). The combination and comparison of the sequences produced in these studies with ours allowed to propose a nomenclatural correspondence, enabling a higher consistency and thus a comparison (Supplementary material Table 2). The nomenclature used here was consistent with the one employed in previous studies (King et al., 2008; Porco et al., 2013; Torres-Leguizamon et al., 2014).

This assemblage of data allowed the detection of a category in published sequences showing inconsistency with the species identification of the MOTUs established in this study. This is the case concerning the sequence JN850542.1, which stands for the morphological identification *A. longa* (Pérez-Losada et al., 2012), but clusters with the MOTU identified as *A. terrestris* in our study. This case could be either due to misidentification or cross-contamination in the samples, or to the fact that the diagnostic criterion might be prone to polymorphism in this tandem of species which might still experience gene flow (see 'MOTUs specific status' paragraph).

The second case was a group of sequences (JN869874.1-JN869876.1; JN869878.1; JN869881.1; JN869882.1) linked to the species identification *A. tuberculata* and clustered with the MOTU *A. caliginosa* L2. In order to look into this case we reexamined morphologically 81 of the *A. caliginosa* specimens from all three lineages, including some specimens reported in this study and some Canadian specimens reported by Porco et al. (2013). Although we found specimens of *A. caliginosa* L2 that differed from the other *A. caliginosa* specimens by the absence of genital tumescences in xxxiii and would thus consistently fall into the concept of *A. tuberculata* (Gates, 1972a, 1972b; Reynolds, 1977), not all L2 had the "tuberculata-type" genital tumescences: out of the 18 *A. caliginosa* L2 specimens whose genital tumescences were unambiguously recognizable, two had genital tumescences in xxxiii, a characteristic of *A. caliginosa*. Gates (1972a), although considering *A. caliginosa* and *A. tuberculata* as separate species, noted that an absence of genital tumescences in xxxiii of *A. caliginosa* 'would not be surprising', but appearance of genital tumescences in xxxiii of *A. tuberculata* would be 'less likely'. However, our finding suggests that it is not unusual to find genital tumescences in xxxiii in a lineage that would be routinely identified as *A. tuberculata*. This suggests that the presence or absence of genital tumescences in xxxiii is plastic, and that specimens widely identified as *A. tuberculata* in North America and Europe may belong to *A. caliginosa* L2.

The other category concerns studies that focused on a single species which was shown here as well as in previous surveys to be a complex of MOTUs. In such situations, applying a consistent nomenclature would be useful to know if the study refers to one or several of these MOTUs and to which one the results should be assigned. This is the case for the study focusing on the mitochondrial diversity of *D. octaedra* (Knott and Haimi, 2010). The comparison with our annotated dataset allowed to assign the mitochondrial diversity investigated in this paper to the MOTU *D. octaedra* L1, thus unambiguously attributing it to inter-populational variability.

These few exemplary cases highlight clearly the interest of implementing a consistent nomenclature across literature in both past and future studies. The accumulation of data for each of these discrete genetic entities would help clarifying the so far blind accretion of data under a single species name potentially representing several different biological realities (Porco et al., 2012; Decaëns et al., 2013). Therefore, in order to unravel this intricate situation and to avoid reproducing it, we suggest the systematic sequencing of COI before experimentations take place.

## 5. Conclusion

The present barcoding survey of the Lumbricidae species in a limited area (Upper Normandy) illustrates what the routine use of molecular taxonomy tools can bring in diversity appraisal, even for a well-known group (concerning taxonomy and ecology) in a region which has been well sampled for decades. Indeed numerous cryptic MOTUs were found in common species, thus confirming and generalizing the previous findings for the family. On a lower estimate, taking into account this cryptic diversity actually raises the diversity of the area by more than 27%. This extra diversity can genuinely be called cryptic as we found no morphological criteria to match these MOTUs with the different morphotypes and varieties previously described. The results concerning the status of these MOTUs, analyzed with the recent findings in the literature, suggested that some of them are inter-fertile to some extent, and that different levels of speciation could actually be occurring among MOTUs.

Species names are basal hypotheses implying that 'end-users' deal with entities presenting homologous properties among individuals. Many disciplines are currently using these hypotheses routinely without any further questioning. The set of elements gathered in *L. rubellus* allied with the high level of cryptic diversity detected in this study and in previous ones (e.g. James et al., 2010; King et al., 2008; Pérez-Losada et al., 2005), either at specific or subspecific level, shows that many studies could benefit from accounting for this diversity that has direct physiological thus functional and ecological implications.

