Zoologica Scripta

Phylogeny and species delimitation of North European *Lumbricillus* (Clitellata, Enchytraeidae)

Mårten J. Klinth, Svante Martinsson & Christer Erséus

Submitted: 20 January 2016
Accepted: 31 March 2016
doi:10.1111/zsc.12187

Klinth, M.J., Martinsson, S. & Erséus, C. (2017). Phylogeny and species delimitation of North European *Lumbricillus* (Clitellata, Enchytraeidae). —*Zoologica Scripta*, *46*, 96–110.

The enchytraeid genus *Lumbricillus* comprises about 80 described species of clitellate worms, which are up to a few centimetres long, and they mostly inhabit the littoral zone of non-tropical marine and brackish waters world-wide. The phylogeny of this genus is poorly studied, but previous work has suggested that *Lumbricillus* is a non-monophyletic group. In this study, species boundaries and the phylogeny of this genus is re-assessed using more than 300 DNA-barcoded specimens (using COI mtDNA), part of which was also sequenced for two additional mitochondrial and four nuclear molecular markers. Statistical and coalescent based applications were used for the delimitation of a total of 24 species, of which 20 were identified as belonging to 17 described morphospecies; one morphospecies was found to be a complex of four delimited species, and another four delimited species could not be matched with any described species. Furthermore, gene trees, concatenation and multispecies coalescent based species trees were estimated using Bayesian inference. The estimated phylogenies confirm a non-monophyletic *Lumbricillus* as *L. semifuscus* is clearly excluded from the genus. Furthermore, the placement of a monophyletic clade consisting of *L. arenarius*, *L. dubius*, and an unidentified species varies between analyses; they are either found as the sister-group to the genus *Grania* or as sister-group to the remaining *Lumbricillus*, where the latter relationship is supported by the multispecies coalescent, which we consider as the most reliable method.

Corresponding author: *Mårten Klinth, Department of Biological & Environmental Sciences, University of Gothenburg, P.O. Box 463, SE-405 30 Gothenburg, Sweden. E-mail: mjeriksson90@gmail.com*

Mårten J. Klinth, Svante Martinsson, and Christer Erséus, Department of Biological & Environmental Sciences, University of Gothenburg, P.O. Box 463, SE-405 30 Gothenburg, Sweden. E-mails: mjeriksson90@gmail.com, svante.martinsson@bioenv.gu.se, christer.erseus@bioenv.gu.se

Introduction

The clitellate family Enchytraeidae (Annelida) contains more than 700 species (Schmelz & Collado 2015). Members of the family are mainly found in terrestrial soils but also occur in many aquatic habitats. Enchytraeids are usually between 5 and 20 mm long and have both a thick cuticle of collagen as well as solid chaetae making them resemble small earthworms. A sister relationship to earthworms (Crassiclitellata) has been supported by some phylogenetic analyses (Siddall *et al.* 2001; Erséus & Källersjö 2004) but another study supports the family Propappidae as the sister to Enchytraeidae (Marotta *et al.* 2008). Thus, the true phylogenetic relationship of Enchytraeidae to other Clitellata remains controversial. One of the about 30 genera within the family Enchytraeidae is *Lumbricillus* Ørsted, 1844 which contains about 75–80 described species, depending on the validity of some subspecies (Nakamura 2000; Schmelz & Collado 2010a). This genus is mainly characterized by having sperm developed in several large, usually regularly lobed, seminal vesicles (testis sacs) as well as some nephridial and spermathecal characters (Nielsen & Christensen 1959; Schmelz & Collado 2010b). A morphology-based phylogeny recovered the small genus *Randidrilus* Coates & Erséus, 1985, together with one species of *Lumbricillus*, as sister group of remaining *Lumbricillus* species (Coates 1989). Another genus, *Grania* Southern, 1913, which was not included in that study, was placed as closely related to *Lumbricillus* in a molecular phylogeny of Enchytraeidae estimated by Erséus *et al.* (2010). In fact, the same study suggested *Lumbricillus* to be paraphyletic, with one of

the six included species [*L. arenarius* (Michaelsen, 1889)] placed as sister to *Grania* rather than together with the remaining *Lumbricillus* species. Erséus *et al.* (2010) did not include any specimens from the genus *Randidrilus*. As that study only included six species of *Lumbricillus* the phylogenetic relationships within *Lumbricillus* remain largely unknown.

The morphological variation among most species within *Lumbricillus* is limited and several species lack sufficient morphological descriptions. This, in combination with conflicting views on the validity of some species and synonymization of nominal taxa makes the delimitation of species problematic. Furthermore, it is well-known that cryptic speciation is common in clitellates (see review by Erséus & Gustafsson 2009), including in enchytraeids (e.g. De Wit & Erséus 2010; Matamoros *et al.* 2012; Martinsson & Erséus 2014). Cryptic species are morphologically indistinguishable, or closely resemble each other and therefore have been identified as the same taxon (Bickford *et al.* 2007).

Species delimitation is a fundamental part of systematic biology, and one of the biggest challenges has been the conflicting views on which species concept/concepts should be used for defining species. De Queiroz (2007) proposed a unified species concept where the only criterion for species is that they are "separately evolving metapopulation lineages" for which previous species concepts may be used as secondary criteria. The more secondary criteria that support the separation of two lineages the stronger the evidence for the lineages belonging to different species (De Queiroz 2007). DNA-barcoding is a popular tool for species delimitation where a short and variable DNA sequence (cytochrome c oxidase subunit 1 or COI for animals) is used to find clusters of organisms which can be considered as putative species (Kress *et al.* 2015). However, using a single genetic marker for species delimitation is known to often overestimate the number of species (e.g. Dasmahapatra *et al.* 2010; Hogner *et al.* 2012; Achurra & Erséus 2013; Martinsson *et al.* 2013, 2015). Furthermore, as COI is a mitochondrial gene it is, with few exceptions, inherited maternally only and thus produces clusters not showing evidence of paternal ancestry (Birky *et al.* 1989). Therefore, preliminary delimitations of individuals into lineages should be validated using further lines of evidence such as more genetic markers or morphological differences (De Queiroz 2007).

In the early days of molecular systematics, phylogenetic studies used single genes with their corresponding gene trees to infer the evolutionary relationships between different species. In other words the gene trees were directly used as species trees (Degnan & Rosenberg 2009). With the increased availability of genetic markers, either the most prevalent gene tree topology or the tree yielded from a concatenation of all DNA sequences has been used as the species tree (Bull *et al.* 1993; Nylander *et al.* 2004). A major problem with inferring species trees directly from gene trees is that the genealogies, meaning the histories of inheritance of the genes, often differ from each other and, most importantly, from the species tree. It has been shown that when the true species tree contains short internal branches, the probability that any sampled gene tree has a topology matching that of the species tree is low (Pamilo & Nei 1988). Consequently, using the most common gene tree or concatenating data can lead to erroneous estimation of the species tree (Degnan & Rosenberg 2006). One way to overcome these shortcomings is to use the multispecies coalescent model (Rannala & Yang 2003). It is based on coalescent theory, which describes the random process in which lineages merge or coalesce when finding their common ancestor, moving backwards in time (Nordborg 2007). The coalescent is efficient to model because instead of following the history of the entire population backwards in time it only generates the random genealogy of the sampled individuals. The coalescent theory has been extended to be used with several genes to estimate the phylogeny of a group of species, in what is known as the multispecies coalescent model. This was first performed by Rannala & Yang (2003) who used the fact that the divergence times between species have to be more recent than the coalescent times for any genes shared between them, assuming no genetic transfer after speciation. This model cannot only be used for estimating species trees, but also population sizes and divergence times (Rannala & Yang 2003). It is also able to handle incongruence between the gene trees (Degnan & Rosenberg 2009). In the multispecies coalescent model, species are treated as independently evolving lineages, defined by abrupt speciation and no gene flow between lineages after separation (Aydin *et al.* 2014), which is consistent with the unified species concept suggested by De Queiroz (2007).

The aim of this study is to delimit the poorly studied North European species of *Lumbricillus*, and at the same time estimate a phylogeny for them, using seven genetic markers. DNA barcoding is used as a tool for preliminary species delimitation which is later evaluated using statistical and coalescent based methods. The phylogeny of *Lumbricillus* and its closest relatives is estimated using concatenated and multispecies coalescent based Bayesian inference. The formal taxonomic revision of genera and species in this complex will be published elsewhere.

Methods

Collection, preservation and preparation of specimens

Enchytraeids used in this study were collected in marine, brackish and limnic habitats, mainly in Norway and

Sweden (Table S1). The posterior parts of the specimens were used for DNA-extraction, whereas the anterior ends were stained in paracarmine, dehydrated in xylene and mounted in Canada balsam (as described by Erséus 1994), for morphological examination and to serve as vouchers. The morphology of the mounted specimens was examined under a compound microscope and specimens were identified to species using original descriptions and the revision by Nielsen & Christensen (1959). In total, 310 specimens of *Lumbricillus* were sequenced for at least one of the molecular markers and included in this study. In addition, the following 11 taxa were used as outgroups: *Achaeta bibulba* Graefe, 1989, *Cernosvitoviella minor* Dózsa-Farkas, 1990, *Chamaedrilus cognettii* (Issel, 1905), *Grania crassiducta* Coates, 1990, *Grania galbina* De Wit & Erséus, 2007, *Grania ovitheca* Erséus, 1977, *Grania pusilla* Erséus, 1974, *Grania variochaeta* Erséus & Lasserre, 1976, *Henlea ventriculosa* (Udekem, 1854), *Marionina communis* Nielsen & Christensen, 1959 and *Mesenchytraeus flavus* (Levinsen, 1883). The outgroups were selected based on the phylogeny by Erséus *et al.* (2010). Some sequences were downloaded from GenBank but most were newly generated for this study as listed in Table S1. Vouchers are deposited in the Swedish Museum of Natural History (SMNH), Stockholm, Sweden, or in the University Museum of Bergen (ZMBN), Bergen, Norway.

