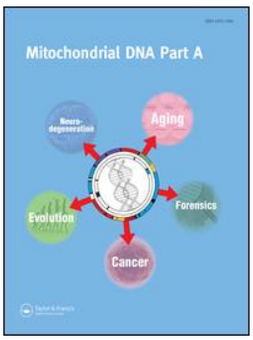

Taylor & Francis
Taylor & Francis Group

Mitochondrial DNA Part A

DNA Mapping, Sequencing, and Analysis

ISSN: 2470-1394 (Print) 2470-1408 (Online) Journal homepage: http://www.tandfonline.com/loi/imdn21

Barcoding gap, but no support for cryptic speciation in the earthworm *Aporrectodea longa* (Clitellata: Lumbricidae)

Svante Martinsson, Caroline Rhodén & Christer Erséus

To cite this article: Svante Martinsson, Caroline Rhodén & Christer Erséus (2017) Barcoding gap, but no support for cryptic speciation in the earthworm *Aporrectodea longa* (Clitellata: Lumbricidae), Mitochondrial DNA Part A, 28:2, 147-155, DOI: 10.3109/19401736.2015.1115487

To link to this article: https://doi.org/10.3109/19401736.2015.1115487

MITOCHONDRIAL DNA PART A, 2017
VOL. 28, NO. 2, 147–155
http://dx.doi.org/10.3109/19401736.2015.1115487

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

Barcoding gap, but no support for cryptic speciation in the earthworm *Aporrectodea longa* (Clitellata: Lumbricidae)

Svante Martinsson, Caroline Rhodén and Christer Erséus

Division of Systematics and Biodiversity, Department of Biological and Environmental Sciences, University of Gothenburg, Göteborg, Sweden

ABSTRACT

DNA-barcoding, using the mitochondrial marker COI, has been found successful for the identification of specimens in many animal groups, but may not be suited for species discovery and delimitation if used alone. In this study, we investigate whether two observed COI haplogroups in the earthworm *Aporrectodea longa* correspond to two cryptic species or if the variation is intraspecific. This is done by complementing *COI* with two nuclear markers, ITS2 and Histone 3. The variation is studied using distance methods, parsimony networks and Bayesian coalescent trees, and the statistical distinctness of the groups is tested on gene trees using the genealogical sorting index, Rosenberg's P_{AB} and Rodrigo et al.'s $P_{(RD)}$. We also applied multilocus species delimitation based on the multispecies coalescence model. The two haplogroups were found in *COI*, and all tests except $P_{(RD)}$ found them to be significantly distinct. However, in ITS2, the same groups were not recovered in any analyses or tests. H3 was invariable in *A. longa*, and was, therefore, included only in the multilocus analysis, which preferred a model treating *A. longa* as one species over a model splitting it into two. We also compared two measurements of size, body length, and no. of segments between the groups. No difference in body length was found, and although a significant difference in no. of segments was noted the haplogroup with the lower mean showed both the highest and the lowest value. When combined, these results led us to the conclusion that there is no support for the separation of *A. longa* into two cryptic species. This study again highlights the importance of complementing mitochondrial barcodes with more data when establishing species boundaries.

ARTICLE HISTORY
Received 1 September 2015
Revised 26 October 2015
Accepted 29 October 2015
Published online
22 December 2015

KEYWORDS
COI, cryptic species, DNA-barcoding, ITS, Oligochaeta, species delimitation