In practice, the outcomes of the present study lead us to recommend (1) the examination of the distribution of distance pairwise comparisons which could give valuable hints on a potential ongoing speciation in the groups studied, (2) the systematic genotyping of specimens used in surveys and experiments and (3) the consistent annotation of the discrete MOTUs uncovered.

## Acknowledgments

This study was supported by grants from the ADEME through the GESSOL-3 call (GENOSOIL project; grant # 1060C0092), from the Regional Council of the Haute-Normandie region through the GRR-SER call (ESTER project). The molecular analytical part of the survey was supported by grants from NSERC, Genome Canada and the Ontario Genomics Institute through the funding of the iBOL project. D. Porco received a one-year post-doctoral fellowship from the Upper Normandy regional council and another one-year post-doctoral fellowship from the GENOSOIL project. S. James and C.-H. Chang were supported by USA National Science Foundation NSF awards DEB- 1136604 and EF-1550795, respectively.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.apsoil.2017.11.001.

## References

Andre, J., King, R.A., Sturzenbaum, S.R., Kille, P., Hodson, M.E., Morgan, A.J., 2010a. Molecular genetic differentiation in earthworms inhabiting a heterogeneous Pb-polluted landscape. Environ. Pollut. 158, 883–890.
Andre, J., Sturzenbaum, S.R., Kille, P., Morgan, A.J., Hodson, M.E., 2010b. Metal bioaccumulation and cellular fractionation in an epigeic earthworm (Lumbricus rubellus): The interactive influences of population exposure histories, site-specific geochemistry and mitochondrial genotype. Soil Biol. Biochem. 42, 1566–1573.
Aubert, M., Hedde, M., Decaëns, T., Bureau, F., Margerie, P., Alard, D., 2003. Effects of tree canopy composition on earthworms and other macro-invertebrates in beech forests of Upper Normandy (France). Pedobiologia 47, 904–912.
Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S., 2009. Mixtools: an r package for analyzing finite mixture models. J. Stat. Softw. 32, 1–29.
Bienert, F., DE Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J.J., Taberlet, P., 2012. Tracking earthworm communities from soil DNA. Mol. Ecol. 21, 2017–2030.
Bortolus, A., 2008. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37, 114–118.
Bouché, M.B., 1972. Lombriciens de France. Ecologie et systematique. Institut natn Rech agron Publ, pp. 1–671.
Bundy, J.G., Sidhu, J.K., Rana, F., Spurgeon, D.J., Svendsen, C., Wren, J.F., Sturzenbaum, S.R., Morgan, A.J., Kille, P., 2008. 'Systems toxicology' approach identifies coordinated metabolic responses to copper in a terrestrial non-model invertebrate, the earthworm Lumbricus rubellus. BMC Biol. 6.
Clause, J., Barot, S., Forey, E., 2016. Earthworms promote greater richness and abundance in the emergence of plant species across a grassland-forest ecotone. J. Plant Ecol. 9, 703–711.
Decaëns, T., Dutoit, T., Alard, D., 1997. Earthworm community characteristics during afforestation of abandoned chalk grasslands (Upper Normandy, France). Eur. J. Soil Biol. 33, 1–11.
Decaëns, T., Bureau, F., Margerie, P., 2003. Earthworm communities in a wet agricultural landscape of the Seine Valley (Upper Normandy, France). Pedobiologia 47, 479–489.
Decaëns, T., Margerie, P., Aubert, M., Hedde, M., Bureau, F., 2008. Assembly rules within earthworm communities in North-Western France − A regional analysis. Appl. Soil Ecol. 39, 321–335.
Decaëns, T., Margerie, P., Renault, J., Bureau, F., Aubert, M., Hedde, M., 2011. Niche overlap and species assemblage dynamics in an ageing pasture gradient in north-western France. Acta Oecol.-Int. J. Ecol. 37, 212–219.
Decaëns, T., Porco, D., Rougerie, R., Brown, G.G., James, S.W., 2013. Potential of DNA barcoding for earthworm research in taxonomy and ecology. Appl. Soil Ecol. 65, 35–42.
Dupont, L., Lazrek, F., Porco, D., King, R.A., Rougerie, R., Symondson, W.O.C., Livet, A., Richard, B., Decaëns, T., Butt, K.R., Mathieu, J., 2011. New insight into the genetic structure of the Allolobophora chlorotica aggregate in Europe using microsatellite and mitochondrial data. Pedobiologia 54, 217–224.
Dupont, L., Gresille, Y., Richard, B., Decaëns, T., Mathieu, J., 2015. Dispersal constraints and fine-scale spatial genetic structure in two earthworm species. Biol. J. Linn. Soc. 114, 335–347.
Dupont, L., Porco, D., Symondson, W.O.C., Roy, V., 2016. Hybridization relics complicate barcode-based identification of species in earthworms. Mol. Ecol. Resources 16, 883–894.