DNA extraction/amplification/sequencing

For the majority of the material studied, DNA was extracted from tissue samples following instructions from the manufacturers using either the QuickExtract DNA Extraction Solution 1.0 (Epicentre, Madison, WI, USA), EZNA Tissue DNA kit (Omega Bio-Tek, Norcross, GA, USA), or DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany). The extracts were later used for DNA amplification by PCR, using either the PuRe-Taq Ready-To-Go (GE Healthcare, Chalfont St. Giles, UK) or Red Taq DNA Polymerase Master Mix (VWR, Haasrode, Belgium) in 25 μL reactions. From the mitochondrial genome, parts of the following genes: *cytochrome oxidase C subunit I* (COI), *12S ribosomal RNA* (12S) and *16S ribosomal RNA* (16S) were amplified and from the nuclear genome the complete *Internal Transcribed Spacer region* (ITS), and 18S *ribosomal RNA gene* (18S), as well as parts of *28S ribosomal RNA* (28S) and *Histone 3 genes* (H3) were amplified, using primers and PCR protocols described in Table S2. COI was amplified for all specimens, and the other markers were amplified for subsets of worms only (Table S1). The PCR products were examined using 1% agarose gel electrophoresis, and successful products were purified using exonuclease I and FastAP thermosensitive alkaline phosphatase. The purified products were sequenced by MWG Eurofins Operon (Edersberg, Germany) or Macrogen (Geumcheon-Gu, Seoul, Korea). Some material was COI barcoded by the Canadian Centre for DNA barcoding (CCDB) (Guelph, Canada). Resulting trace files were assembled in Geneious 6.1.8 (created by Biomatters; available from http://www.geneious.com). All sequences were aligned using default settings in MAFFT (Katoh *et al.* 2002) as implemented within Geneious.

Recombination test

Alignments in FASTA format of all genes and concatenations were opened in SplitsTree4 (Huson & Bryant 2006), a program that computes phylogenetic networks and can be used to detect hybridization and recombination. Phi-tests were conducted to search for data heterogeneity which could be due to recombination.

Species delimitation

Sequences of COI, ITS and H3 were used for species delimitation, each alignment representing different numbers of specimens. The alignments were analysed in MEGA 5.10 (Tamura *et al.* 2011), first for selection of evolutionary model, using the Bayesian Information Criterion (BIC), second for pairwise distances both with raw distances (p-distance) and distances based on the appropriate evolutionary model. If the selected evolutionary model was not implemented in MEGA, the closest model in terms of number of parameters was chosen. All analyses in MEGA used partial deletion. A gamma distribution of rates among sites was selected, other parameters were used with default settings. The following evolutionary models were selected for the distance analyses: COI: Tamura-Nei+Gamma (TN93), ITS: Tamura 3-parameter+Gamma (T92), H3: Tamura 3-parameter+Gamma (T92).

The distances were analysed in ABGD (Automatic Barcode Gap Discovery) (Puillandre *et al.* 2012) using default values except for the following: *Pmin*: 0.01, *Pmax*: 0.2 and *Steps*: 20, using the initial partition to delimit the specimens into molecular operational taxonomic units (MOTUs) (Floyd *et al.* 2002). As only COI showed a clear barcode gap and was consistent in the number of MOTUs yielded from the ABGD analysis, these results were used for initial species hypotheses to be further tested.

Estimated gene trees for COI, H3 and ITS (see Results) were imported into Geneious and the species hypotheses from ABGD were tested using the species delimitation plug-in (Masters *et al.* 2011). The hypotheses were evaluated using the results from the *Rosenberg's* P_{AB} and the *P (Randomly Distinct)* statistical tests. *Rosenberg's* P_{AB} tests whether the monophyly of the selected group in the gene tree is due to an evolutionary process or if it is the result of random branching of the tree caused by an insufficient

sample size (Rosenberg 2007). *P(Randomly Distinct)* uses the ratio between the length of the branches within a selected group and the length between selected groups to estimate whether the groups' distinctiveness is due to an evolutionary process or to random coalescent processes. This test cannot be and, therefore, was not performed for species with only one sampled individual (Rodrigo *et al.* 2008).

Further species delimitation of the species hypotheses from ABGD was performed in DISSECT (Division of Individuals into Species using Sequences and Epsilon-Collapsed Trees), a novel method that uses Bayesian inference to co-estimate the species tree and the species delimitation by setting a prior distribution on node heights with a high peak close to zero; the node will only be treated as split into two species if the posterior distribution of a node is significantly separated from this peak (Jones *et al.* 2015). Specimens sequenced for COI, H3 and ITS, in total 72 specimens grouped into 32 preliminary species hypotheses (including outgroups), were included in the DISSECT analysis. Specimens from *L. tuba* Stephenson, 1911, CE838 *Cernosvitoviella minor* and CE699 *Grania ovitheca* all lacked sequences of H3 and could thus not be included. The DISSECT analysis was run through *BEAST (Heled & Drummond 2010) using the same settings and priors as in the species tree estimation (see below) with the species tree prior changed into a birth-death process. In the xml-file the ploidy level was changed to allow for hermaphrodites (see Species tree estimation below), birthDeathModel was replaced with a birthDeathCollapseModel, and an operator was added for the origin height, according to the supplementary material by Jones *et al.* (2015). The ε-parameter, which determines the collapsing node height, was assigned the recommended default value of 0.0001. The mcmc was set to run for 500 million generations sampling every 50 000, thus generating 10 000 trees. Resulting log-file was evaluated in Tracer v1.5 (Rambaut & Drummond 2007), and the sampled trees were run through SpeciesDelimitationAnalyser (Jones *et al.* 2015) using default values and removing the first 1000 trees as burn-in (determined sufficient after examining the log-file).

Gene tree estimation

Gene trees were estimated for each of the seven genes in MrBayes v.3.2.4 (Ronquist *et al.* 2012). COI and H3 were partitioned according to codon position and ITS was by eye roughly partitioned into ITS1, 5.8S and ITS2. All partitions were unlinked to allow for different base frequencies, shape of the gamma distribution, proportion of invariable sites and substitution rates, using nst = mixed, rates = invgamma and brlenspr = unconstrained:Exp(100). The mcmc was set to run for 10 million generations sampling every 10 000 generations. Consensus trees were summarized with 25% discarded as burnin. Resulting p-files were examined in Tracer v1.5 to evaluate convergence and to ensure sufficient burn-in for the trees. Consensus trees were viewed in TreeGraph 2 (Stöver & Müller 2010) and edited in GIMP 2.8.10.

Concatenated phylogenetic analyses

Three separate concatenated analyses were set to run in MrBayes v.3.2.4 using the combined information of the mitochondrial, nuclear and all genes, respectively. In the mitochondrial analysis 49 specimens were selected (including outgroups) and concatenated with sequences for 12S, 16S and COI, with exceptions: CE3502 *L. rutilus* Welch, 1914 and CE2549 *L. rubidus* Finogenova & Streltsov, 1978, both lacking 12S. The mitochondrial concatenation was partitioned according to gene and COI was further partitioned according to codon position (five partitions in total). In the nuclear ribosomal analysis 48 specimens were selected (including outgroups) and concatenated with sequences for 18S, ITS and 28S. The nuclear ribosomal concatenation was partitioned according to gene and ITS was further roughly partitioned into ITS1, 5.8S and ITS2 (five partitions in total). In the total concatenated analysis: 47 specimens were selected (including outgroups) and concatenated with sequences for 12S, 16S, 18S, 28S, COI, H3 and ITS, with exceptions: CE3502 *L. rutilus* and CE2549 *L. rubidus* both lacking 12S; CE2246 *L. sp. G*, CE838 *Cernosvitoviella minor*, CE699 *Grania ovitheca*, CE879 *L. tuba*, CE664 *L. lineatus* (Müller, 1774) and CE986 *L. lineatus*, all lacking H3. The total concatenation was partitioned according to gene, ITS was further roughly partitioned into ITS1, 5.8S and ITS2, COI and H3 were further partitioned according to codon position (13 partitions in total). For all the concatenations the partitions were unlinked, used nst = mixed, and the same settings were applied as for the gene tree estimations (see above) except for running 20 million generations with sampling every 20 000 generations (Mitochondrial and Nuclear ribosomal), or 50 million generations with sampling every 50 000 generations (Total concatenation). Consensus trees were summarized with 25% discarded as burnin. Resulting p-files were examined in Tracer v1.5 to evaluate convergence and to ensure sufficient burn-in for the trees. Consensus trees were viewed in TreeGraph 2 and edited in GIMP 2.8.10.