Introduction

DNA barcoding facilitates identification of organisms, by matching the sequence of a short standardized marker for identification, for animals a part of the mitochondrial gene *cytochrome oxidase subunit 1* (COI), with a reference library (Hebert et al., 2003). This approach has been found successful in many animal groups (Waugh, 2007). It has been proven good for matching specimens of different life stages and/or sexes and thereby increasing the number of individuals that can be identified to species level compared with only using morphology, where often only adults, and in cases with non-hermaphroditic animals, where often only one sex can be reliably identified (Ekrem et al., 2010; Richard et al., 2010; Stur & Ekrem, 2011). However, the usefulness of DNA barcoding depends on the existence of a good reference library to match the unknown sequences, and the importance of taking intraspecific variation into account should not be underestimated (Bergsten et al., 2012; Ekrem et al., 2007; Kvist, 2013). DNA barcoding has been shown to be beneficial for the studies of earthworms and other clitellates; it has been used to study invasive species (Martinsson et al., 2015; Porco et al., 2013), to test model organisms used in ecotoxicology (Römbke et al., 2015), and together with other data, it has been used to discover cryptic species (King et al., 2008; James et al., 2010; Martinsson & Erséus, 2014). However, when studying species boundaries and delimiting of taxa, using *COI* alone is known to often overestimate the number of species, and then more data are needed to confirm the result of DNA barcoding (Dasmahapatra et al., 2010). This has also been found to be the case for some clitellates (Achurra & Erséus, 2013; Martinsson et al., 2013). One of the most commonly used additional markers is, the whole or a part of, the ITS (internally transcribed spacer) region, consisting of ITS1, 5.8S ribosomal RNA gene (5.8S) and ITS2, and which has been found useful for species identification and delimitation of animals, as well as other organisms (Blouin, 2002; Yao et al., 2010). Due to its usefulness, ITS has been suggested as the standard barcoding region for fungi (Schoch et al., 2012), and ITS2 has been suggested as a complementary locus for the identification of animals along with the standard COI barcode (Yao et al., 2010). Another, possible additional marker is Histone 3 (H3), a more slowly evolving marker, and which has been used to study species boundaries in various animal groups (Martinsson & Erséus,

CONTACT Svante Martinsson svante.martinsson@bioenv.gu.se Division of Systematics and Biodiversity, Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, Göteborg SE-405 30, Sweden
Supplemental data for this article can be accessed http://dx.doi.org/10.3109/00365513.2015.1115487.

2014; Nygren et al., 2009; Padula et al., 2014; Rossi & Mantelatto, 2013).

During the course of a study of earthworm diversity in Scandinavia using DNA barcoding, the senior author (C. E.) found large genetic diversity within the morphospecies *Aporrectodea longa* (Ude, 1885) (Lumbricidae). This taxon comprises two COI haplogroups, with genetic distances of up to 7–8% between each other. *Aporrectodea longa* is an anecic earthworm, i.e., a species that makes deep vertical burrows, and it feeds on litter from the surface and organic material in the soil (Brown et al., 2000). The species is widespread in northern and central Europe and introduced to North America and Australasia (Sims & Gerard, 1985). At least two more mitochondrial lineages identified as *A. longa* are found in GenBank. However, at least one of them seems to consist of misidentified *A. nocturna*. If these lineages are included in *A. longa*, the species would be paraphyletic vis-a-vis at least A. *giardi*, but probably also to *A. nocturna* (Dominguez et al., 2015; Fernandez et al., 2012). Due to this, in this study, we assume that *A. longa* sensu (Ude, 1885), originally described from Germany, is the lineage that is the sister-group to *A. giardi*, i.e., the two groups that are found in Scandinavia.

The aim of this study is to test whether the two observed haplogroups represent different species, or if this is a case of deep sympatric intraspecific divergence. We will complement the COI data with two nuclear markers, the Internally Transcribed Spacer 2 (ITS2) and Histone 3 (H3). We will analyze the data sets alone using parsimony haplotype networks and Bayesian coalescent trees, on which we will statistically test the distinctness of the two haplogroups. Moreover, in a mutilocus approach, we will use the multispecies coalescence (MSC) to test if the groups belong to the same species or not. Lastly, we will test whether the specimens of the two haplogroups of *A. longa* differ in body size. If the haplogroups represent different species, we expect that the two COI clusters are well separated in both the tree and the haplotype network, and that the groups are significantly distinct. In the nuclear datasets, we expect the two groups to be at least significantly sorted, but preferably reciprocally monophyletic. We also expect the MSC analysis to prefer a model where the haplogroups are treated as two species over one that treats them as one. If we do not observe this, we will not be able to reject the null hypothesis that *A. longa* constitutes a single species.

Material and methods

Sampling

This study includes 34 specimens of *A. longa* collected in Sweden, Norway, and Denmark during 2008–2015. Collection information is listed in Table 1. The specimens were selected from a much larger set of specimens DNA barcoded during CE's study of Scandinavian earthworms. We include all available individuals of haplogroup 1, which is the rarer of the two groups, as well as a good representation of the COI diversity of haplogroup 2, in cases where both groups were found at the same locality, specimens of both groups were included. All specimens are preserved in 95% ethanol as physical vouchers. The vouchers are deposited in the Swedish museum of Natural History (SMNH), Stockholm, Sweden, and in the University Museum of Bergen (ZMBN), Bergen, Norway; see Table 1 for voucher numbers.