Dupont, L., Torres-Leguizamon, M., Rene-Corail, P., Mathieu, J., 2017. Landscape features impact connectivity between soil populations: a comparative study of gene flow in earthworms. Mol. Ecol. 26, 3128–3140.
Dutoit, T., Decaëns, T., Alard, D., 1997. Successional changes and diversity of soil macrofaunal communities on chalk grasslands in Upper-Normandy (France). Acta Oecol.-Int. J. Ecol. 18, 135–149.
Fernández, R., Almodovar, A., Novo, M., Simancas, B., Cosin, D.J.D., 2012. Adding complexity to the complex: new insights into the phylogeny, diversification and origin of parthenogenesis in the Aporrectodea caliginosa species complex (Oligochaeta, Lumbricidae). Mol. Phylogenet. Evol. 64, 368–379.
Fernández, R., Novo, M., Marchan, D.F., Cosin, D.J.D., 2016. Diversification patterns in cosmopolitan earthworms: similar mode but different tempo. Mol. Phylogenet. Evol. 94, 701–708.
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. Mol. Mar. Biol. Biotechnol. 3, 294–299.
Gates, G.E., 1972a. Burmese Earthworms: an introduction to the systematics and biology of Megadrile oligochaetes with special reference to South-East Asia. Trans. Am. Philos. Soc. 67, 1–326.
Gates, G.E., 1972b. Contributions to North American earthworms. Toward a revision of the earthworm family Lumbricidae IV. The trapezoids species group. Bull. Tall Timbers Res. Station 12, 1–146.
Giska, I., Sechi, P., Babik, W., 2015. Deeply divergent sympatric mitochondrial lineages of the earthworm Lumbricus rubellus are not reproductively isolated. BMC Evol. Biol. 15.
Gotelli, N.J., Entsminger, G.L., 2006. EcoSim: Null Models Software for Ecology, 7 ed. Acquired Intelligence and Kesey-Bear, Jericho, Vermont, USA.
Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. U. S. A. 103, 968–971.
Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 41, 95–98.
Hebert, P.D.N., Cywinska, A., Ball, S.L., Dewaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. London Series B-Biol. Sci. 270, 313–321.
Hedde, M., Aubert, M., Bureau, F., Margerie, P., Decaens, T., 2007. Soil detritivore macroinvertebrate assemblages throughout a managed beech rotation. Ann. For. Sci. 64, 219–228.
Ivanova, N.V., Dewaard, J.R., Hebert, P.D.N., 2006. An inexpensive: automation-friendly protocol for recovering high-quality DNA. Mol. Ecol. Notes 6, 998–1002.
James, S.W., Porco, D., Decaëns, T., Richard, B., Rougerie, R., Erséus, C., 2010. DNA barcoding reveals cryptic diversity in Lumbricus terrestris L., 1758 (Clitellata): resurrection of L. herculeus (Savigny, 1826). PLoS One 5, e15629.
Jones, C.G., Lawton, J.H., Shachak, M., 1994. Organisms as ecosystem engineers. Oikos 69, 373–386.
Jones, G.L., Wills, A., Morgan, A.J., Thomas, R.J., Kille, P., Novo, M., 2016. The worm has turned: behavioural drivers of reproductive isolation between cryptic lineages. Soil Biol. Biochem. 98, 11–17.
Kautenburger, R., 2006. Genetic structure among earthworms (*Lumbricus terrestris* L.) from different sampling sites in western Germany based on random amplified polymorphic DNA. Pedobiologia 50, 257–266.
Kille, P., Andre, J., Anderson, C., Ang, H.N., Bruford, M.W., Bundy, J.G., Donnelly, R., Hodson, M.E., Juma, G., Lahive, E., Morgan, A.J., Sturzenbaum, S.R., Spurgeon, D.J., 2013. DNA sequence variation and methylation in an arsenic tolerant earthworm population. Soil Biol. Biochem. 57, 524–532.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. J. Mol. Evol. 16, 111–120.