Coalescent based species tree estimation

Alignments were prepared with 46 specimens (including outgroups) for the seven gene markers. The following specimens lacked some of the genes: CE3502 *L. rutilus* and CE2549 *L. rubidus* both lacking 12S, CE2248 *L. semifuscus* (Claparéde, 1861) lacked 18S, CE2497 *L. pagenstecheri A* [one of four species identified as *L. pagenstecheri* (Ratzel,

1869)] lacked COI, CE2246 *L. sp. G*, CE664 *L. lineatus* and CE986 *L. lineatus*, all lacking H3. *Lumbricillus tuba* lacked H3 due to sequencing problems and could thus not be included in the *BEAST analysis. The same occurred for the two outgroups CE838 *Cernosvitoviella minor* and CE699 *Grania ovitheca*. The alignments were imported into BEAUti v1.8.0, part of the BEAST-package (Drummond *et al.* 2012), where the *BEAST option was selected to allow for species tree estimation under the multispecies coalescence model (Heled & Drummond 2010). The specimens were grouped into 32 species according to the results from the species delimitation, using a prepared traits file. Each gene was kept unlinked for site, clock and tree models. Default settings and priors were used with the following exceptions: Sites: Each partition was given the substitution model selected by both the model test in MEGA and the results from the reversible model jump in MrBayes. When two different models were favoured or tied for best fit, the more complex of the two was selected. The selection was as follows: GTR+G+I for 12S, 16S and COI. TN93 + G+I for 18S, 28S, ITS and H3. Base frequencies were estimated. Clocks: The evolutionary rate of COI was set to 1 and the rates of all other genes estimated in relation to this. All clock models were set to lognormal relaxed uncorrelated clocks. Trees: The Yule process was selected for species tree prior, and piecewise linear & constant root for population size model. For 12S, 16S and COI genes ploidy type was changed to mitochondrial and all genes were set with UPGMA starting trees. Priors: Clock priors (ucld.mean) were set as uniform ranging from 0 to 2 with an initial value of 1 for 12S, 16S and ITS, but ranging from 0 to 1 with an initial value of 0.5 for 18S, 28S and H3 (all in relation to the set rate of 1 for COI). These settings were based on previous studies on insects (Lin & Danforth 2004; Danforth *et al.* 2005) confirmed for clitellates (Martinsson & Erséus 2014). The priors for species.popMean and species.yule.birthRate were changed from the improper default settings, which do not sum up to 1, to lognormal distributions with default values. The effective population size for the mitochondrial markers was scaled to half of that of the nuclear, by changing the "ploidy" level manually, this as these worms are hermaphrodites and thus have two potential mothers which can contribute the mitochondrial genes to the next generation, making the effective population size twice that of mitochondrial genes from non-hermaphrodites, but still only half of that of nuclear markers. The xml-file was run for 500 million generations in BEAST v1.8.1 (Drummond *et al.* 2012) sampling every 500 000 generations. Resulting log-file was evaluated in Tracer v1.5, and the tree file was run through Treeannotator v1.7.4. (Drummond *et al.* 2012) removing the initial 10% as burn-in (determined sufficient after examining the log-file) and later viewed in Figtree v1.4.0. (Rambaut 2009) and edited in GIMP 2.8.10.

Results

DNA sequencing

Sequences were successfully obtained for the majority of the seven genes from all putative *Lumbricillus* species and some outgroups (Table S1). 304 sequences were obtained for COI, 87 for ITS, 71 for H3, 34 for 28S, 33 for 18S, 66 for 16S and 32 for 12S. There were some problems with sequencing H3 for *L. verrucosus* (Claparéde, 1861) and *L. buelowi* Nielsen & Christensen, 1959; reads from the reverse primer H3R were obtained from both species, but reads from the forward primer (H3F) were only rarely obtained for *L. verrucosus* and never obtained for *L. buelowi*. Furthermore, sequences for H3 were obtained for *L. tuba*, but they were not possible to align with the remaining H3 sequences and therefore were excluded from the analyses.

Recombination test

The phi-test did not find statistically significant evidence for recombination in any of the genes or concatenated alignments.

Species delimitation

A global barcoding gap was found for the pairwise distances of COI approximately between 6.5 and 9% of genetic distance (Fig. S1A). ITS showed a similar tendency between 7 and 13% but without any clear gap (Fig. S1C), whereas H3 showed no indication of a barcoding gap (Fig. S1B). The initial partitions of the ABGD analyses for COI yielded 24 groups when using both p-distances and TN93-distances, presented as species hypotheses in Table 1. The same 24 groups were found monophyletic (where applicable, that is, not for singletons) in both the COI and ITS gene trees which are further presented below (Figs 1 and 2). The results from the statistical tests, *Rosenberg's* P_{AB} and *P(Randomly Distinct)*, in the species delimitation plug-in in Geneious are also presented in Table 1. Most of the proposed species hypotheses provided by ABGD are supported by the *Rosenberg's* P_{AB} and *P(Randomly Distinct)* statistical tests (Table 1). In H3, some species hypotheses were found paraphyletic and thus could not be statistically tested. Species hypotheses for groups with only one sequenced specimen for the gene could not be tested. In these cases, *P(Randomly Distinct*) is unusable and *Rosenberg's* P_{AB} requires the selected specimen to have a large distance to its closest relative to get statistical significance.

The results from the species delimitation using DISSECT are also presented in Table 1. Delimitations for all the included species were found with the highest support,

Table 1 Results from the species delimitation plug-in implemented in Geneious. Species hypotheses yielded from ABGD using genetic distances in COI listed under Species. *P*-values from the statistical tests of the species delimitation plug-in displayed for each of the three genes as well as posterior probabilities from DISSECT. Grey shade indicates statistical significance (*P*-values ≤ 0.05 or PP ≥ 0.95) for P(Randomly Distinct) (P_{RD}) and Rosenberg's P_{AB} or the highest support in DISSECT

	COI			H3			ITS			
Species	M	P_{RD}	P_{AB}	M	P_{RD}	P_{AB}	M	P_{RD}	P_{AB}	DISSECT
L. arenarius (Michaelsen, 1889)		0.05	<0.001		0.05	0.01		0.05	0.17	1
L. buelowi Nielsen & Christensen, 1959		0.05	<0.001		0.87	<0.001		0.05	<0.001	1
L. dubius (Stephenson, 1911)		0.07	<0.001		0.05	0.17		0.05	0.004	1
L. fennicus Nurminen, 1964		0.05	<0.001	S		0.1 S	S	NA	0.17	1
L. cf. helgolandicus (sensu Nielsen & Christensen, 1959 nec Michaelsen, 1934)		0.05	<0.001		0.18	<0.001		0.05	<0.001	1
L. kaloensis Nielsen & Christensen, 1959		0.05	<0.001		0.77	<0.001		0.05	0.17	1
L. knoellneri Nielsen & Christensen, 1959		0.05	<0.001		NA	NA		0.05	<0.001	1
L. lineatus (Müller, 1774)		0.05	<0.001		NA	NA		0.05	<0.001	1
L. pagenstecheri A (Ratzel, 1869)		0.05	0.02		0.05	0.05		0.05	0.01	1
L. pagenstecheri B		0.05	0.01		0.05	0.05		0.05	0.005	1
L. pagenstecheri C		0.05	0.02		0.05	0.05		0.05	0.003	1
L. pagenstecheri D		0.05	<0.001		NA	NA		0.05	0.01	1
L. pumilio Stephenson, 1932		0.05	<0.001	S	NA	0.33	S	NA	0.33	1
L. rivalis (Levinsen, 1883)		0.05	<0.001		0.05	<0.001		0.05	0.004	1
L. rubidus Finogenova & Streltsov, 1978		0.63	<0.001		0.05	0.33		0.05	0.33	1
L. rutilus Welch, 1914		0.05	<0.001		0.05	<0.001		0.05	<0.001	1
L. semifuscus (Claparéde, 1861)		0.05	<0.001		0.05	<0.001		0.05	<0.001	1
L. tuba Stephenson, 1911		0.05	<0.001	–	–	–		0.05	<0.001	1
L. verrucosus (Claparéde, 1861)		0.05	<0.001		NA	NA		0.05	0.004	1
L. viridis Stephenson, 1911		0.98	<0.001		0.05	0.05		0.05	<0.001	1
L. sp. E		0.05	<0.001		0.07	0.001		0.05	<0.001	1
L. sp. F		0.05	<0.001	S	NA	0.1	S	NA	0.001	1
L. sp. G		0.09	<0.001	S	NA	0.003		0.05	<0.001	1
L. sp. H		0.05	<0.001	S	NA	0.17	S	NA	0.17	1

M stands for monophyly where a grey shade indicates monophyly of the clade in the gene tree, not applicable (NA) for species with only a single sampled individual (marked with an S) for which the P(Randomly Distinct) test is unusable. H3 was completely lacking for *L. tuba*, indicated by a dash (–).

including the singletons, which had limited or no support in the statistical tests implemented in Geneious. The resulting species tree from the DISSECT analysis matched the topology of the species tree generated using all seven genes well (see Species tree estimation below). The only differences were in the placement of the outgroups which could be explained by the inclusion of the additional mitochondrial and nuclear ribosomal genes in the latter analysis, providing a better resolution of the deeper branching's in the phylogeny. Out of the 24 delimited species, 20 were morphologically identified with existing nominal names. The species delimitation allowed for some specimens previously assigned to *L. lineatus* to be separated and identified as *L. verrucosus*, a species previously synonymized with *L. lineatus*. Cryptic lineages were found in *L. pagenstecheri* where the four delimited species all morphologically match the description of *L. pagenstecheri*. Additionally, four of the delimited species could not be matched with any known species, ether due to lack of good material, or because they may represent undescribed species. A list of the identified species can be found in Table 1.