King, R.A., Tibble, A.L., Symondson, W.O.C., 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. Mol. Ecol. 17, 4684–4698.
Klarica, J., Kloss-Brandstatter, A., Traugott, M., Juen, A., 2012. Comparing four mitochondrial genes in earthworms − implications for identification, phylogenetics, and discovery of cryptic species. Soil Biol. Biochem. 45, 23–30.
Knott, K.E., Haimi, J., 2010. High mitochondrial DNA sequence diversity in the parthenogenetic earthworm Dendrobaena octaedra. Heredity 105, 341–347.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Lavelle, P., Spain, A.V., 2001. Soil Ecology. Kluwer Academic Publishers.
Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., Rossi, J.P., 2006. Soil invertebrates and ecosystem services. Eur. J. Soil Biol. 42, S3–S15.
Lemon, J., 2006. Plotrix: a package in the red light district of R. R-News 6, 8–12.
Letunic, I., Bork, P., 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23, 127–128.
Liebeke, M., Bruford, M.W., Donnelly, R.K., Ebbels, T.M.D., Hao, J., Kille, P., Lahive, E., Madison, R.M., Morgan, A.J., Pinto-Juma, G.A., Spurgeon, D.J., Svendsen, C., Bundy, J.G., 2014. Identifying biochemical phenotypic differences between cryptic species. Biol. Lett. 10.
Margerie, P., Decaëns, T., Bureau, F., Alard, D., 2001. Spatial distribution of earthworm species assemblages in a chalky slope of the Seine Valley (Normandy, France). Eur. J. Soil Biol. 37, 291–296.
Martinsson, S., Erseus, C., 2017. Cryptic speciation and limited hybridization within Lumbricus earthworms (Clitellata: Lumbricidae). Mol. Phylogenet. Evol. 106, 18–27.
Novo, M., Almodovar, A., Diaz-Cosin, D.J., 2009. High genetic divergence of hormogastrid earthworms (Annelida, Oligochaeta) in the central Iberian Peninsula: evolutionary and demographic implications. Zool. Scripta 38, 537–552.
Oksanen, J., Blanchet, G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2012. Vegan: Community Ecology Package. R package version 2. 0-3.
Pérès, G., Vandenbulcke, F., Guernion, M., Hedde, M., Beguiristain, T., Douay, F., Houot, S., Piron, D., Richard, A., Bispo, A., Grand, C., Galsomies, L., Cluzeau, D., 2011. Earthworm indicators as tools for soil monitoring, characterization and risk assessment. An example from the national Bioindicator programme (France). Pedobiologia 54, S77–S87.
Pérez-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.
Pérez-Losada, M., Ricoy, M., Marshall, J.C., Dominguez, J., 2009. Phylogenetic assessment of the earthworm Aporrectodea caliginosa species complex (Oligochaeta: Lumbricidae) based on mitochondrial and nuclear DNA sequences. Mol. Phylogenet. Evol. 52, 293–302.
Pérez-Losada, M., Bloch, R., Breinholt, J.W., Pfenninger, M., Dominguez, J., 2012. Taxonomic assessment of Lumbricidae (Oligochaeta) earthworm genera using DNA barcodes. Eur. J. Soil Biol. 48, 41–47.
Pansu, J., DE Danieli, S., Puissant, J., Gonzalez, J.M., Gielly, L., Cordonnier, T., Zinger, L., Brun, J.J., Choler, P., Taberlet, P., Cecillon, L., 2015. Landscape-scale distribution patterns of earthworms inferred from soil DNA. Soil Biol. Biochem. 83, 100–105.
Paoletti, M.G., 1999. The role of earthworms for assessment of sustainability and as bioindicators. Agric. Ecosyst. Environ. 74, 137–155.
Plaisance, L., Caley, M.J., Brainard, R.E., Knowlton, N., 2011. The diversity of coral reefs: what are we missing? PLoS One 6.
Porco, D., Bedos, A., Greenslade, P., Janion, C., S.K.A.RŻyński, D., Stevens, M.I., Jansen Van Vuuren, B., Deharveng, L., 2012. Challenging species delimitation in Collembola: cryptic diversity among common springtails unveiled by DNA barcoding. Invertebrate Syst. 26, 470–477.