Gene tree estimation

All gene trees showed high ESS (Effective Sample Size) values and good convergence. COI is highly variable and clearly grouped individuals into morpho-species. To some extent it was able to determine close relationships between species but the COI tree showed poor resolution for deeper branches in the phylogeny (Fig. 1). Therefore, it will not be included when discussing the phylogenetic results of the gene trees. *Lumbricillus* was not found monophyletic in any of the gene trees, *L. semifuscus* was always found among

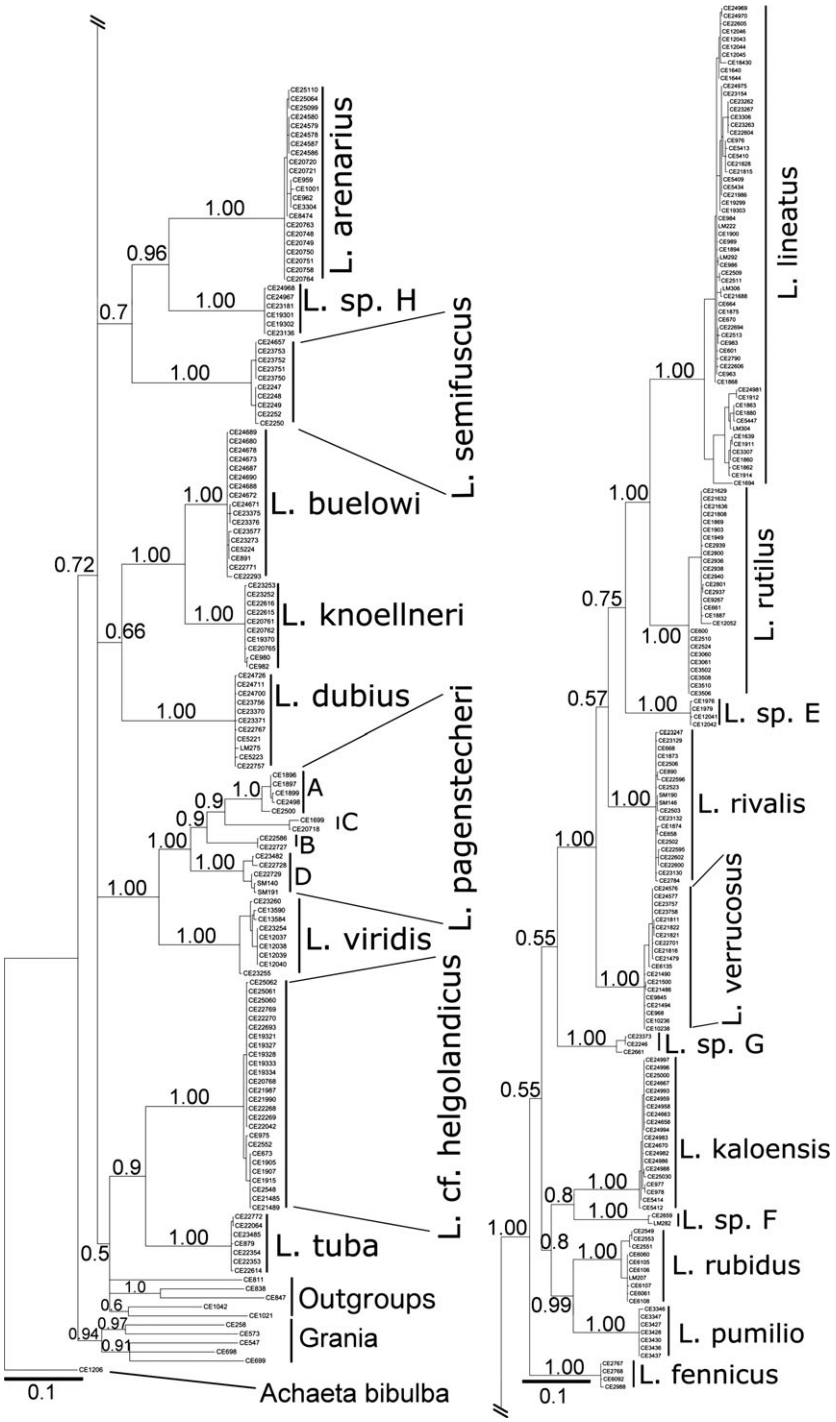

Fig. 1 COI gene tree of all COI-barcoded specimens in this study, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

the outgroups, well separated from all other *Lumbricillus*, here referred to as *Lumbricillus* s. lat. However, all gene trees supported the majority of the *Lumbricillus* s. lat. species as one monophyletic group, here referred to as *Lumbricillus* s. str. (Figs 1, 2 and S2–S6). This group includes all species except *L. arenarius*, *L. dubius* (Stephenson, 1911) and *L.* sp. H (together referred to as the *L. arenarius* group). The *L. arenarius* group was found as a more

© 2016 Royal Swedish Academy of Sciences, **46**, 1, January 2017, pp 96–110

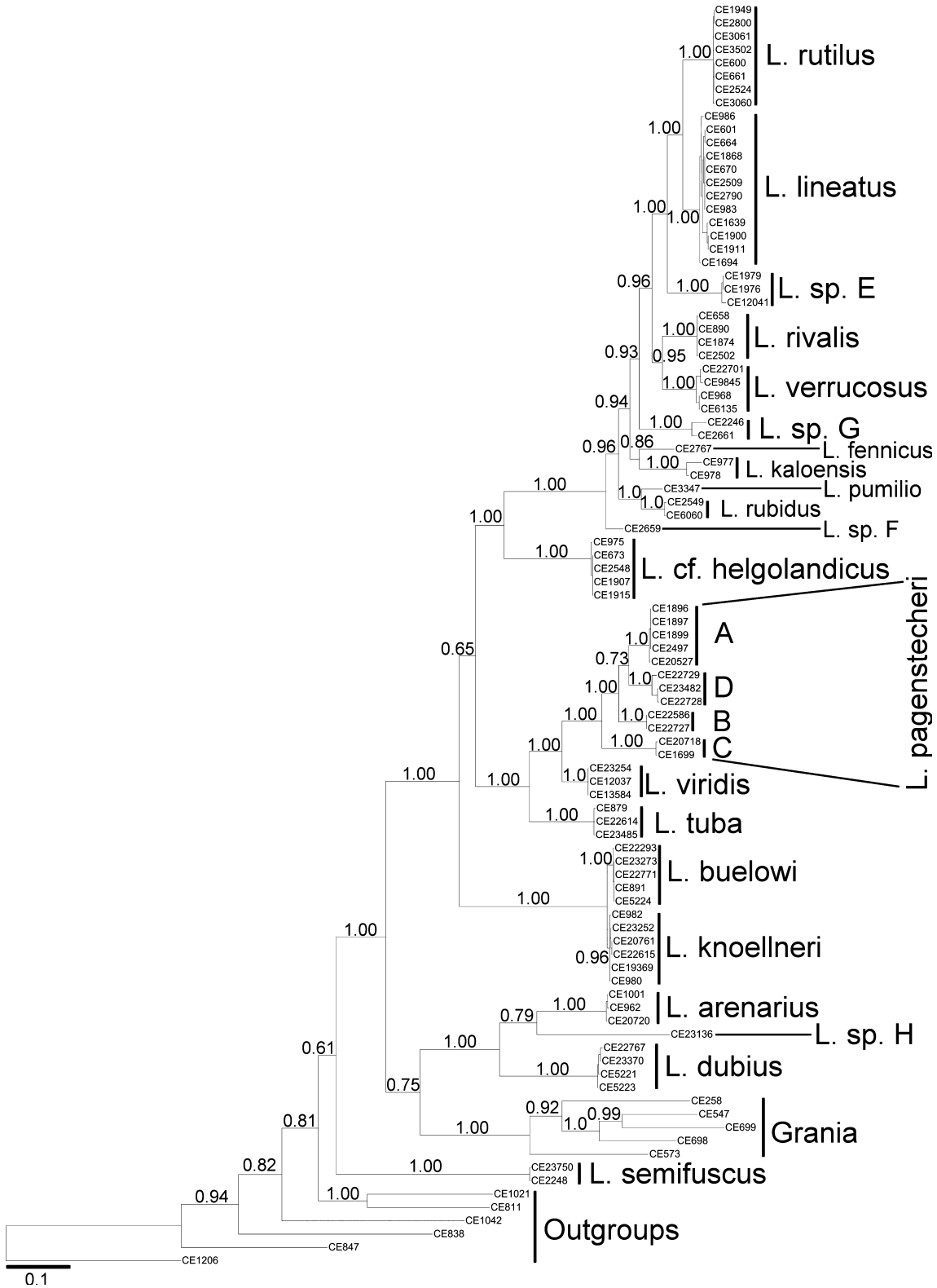

Fig. 2 ITS gene tree, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

or less well supported monophyletic group in all gene trees, except for 12S (Fig. S2). However, the placement of this group in the phylogeny varied; most trees placed it as sister to *Grania* (ITS, 12S, 18S and 28S; Figs 2, S2, S4 and S5), whereas in two trees (16S and H3) it was placed as sister to *Lumbricillus* s. str. (Figs S3 and S6). Furthermore, these differing topologies were mostly well supported in the respective trees. When the *L. arenarius* group was placed together with *Grania*, these together made up the sister of *Lumbricillus* s. str., but when the *L. arenarius* group was placed as sister to *Lumbricillus* s. str., *Grania* was not found to be the closest outgroup to the two. In the 16S tree *Achaeta bibulba* was found within *Grania*, rendering the latter paraphyletic (Fig. S3). Within *Lumbricillus* s. str. most gene trees placed *L. buelowi* and *L. knoellneri* Nielsen & Christensen, 1959 as sister to the rest, except in 12S where *L. buelowi* and *L. knoellneri* were nested within one of three groups that made up a basal trichotomy of *Lumbricillus* s. str. (Fig. S2), and in H3, where *L. viridis* Stephenson, 1911 and the *L. pagenstecheri* species complex (including *L. pagenstecheri* A–D) were sister to the rest (Fig. S6). The support for *L. buelowi* and *L. knoellneri* as sister to the rest of *Lumbricillus* s. str. was low in all but the 16S tree (Fig. S3). All gene trees further contained a group within *Lumbricillus* s. str. containing all the species without distinct ampulla in their spermathecae, viz.: *L. fennicus* Nurminen, 1964, *L. kaloensis* Nielsen & Christensen, 1959, *L. lineatus*, *L. pumilio* Stephenson, 1932, *L. rivalis* (Levinsen 1883), *L. rubidus*, *L. rutilus*, *L. verrucosus*, *L.* sp. E, *L.* sp. F and *L.* sp. G. This grouping was usually well supported, except in the 28S tree (Fig. S5), and has *L.* cf. *helgolandicus* (sensu Nielsen & Christensen, 1959 nec Michaelsen, 1934) as sister to it in all but the 12S and 16S trees (Figs S2 and S3).

Concatenated phylogenetic analyses

All of the three concatenated analyses showed high ESS-values and good convergence. Similarly to the gene trees, *Lumbricillus* was not found monophyletic in any of the concatenated sequence trees, i.e., *L. semifuscus* was always found together with the outgroups, although its placement varied and was never well supported (Figs 3 and S7–S8). The *L. arenarius* species group (Figs 3 and S7–S8, Clade C,) formed a monophyletic clade with maximum support in all trees. In the mitochondrial concatenation this group was the sister-group to *Lumbricillus* s. str. (Fig. S7), whereas the nuclear ribosomal concatenation supported it as sister to *Grania* where the two in turn made up the sister of *Lumbricillus* s. str. (Fig. S8). Interestingly, the mitochondrial concatenation did not even support *Grania* as the sister to *Lumbricillus* s. lat. (Fig. S7). The concatenation of all genes (Fig. 3) supported the grouping of the *L. arenarius* species group together with *Grania*, where these two in turn made up the sister of *Lumbricillus* s. str. The *Lumbricillus* s. str. clade was found monophyletic with maximum support in all three trees (Figs 3 and S7–S8, Clade A). Within this group, *L. buelowi* and *L. knoellneri* formed a clade that was always found as sister-group to the rest with good support. This was followed by another well supported clade which was the sister-group to the remaining *Lumbricillus* s. str.; it contained *L. tuba*, *L. viridis* and the *L. pagenstecheri* species complex. In all three concatenations *L. viridis* was found as sister to the *L. pagenstecheri* species complex with *L. tuba* sister to these. As with the gene trees, all three concatenations also found a well-supported clade within *Lumbricillus* s. str. containing all species without distinct ampulla in their spermathecae (Figs 3 and S7–S8, Clade B), and *L.* cf. *helgolandicus* seemed to be the sister to this group, although not well supported in the mitochondrial concatenation (Fig. S7). Within the group lacking distinct spermathecal ampullae (Clade B), the topology of the phylogeny varied between the different concatenated sequence trees, with the exception that *L. pumilio* and *L. rubidus* always formed a well-supported clade.

Species tree estimation

The species tree yielded by *BEAST showed high ESS-values and good convergence for most parameters (likelihood included) but only moderate support for the posterior and prior parameters. Once again *Lumbricillus* was not monophyletic as *L. semifuscus* was found among the outgroups, but any closer placement was not possible (Fig. 4). In contrast to the tree from the concatenation of all genes (Fig. 3), the *BEAST species tree placed the *L. arenarius* species group (Fig. 4, Clade C) as sister to *Lumbricillus* s. str. (Fig. 4, Clade A) and *Grania* as sister to these two (but with no support). *Lumbricillus* s. str. was once again found with high support. As with the concatenation of all genes, it seemed that *L. buelowi* and *L. knoellneri* were sister to the rest of *Lumbricillus* s. str., followed by *L. viridis* together with the *L. pagenstecheri* species complex as sister to the remaining *Lumbricillus* s. str. Within the latter, *L.* cf. *helgolandicus* was the sister to the well-supported group without distinct spermathecal ampullae (Fig. 4, Clade B). The group without distinct ampullae displayed here the same topology as in the tree from the concatenation of all genes (Fig. 3) with varying support. The sister relationship between *L. pumilio* and *L. rubidus* was well supported as was the one between *L. rivalis* and *L. verrucosus*.

Discussion

This is the most extensive phylogeny of the genus *Lumbricillus* to date, considering the amount of included specimens, species and genetic markers. The data used in this study have made it possible both to establish the boundaries and to resolve the phylogeny of the included species.

© 2016 Royal Swedish Academy of Sciences, **46**, 1, January 2017, pp 96–110

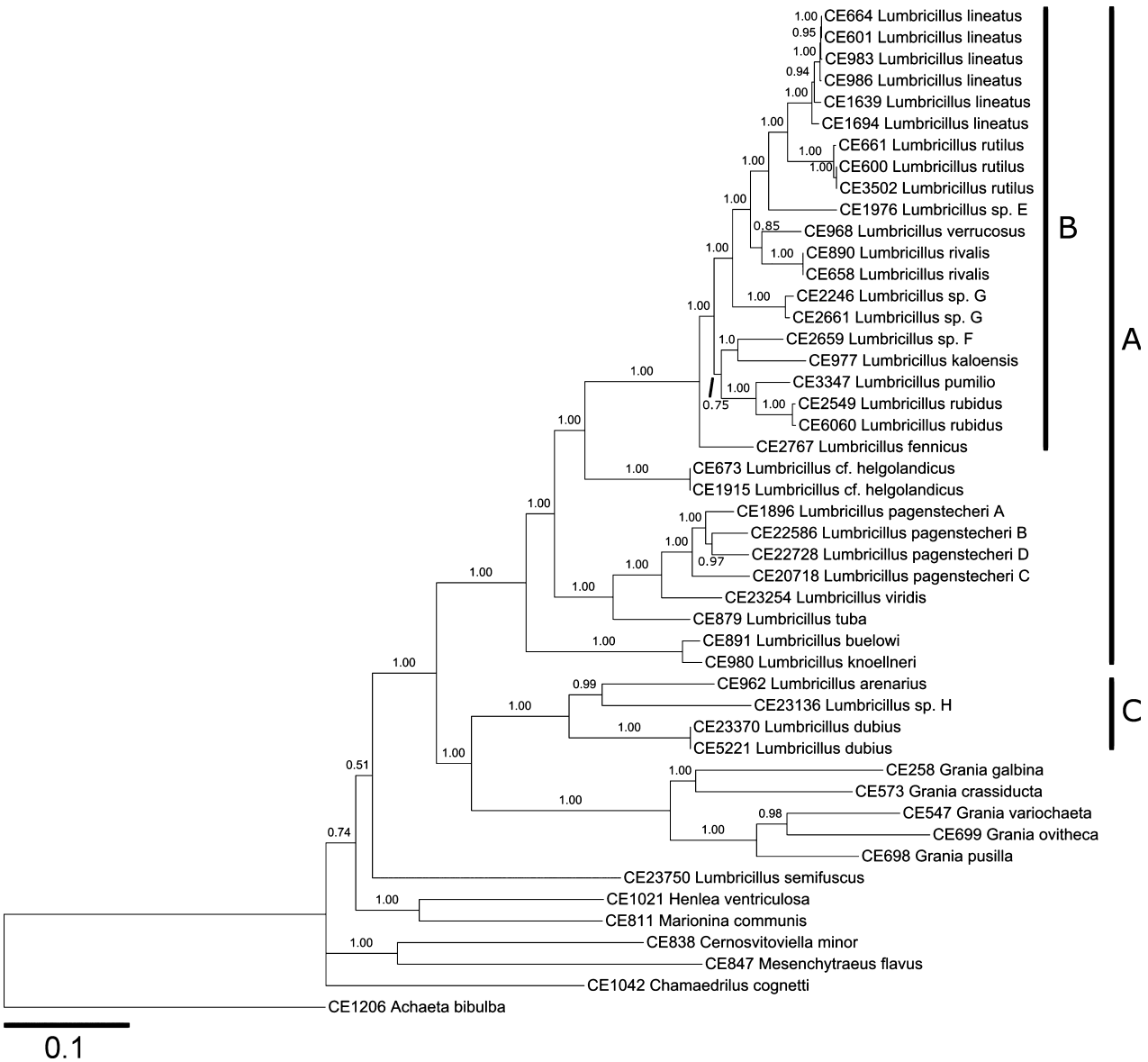

Fig. 3 Concatenated sequence tree of 12S, 16S, COI, 18S, 28S, ITS and H3 genes, estimated using Bayesian inference. —A. *Lumbricillus* sensu stricto. —B. The clade within *Lumbricillus* sensu stricto with spermathecae lacking distinct ampulla. —C. *L. arenarius* species group. Support values are posterior probabilities. Scale shows expected number of changes per site.