Porco, D., Decaëns, T., Deharveng, L., James, S., S.K.A.RŻyński, D., Erséus, C., Butt, K., Richard, B., Hebert, P.N., 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. Biol. Invas. 15, 899–910.
Pulleman, M., Creamer, R., Hamer, U., Helder, J., Pelosi, C., Peres, G., Rutgers, M., 2012. Soil biodiversity: biological indicators and soil ecosystem services-an overview of European approaches. Curr. Opin. Environ. Sustain. 4, 529–538.
R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Reynolds, J.W., 1977. The Earthworms (Lumbricidae and Sparganophilidae) of Ontario. Toronto Royal Ontario Museum.
Richard, B., Decaëns, T., Rougerie, R., James, S.W., Porco, D., Hebert, P.D.N., 2010. Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding. Mol. Ecol. Resources 10, 606–614.
Richard, B., Legras, M., Margerie, P., Mathieu, J., Barot, S., Caro, G., Desjardins, T., Dubs, F., Dupont, L., Decaëns, T., 2012. Spatial organization of earthworm assemblages in pastures of northwestern France. Eur. J. Soil Biol. 53, 62–69.
Rocha-Olivares, A., Fleeger, J.W., Foltz, D.W., 2004. Differential tolerance among cryptic species: a potential cause of pollutant-related reductions in genetic diversity. Environ. Toxicol. Chem. 23, 2132–2137.
Rutgers, M., Schouten, A.J., Bloem, J., Van Eekeren, N., DE Goede, R.G.M., Akkerhuis, G., Van Der Wal, A., Mulder, C., Brussaard, L., Breure, A.M., 2009. Biological measurements in a nationwide soil monitoring network. Eur. J. Soil Sci. 60, 820–832.
Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A.M., Neilson, R., Boag, B., Schmidt, O., Murchie, A.K., Blackshaw, R.P., Pérès, G., Cluzeau, D., Guernion, M., Briones, M.J.I., Rodeiro, J., Piñeiro, R., Cosín, D.J.D., Sousa, J.P., Suhadolc, M., Kos, I., Krogh, P.-H., Faber, J.H., Mulder, C., Bogte, J.J., Wijnen, H.J.V., Schouten, A.J., Zwart, D.D., 2016. Mapping earthworm communities in Europe. Appl. Soil Ecol. 97, 98–111.
Saitou, N., Nei, M., 1987. The neighbor-joining method − a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.
Shekhovtsov, S.V., Golovanova, E.V., Peltek, S.E., 2013. Cryptic diversity within the Nordenskiold's earthworm, Eisenia nordenskioldi subsp. nordenskioldi (Lumbricidae, Annelida). Eur. J. Soil Biol. 58, 13–18.
Sims, R.W., Gerard, B.M., 1985. Earthworms: Keys and Notes for the Identification and Study of the Species. Linnean Society of London and the Estuarine and Brackish-Water Sciences Association.
Spurgeon, D.J., Hopkin, S.P., Jones, D.T., 1994. Effects of cadmium, copper, lead and zinc on growth, reproduction and survival of the earthworm eisenia-foetida (Savigny) − assessing the environmental-impact of point-source metal contamination in terrestrial ecosystems. Environ. Pollut. 84, 123–130.
Stürzenbaum, S.R., Andre, J., Kille, P., Morgan, A.J., 2009. Earthworm genomes, genes and proteins: the (re)discovery of Darwin's worms. Proc. R. Soc. B: Biol. Sci. 276, 789–797.
Sturmbauer, C., Opadiya, G.B., Niederstatter, H., Riedmann, A., Dallinger, R., 1999. Mitochondrial DNA reveals cryptic oligochaete species differing in cadmium resistance. Mol. Biol. Evol. 16, 967–974.
Torres-Leguizamon, M., Mathieu, J., Decaëns, T., Dupont, L., 2014. Genetic structure of earthworm populations at a regional scale: inferences from mitochondrial and microsatellite molecular markers in *Aporrectodea icterica* (Savigny 1826). PLoS One 9, e101597.