Main conclusions, phylogeny

As previously found by other authors (e.g. Kossmagk-Stephan 1983; Erséus *et al.* 2010) *Lumbricillus* is non-monophyletic and in need of revision. Our results clearly support the monophyly of both *Lumbricillus* s. str. and the *L. arenarius* species group. However, the *L. arenarius* species group is not conclusively found as the sister to *Lumbricillus* s. str. as some gene trees and the concatenated species tree supports this group as sister to *Grania*. The species tree based on the multispecies coalescent, which should be the most reliable method, does support a sister relationship between *Lumbricillus* s. str. and the *L. arenarius* species group. Nevertheless, the observed incongruence should not be ignored and the addition of more molecular markers may be required to find the true relationship between *Grania*, *Lumbricillus* s. str. and the *L. arenarius* species group. The reason for this uncertainty in the phylogeny originates in the incongruence of the gene trees, possibly due to incomplete lineage sorting remaining from a rapid diversification between the three groups. Unfortunately, there were no species of *Randidrilus* included in this study, a genus which based on its morphology has been considered as

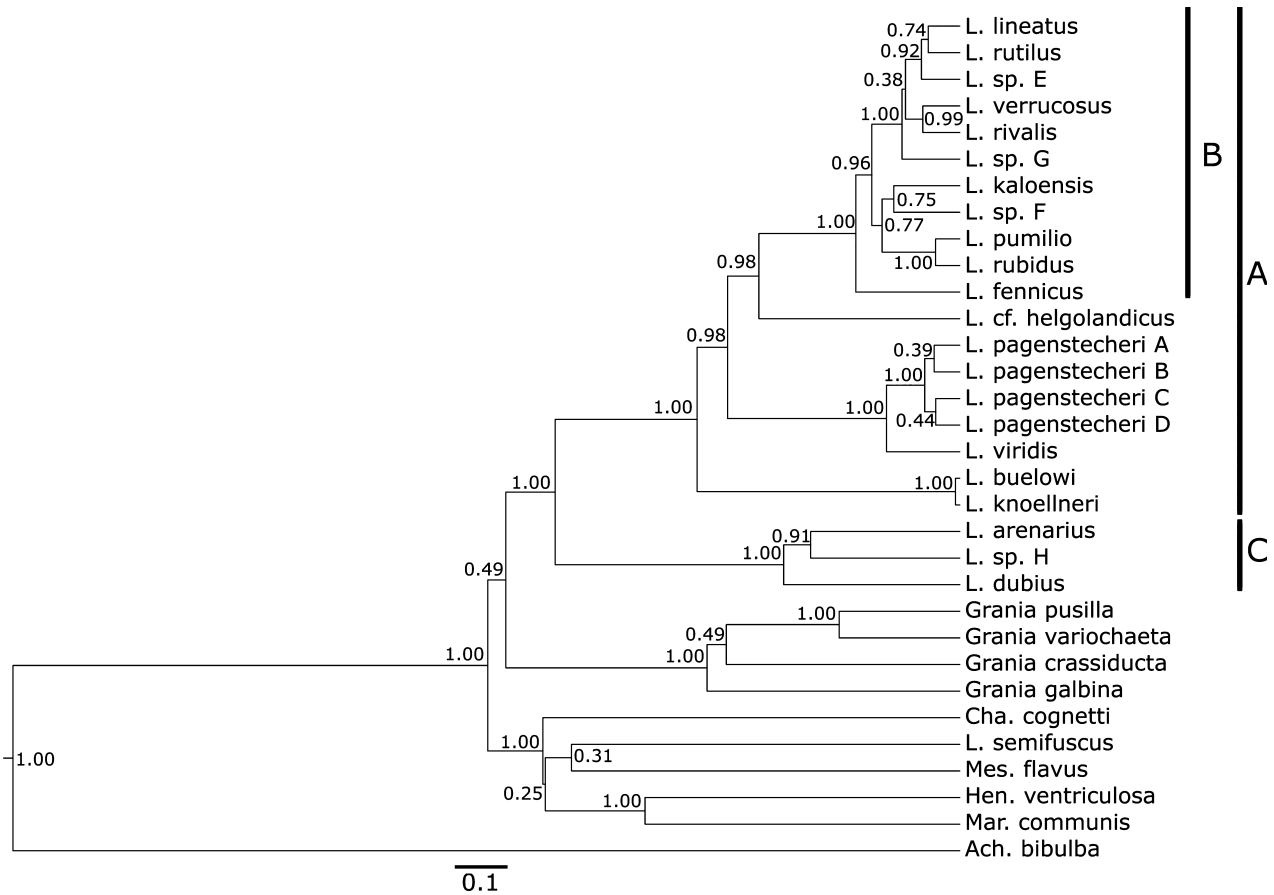

Fig. 4 Species tree based on 12S, 16S, COI, 18S, 28S, ITS and H3 genes, estimated using Bayesian inference under the multispecies coalescent model in *BEAST. —A. *Lumbricillus* sensu stricto. —B. The clade within *Lumbricillus* sensu stricto with spermathecae lacking distinct ampulla. —C. *L. arenarius* species group. Support values are posterior probabilities. Scale shows expected number of changes per site in COI with all other genes relative to it.

closely related to *Lumbricillus* (Coates 1989). If *Randidrilus* indeed is closely related to *Lumbricillus* s. lat. and/or to *Grania*, the addition of members from it could provide important missing pieces to better resolve the phylogeny of all these groups. This study has also confirmed genetically that *L. semifuscus* is not a member of *Lumbricillus* s. lat., but none of the included outgroups seems to be closely related to this species and a more thorough sampling of outgroups is required to find out where this species truly belongs.

Gene trees and incongruence

The gene trees showed similar topologies concerning the general groupings but with a lot of, mostly poorly supported, incongruences within some of them. The only major incongruence between all the gene trees regards the placement of the *L. arenarius* species group (also containing *L. dubius* and *L.* sp H) where 16S and H3 support this group as sister to *Lumbricillus* s. str., whereas 12S, 18S, 28S and ITS support a sister relationship with *Grania* instead. We expected 18S, 28S and ITS to show the same pattern as they are all nuclear ribosomal genes with a linked inheritance. Strangely, the same pattern was not observed between 12S and 16S which are both ribosomal genes of the mitochondrial genome and thus should share the same evolutionary history. Here it should be noted that in the 16S gene tree (Fig. S3), *Grania* was found paraphyletic as it included *Achaeta bibulba*. A closer look at the alignment for this gene showed no clear resemblance between the *Grania* species and *Achaeta bibulba*, or any suggestions of a misalignment. Further, the *A. bibulba* 16S sequence is more similar to other *Achaeta* spp. than to any of those of other *Grania* spp. (C. Erséus, unbublished data). The 16S gene tree shows *Achaeta bibulba* with by far the longest branch length and its placement within *Grania* can probably be explained as a case of long branch attraction (see Bergsten 2005). Incongruence among gene trees is well known and the cases reported in our study could (alternatively) be explained by incomplete lineage sorting, perhaps due to a

rapid diversification between the *Lumbricillus* s. lat. species. Other possible sources for incongruence include recombination, hybridization and gene duplication. We did not find any support for recombination within or between the included genes based on the networks and statistical tests calculated in SplitsTree 4. Furthermore, there were no clear cases where the position of a species alternated between two different well-supported sister relationships, something we would have expected for species of hybrid origin. This does not prove that there are no hybrid species present within *Lumbricillus* s. lat., but additional individuals and genes are required to discover such patterns. Finally, there have not been any recent studies regarding the ploidy level of the species within *Lumbricillus* s. lat., since Christensen (1961) reported polyploidy in only one (*L. lineatus*) of 10 species examined from this genus. We found no clear distinctions between diploid and triploid specimens of *L. lineatus* in any of our phylogenetic trees, and therefore we did not examine this factor further.

Phylogeny and morphology

The phylogeny within *Lumbricillus* s. str. (the group excluding the *L. arenarius* group and *L. semifuscus*) provides interesting insights related to the morphology of the included species. For instance, the species representing long basal branches constitutes a group of species with distinct spermathecal ampullae, whereas the species without such ampullae forms a well-supported monophyletic group (Figs 3–4, Clade B). Furthermore, the closely related *L. buelowi* and *L. knoellneri*, which were found to be sister group to the remaining *Lumbricillus* s. str., have testes which lack the fan-shaped arrangement of lobes of the seminal vesicles that are typical for most species of the genus. This suggests that unlobed seminal vesicles are the ancestral state in the genus, also supported by the fact that the closest outgroups share this condition. Thus, the characteristically lobed seminal vesicles are a synapomorphic state found in the remaining *Lumbricillus* s. str. Unfortunately, this makes defining the genus using morphological characters much harder and placement of other species not studied in this study difficult. Previously described species that lack the characteristic lobed seminal vesicles cannot simply be placed outside *Lumbricillus* s. str. as they could be closely related to *L. buelowi* and *L. knoellneri*. A solution would be to further divide the genus into a more exclusive *Lumbricillus* s. str. without *L. buelowi* and *L. knoellneri* which would instead make up another genus, supported by the monophyly of both groups. However, conclusions about such a division and about the still uncertain topology within *Lumbricillus* s. str., more specifically within the group of species without distinct ampulla of their spermathecae, could be premature as the phylogeny is likely to change when more species are added. Therefore, we do not support the further division of *L. buelowi* and *L. knoellneri* into a separate genus. The well-supported *L. arenarius* species group (containing *L. arenarius*, *L. dubius* and *L.* sp H) all have irregularly lobed seminal vesicles, lacking the fan-shaped arrangement observed in *Lumbricillus* s. str. (excluding *L. buelowi* and *L. knoellneri*). Furthermore, all three species have relatively long sperm funnels, few chaetae and *L. dubius* as well as *L.* sp H both have penial bulbs that are more or less bilobed. It should be kept in mind that only 17 out of 80 described morphospecies have been included so far, and there are probably still many undescribed species of this group to discover, especially in the non-European parts of the Holarctic, and in the Antarctic region.

Species delimitation

Our use of DNA barcoding proved to be a valuable tool for a preliminary clustering of the individuals into putative species, which facilitated species delimitation, phylogenetic inference and morphological examination. Most of the recognized species were well-supported in at least one of the statistical tests (*Rosenberg's* P_{AB} and *P(Randomly Distinct)*; both in Geneious) for COI and ITS, the exceptions being taxa from which only one individual had been sampled. Furthermore, the coalescent based delimitation provided by DISSECT found the highest support (posterior probabilities of 1) for all 24 species, including the ones with only one sampled individual. The power of this combined evidence approach can be exemplified by the genetic corroboration of the separation of the two similar 'morphospecies', *L. buelowi* and *L. knoellneri*. Furthermore, what was morphologically determined as *L. pagenstecheri* appears to be a complex of four different species, and it is possible that some of them may be matched with other nominal taxa that either have been synonymized with or regarded as closely related to *L. pagenstecheri* (see e.g. Erséus 1976; Coates & Ellis 1981). Similarly, we first had identified a large group of worms as *L. lineatus*, but they were found to represent two genetically different species, which upon further morphological examination we were able to separate into *L. lineatus* and *L. verrucosus*. Our resurrection of *L. verrucosus*, which has long been considered a synonym to *L. lineatus* (Nielsen & Christensen 1959), is based on slight differences in blood color, number of chaetae and sperm funnel length/width ratio (to be treated in more detail elsewhere). However, despite these observed differences, most other characters were found to overlap and/or vary considerably, making these two species difficult to distinguish morphologically. Interestingly, *L. verrucosus* was not supported as the closest sister of *L. lineatus* in either of the phylogenetic trees but rather as the sister of *L. rivalis* (Figs 3 and 4). Another case involves the unidentified

species *L.* sp H and *L. arenarius* where the two, although genetically supported as distinct species (Table 1), were found as sister taxa in the phylogenetic species trees (Figs 3–4). The reason why we did not refer to the two species as *L. arenarius* A and B is that we observed a morphological difference in the penial bulb, making *L.* sp H distinct from the description of *L. arenarius*. Nielsen & Christensen (1959) provided a well needed revision of the genus, but several species were then synonymized without any discussion or explanation, and only a fraction of the described species of *Lumbricillus* were treated. It is apparent that a further taxonomic revision is required for this genus and its many species.

H3 for species delimitation

Considering the observed low variation in H3 it appears that this gene is of limited use for delimiting closely related species, at least in this genus. However, the failure with sequencing H3 for some species may provide additional support for their separation from other taxa. This concerns the two pairs of taxa: *L. verrucosus* + *L. lineatus* and *L. buelowi* + *L. knoellneri*. H3 was successfully sequenced for *L. lineatus*, but out of the five attempts for *L. verrucosus* merely two sequences were obtained, and one of which from the forward primer read only. Similarly, H3 was successfully sequenced for *L. knoellneri*, but out of the five attempts for *L. buelowi* only reads from the forward primer were obtained. This suggests a genetic difference in the reverse H3 primer binding site between the two species in each pair of taxa, at least between *L. knoellneri* and *L. buelowi*.

Conclusion

In this study, we included 24 enchytraeid species by us recognized as members of *Lumbricillus* sensu Nielsen & Christensen (1959) and subsequent workers, and all of them were delimited as good species with high support. The estimated phylogenies of these species strongly support that *Lumbricillus* is a non-monophyletic taxon, at least as long as *L. semifuscus* remains in it. However, a large majority of the investigated species (including the type species, *L. lineatus*) appear to comprise a monophyletic group, which still requires its proper delimitation. *Lumbricillus semifuscus* should be excluded from the genus as it is not closely related to it, and its systematic position remains unknown. We found a discrepancy between the phylogenies as to the placement of *L. arenarius*, *L. dubius* and *L.* sp. H that together make up a monophyletic group, which depending on the kind of analysis is either sister to the remaining *Lumbricillus* species or sister to *Grania*. The species tree based on the multispecies coalescent, which we consider as the most reliable method, places this group as sister to *Lumbricillus* s. str., but additional molecular markers and sampled taxa might be needed to find the true relationship between these groups. Finally, 20 of the 24 included species have been identified, using original descriptions, and formal taxonomic revisions of them will be published elsewhere.

Acknowledgements

We would like to thank Robert Almstrand, Anna Ansebo, Nicholas Bekkouche, Mike Dempsey, Torbjørn Ekrem, Diego Fontaneto, Ton van Haaren, Karstein Hårsaker, Magnus Johansson, Sebastian Kvist, Lisa Matamoros, Erica Mejlon (=E. Sjölin), Urban Olsson, Belén Reboreda Rivera, Emilia Rota, Malin Strand, Per Sundberg, David Templeman, Hong-Zhou Wang, Pierre De Wit and Endre Willassen, for assistance in the field and/or collecting or otherwise providing material; to Anna Ansebo, Daniel Gustafsson, Per Hjelmstedt, Magnus Johansson, Emelie Lindqvist, Maria Lindström, Lisa Matamoros, Urban Olsson and Marcus Svensson for lab assistance; and to Norwegian Barcode of Life (NorBol), and to The Canadian Center of DNA Barcoding, for COI-barcoding some of our specimens. Financial support was given by The Swedish Research Council (VR), Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS), Swedish Taxonomy Initiative (ArtDatabanken, Uppsala), and Norwegian Taxonomy Initiative (Artsdatabanken, Trondheim), to CE; and by Adlerbert Foundation and Royal Society of Arts and Sciences in Göteborg, to SM and CE.

References

Achurra, A. & Erséus, C. (2013). DNA barcoding and species delimitation: the *Stylodrilus heringianus* case (Annelida: Clitellata: Lumbriculidae). *Invertebrate Systematics*, *27*, 118–128.

Aydin, Z., Marcussen, T., Ertekin, A. S. & Oxelman, B. (2014). Marginal likelihood estimate comparisons to obtain optimal species delimitations in *Silene* sect. Cryptoneurae (Caryophyllaceae). *PLoS One*, *9*, e106990.

Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, *21*, 163–193.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., Ingram, K. K. & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*, 148–155.

Birky, C. W., Fuerst, P. & Maruyama, T. (1989). Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, *121*, 613–627.

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L. & Waddell, P. J. (1993). Partitioning and combining data in phylogenetic analysis. *Systematic Biology*, *42*, 384–397.

Christensen, B. (1961). Studies on cyto-taxonomy and reproduction in the Enchytraeidae. *Hereditas*, *47*, 387–450.

Coates, K. A. (1989). Phylogeny and origins of Enchytraeidae. *Hydrobiologia*, *180*, 17–33.

Coates, K. A. (1990). Marine Enchytraeidae (Oligochaeta, Annelida) of the Albany area, Western Australia. In F. E. Wells, D. I. Walker, H. Kirkman & R. Lethbridge (Eds) *Proceedings of the Third International Marine Biological Workshop: The Marine Flora and Fauna of Albany, Western Australia* (pp. 13–41). Perth: Western Australia Museum.

Coates, K. & Ellis, D. V. (1981). Taxonomy and distribution of marine Enchytraeidae (Oligochaeta) in British Columbia. *Canadian Journal of Zoology*, *59*, 2129–2150.

Coates, K. A. & Erséus, C. (1985). Marine enchytraeids (Oligochaeta) of the coastal northwest Atlantic (northern and mid USA). *Zoologica Scripta*, *14*, 103–116.

Danforth, B. N., Lin, C. P. & Fang, J. (2005). How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? *Systematic Entomology*, *30*, 549–562.

Dasmahapatra, K. K., Elias, M., Hill, R. I., Hoffman, J. I. & Mallet, J. (2010). Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*, *10*, 264–273.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*, 879–886.

De Wit, P. & Erséus, C. (2007). Seven new species of *Grania* (Annelida: Clitellata: Enchytraeidae) from New Caledonia, South Pacific Ocean. *Zootaxa*, *1426*, 27–50.

De Wit, P. & Erséus, C. (2010). Genetic variation and phylogeny of Scandinavian species of *Grania* (Annelida: Clitellata: Enchytraeidae), with the discovery of a cryptic species. *Journal of Zoological Systematics and Evolutionary Research*, *48*, 285–293.

Degnan, J. H. & Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genetics*, *2*, 762–768.

Degnan, J. H. & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, *24*, 332–340.

Dózsa-Farkas, K. (1990). New enchytraeid species from sphagnum-bogs in Hungary (Oligochaeta: Enchytraeidae). *Acta Zoologica Hungarica*, *36*, 265–274.

Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973.

Erséus, C. (1974). *Grania pusilla* sp. n. (Oligochaeta, Enchytraeidae) from the West coasts of Norway and Sweden with some taxonomic notes on the genus *Grania*. *Sarsia*, *56*, 87–94.

Erséus, C. (1976). Littoral Oligochaeta (Annelida) from Eyjafjörôur, North Coast of Iceland. *Zoologica Scripta*, *5*, 5–11.

Erséus, C. (1977). Marine Oligochaeta from the Koster Area, west coast of Sweden, with descriptions of two new enchytraeid species. *Zoologica Scripta*, *6*, 293–298.

Erséus, C. (1994). The Oligochaeta. In J. A. Blake & B. Hilbig (Eds) *Taxonomic Atlas of the Benthic Fauna of the Santa Maria Basin and Western Santa Barbara Channel Volume 4 Oligochaeta and Polychaeta: Phyllodocida (Phyllodocidae to Paralacydoniidae)* (pp. 5–38). Santa Barbara, CA: Santa Barbara Museum of Natural History.

Erséus, C. & Gustafsson, D. R. (2009). Chapter 3. Cryptic speciation in clitellate model organisms. In D. H. Shain (Ed.) *Annelids in Modern Biology* (pp. 31–46). Hoboken, NJ: Wiley-Blackwell.

Erséus, C. & Källersjö, M. (2004). 18S rDNA phylogeny of Clitellata (Annelida). *Zoologica Scripta*, *33*, 187–196.

Erséus, C. & Lasserre, P. (1976). Taxonomic status and geographic variation of the marine enchytraeid genus *Grania* Southern (Oligochaeta). *Zoologica Scripta*, *5*, 121–132.

Erséus, C., Rota, E., Matamoros, L. & De Wit, P. (2010). Molecular phylogeny of Enchytraeidae (Annelida, Clitellata). *Molecular Phylogenetics and Evolution*, *57*, 849–858.

Floyd, R., Abebe, E., Papert, A. & Blaxter, M. (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology*, *11*, 839–850.

GIMP 2.8.10. GNU Image Manipulation Program. Available via http://www.gimp.org/.

Graefe, U. (1989). Systematische Untersuchungen an der Gattung *Achaeta* (Enchytraeidae, Clitellata) 2. Beschreibung von vier neuen Arten. *Mitteilungen aus den Hamburgischen Zoologischen Museum und Institut. Germany: Hamburg*, *86*, 127–131.

Heled, J. & Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, *27*, 570–580.

Hogner, S., Laskemoen, T., Lifjeld, J. T., Porkert, J., Kleven, O., Albayrak, T., Kabasakal, B., & Johnsen, A. (2012). Deep sympatric mitochondrial divergence without reproductive isolation in the common redstart *Phoenicurus phoenicurus*. *Ecology and Evolution*, *2*, 2974–2988.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254–267.

Issel, R. (1905). Oligocheti inferiori della fauna italica. *Zoologische Jahrbücher. Abteilung für Anatomie und Ontogenie der Tiere, Jena*, *22*, 451–476.

Jones, G., Aydin, Z. & Oxelman, B. (2015). DISSECT: an assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. *Bioinformatics*, *31*, 991–998.

Katoh, K., Misawa, K., Kuma, K. I. & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*, 3059–3066.

Kossmagk-Stephan, K. J. (1983). Marine Oligochaeta from a sandy beach of the island of Sylt (North Sea) with description of four new enchytraeid species. *Mikrofauna des Meeresbodens*, *89*, 1–28.

Kress, W. J., García-Robledo, C., Uriarte, M. & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*, 25–35.

Levinsen, G. M. R. (1883). Systematisk-geografisk Oversigt over de nordiske Annulata, Gephyrea, Chaetognathi og Balanoglossi. II. *Videnskabelige Meddelelser fra den Naturhistoriske Forening i Kjøbenhavn for Aaret*, *1883*, 92–350.

Lin, C. P. & Danforth, B. N. (2004). How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Molecular Phylogenetics and Evolution*, *30*, 686–702.

Marotta, R., Ferraguti, M., Erséus, C. & Gustavsson, L. M. (2008). Combined-data phylogenetics and character evolution of Clitellata (Annelida) using 18S rDNA and morphology. *Zoological Journal of the Linnean Society*, *154*, 1–26.

Martinsson, S. & Erséus, C. (2014). Cryptic diversity in the well-studied terrestrial worm *Cognettia sphagnetorum* (Clitellata: Enchytraeidae). *Pedobiologia*, *57*, 27–35.

Martinsson, S., Achurra, A., Svensson, M. & Erséus, C. (2013). Integrative taxonomy of the freshwater worm *Rhyacodrilus falciformis* sl (Clitellata: Naididae), with the description of a new species. *Zoologica Scripta*, *42*, 612–622.

Martinsson, S., Rhodén, C. & Erséus, C. (2015). Barcoding gap, but no support for cryptic speciation in the earthworm *Aporrectodea longa* (Clitellata: Lumbricidae). *Mitochondrial DNA*, doi: 10.3109/19401736.2015.1115487.

Masters, B. C., Fan, V. & Ross, H. A. (2011). Species delimitation–a Geneious plugin for the exploration of species boundaries. *Molecular Ecology Resources*, *11*, 154–157.

Matamoros, L., Rota, E. & Erséus, C. (2012). Cryptic diversity among the achaetous *Marionina* (Annelida, Clitellata, Enchytraeidae). *Systematics and Biodiversity*, *10*, 509–525.

Nakamura, Y. (2000). Checklist of enchytraeides (Oligochaeta: Enchytraeidae) of the world. *Miscellaneous Publication of the Tohoku National Agricultural Experiment Station (Japan)*, *24*, 29–104.

Nielsen, C. O. & Christensen, B. (1959). The enchytraeidae critical revision and taxonomy of European species. *Natura Jutlandica*, *8–9*, 1–160.

Nordborg, M. (2007). Coalescent theory. In D. J. Balding, M. Bishop & C. Cannings (Eds). *Handbook of Statistical Genetics, Third Edition* (pp. 843–877). Chichester, UK: John Wiley & Sons, Ltd.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P. & Nieves-Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, *53*, 47–67.

Ørsted, A. S. (1844). *De Regionibus Marinis* (pp. 1–89). Hauniae, Københavnː J.C. Scharling.

Pamilo, P. & Nei, M. (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution*, *5*, 568–583.

Puillandre, N., Lambert, A., Brouillet, S. & Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, *21*, 1864–1877.

Rambaut, A. (2009). FigTree v1. 4: Tree figure drawing tool. Available via http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut, A. & Drummond, A. J. (2007). Tracer v1.5. Available via http://beast.bio.ed.ac.uk/Tracer 2007.

Rannala, B. & Yang, Z. (2003). Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics*, *164*, 1645–1656.

Rodrigo, A., Bertels, F., Heled, J., Noder, R., Shearman, H. & Tsai, P. (2008). The perils of plenty: what are we going to do with all these genes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 3893–3902.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S. & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*, 539–542.

Rosenberg, N. A. (2007). Statistical tests for taxonomic distinctiveness from observations of monophyly. *Evolution*, *61*, 317–323.

Schmelz, R. M. & Collado, R. (2010a). An updated checklist of currently accepted species of Enchytraeidae (Oligochaeta, Annelida). *Newsletter on Enchytraeidae*, *12*, 67–87.

Schmelz, R. M. & Collado, R. (2010b). A guide to European terrestrial and freshwater species of Enchytraeidae (Oligochaeta). *Soil Organisms*, *82*, 1–176.

Schmelz, R. M. & Collado, R. (2015). Checklist of taxa of Enchytraeidae (Oligochaeta): an update. *Soil Organisms*, *87*, 149–152.

Siddall, M. E., Apakupakul, K., Burreson, E. M., Coates, K. A., Erséus, C., Gelder, S. R., Källersjö, M. & Trapido-Rosenthal, H. (2001). Validating Livanow: molecular data agree that leeches, branchiobdellidans, and *Acanthobdella peledina* form a monophyletic group of oligochaetes. *Molecular Phylogenetics and Evolution*, *21*, 346–351.

Southern, R. (1913). Clare Island survey part 48: oligochaeta. *Proceedings of the Royal Irish Academy*, *31*, 1–48.

Stöver, B. C. & Müller, K. F. (2010). TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, *11*, 7.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*, 2731–2739.

d'Udekem, J. (1854). Description d'une nouvelle espèce d'*Enchytraeus*. *Bulletins de L'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique. Bruxelles*, *21*, 853–864.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Ranked pairwise p-distances for *Lumbricillus* specimens, the y-axis displaying genetic distance, for (A) COI, (B) H3, and (C) ITS.

Fig. S2. 12S gene tree, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

Fig. S3. 16S gene tree, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

Fig. S4. 18S gene tree, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

Fig. S5. 28S gene tree, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

Fig. S6. H3 gene tree, estimated using Bayesian inference. Support values showing posterior probabilities. Scale bar showing expected number of changes per site.

Fig. S7. Concatenated sequence tree of 12S, 16S and COI, all mitochondrial genes, estimated using Bayesian inference.

Fig. S8. Concatenated sequence tree of 18S, 28S and ITS, all nuclear ribosomal genes, estimated using Bayesian inference.

Table S1. List of specimens used in this study, with specimen identification number, collection data, GPS coordinates (in decimal degrees), GenBank accession numbers (bold numbers are new sequences generated in this study) and voucher numbers.

Table S2. Primer sequences and PCR programs for the 7 markers used in this study.