DNA sequencing and assembly

DNA was extracted from a small piece of the body wall taken from the posterior part of each specimen. The DNA was extracted either using Epicentre's QuickExtract DNA Extraction Solution 1.0 or Qiagen's DNeasyBlood & Tissue Kit (Qiagen, Valencia, CA). The genetic markers were amplified using PCR with the primers and programs listed in Table S1. The PCR was carried out as a 25 µl reaction. To confirm amplification, the PCR products were run on a 1% agarose gel. The PCR products were purified using 5 µl ExoTAP (Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase) (Werle et al., 1994). Sequencing was carried out by Eurofins MWG Operon (Ebersberg, Germany) or Macrogen (Geumcheon-Gu, Seoul, Korea). One specimen (CE10419) was handled by the Canadian Center for DNA Barcoding (CCDB; Guelph, Canada), with data stored at the Barcoding of Life Data Systems (BOLD), for these specimens, only the COI sequence is available. Sequences were assembled into consensus sequences using Geneious v.7.1.8 (Biomatters Ltd., Auckland, New Zealand). The sequences of each marker were aligned using MAFFT (Katoh et al., 2002) as implemented in Geneious v. 7.1.8. In the ITS2 dataset, several individuals showed clear sign of heterozygosity, i.e., showing distinct double peaks on certain positions in the chromatograms. Due to this, we separated the ITS2 alleles using the PHASE algorithm (Stephens & Donnelly, 2003; Stephens et al., 2001) as implemented in DNAsp v.5.10 (Librado & Rozas, 2009), the phasing was run for 100 iterations after 100 initial burn-in iterations, with a thinning interval of 1 using default settings. For homozygous specimens, only one of the two identical haplotypes was kept. All sequences are deposited in GenBank; see Table 1 for accession numbers.

Distance analyses

Pairwise genetic distances were calculated for the COI and ITS2 datasets (the H3 sequences were identical for all specimens; see results) in MEGA v.6.06 (Tamura et al., 2013) using uncorrected *p*-distances. The distances were analyzed using the online version of Automatic Barcode Gap Discovery, ABGD (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) (Puillandre et al., 2012) with default settings in order to delimit clusters, and visualized as ranked genetic distances (Figure 1). ABGD delimits genetic clusters by detecting a significant gap in the pairwise distance distribution. If there is a gap in the distribution, the lower distance values are presumed to reflect intraspecific difference and the higher values to reflect the interspecific differences.

Haplotype networks

Parsimony haplotype networks were constructed for the COI and ITS2 datasets using the Median-joining method (Bandelt et al., 1999) as implemented in PopART v.1 (Leigh & Bryant, 2015), with epsilon set to 0.

Table 1. List of specimens of *Aporrectodea longa* and *A. caliginosa* including in the study, identification number, COI cluster, collection sites and dates, GenBank accession nos., and museum voucher nos.

			GPS-coordinates				GenBank accession numbers			
Id. no:	COI Cluster	Collection locality	N	E	Collection date	Leg.	COI	H3	ITS	Museum voucher no.
CE4219	2	SE: Skåne, Bromölla	56.1058	14.5028	May 2008	C. Erséus	KT924098	KT924134	KT924185	SMNH149654
CE4224	2	SE: Skåne, Bromölla	56.1058	14.5028	May 2008	C. Erséus	KT924099	KT924135	KT924187	SMNH149655
CE4226	1	SE: Skåne, Bromölla	56.1058	14.5028	May 2008	C. Erséus	KT924100	KT924136	KT924186	SMNH149656
CE4890	1	SE: Västergötland, Kinnekulle	58.5583	13.3339	September 2008	C. Erséus	KT924101	KT924137	KT924188	SMNH149657
CE4891	2	SE: Västergötland, Kinnekulle	58.5583	13.3339	September 2008	C. Erséus	KT924102	KT924138	KT924189	SMNH149658
CE4951	2	DK: Midtjylland, Århus	56.2081	10.2133	October 2008	T. Cedhagen	KT924103	KT924139	KT924190/KT924191	SMNH149659
CE5124	2	SE: Bohuslän, Tjärnö	58.8869	11.1442	October 2008	C. Erséus	KT924104	KT924140	KT924192/KT924193	SMNH149660
CE5125	2	SE: Bohuslän, Tjärnö	58.8869	11.1442	October 2008	C. Erséus	KT924105	KT924141	KT924194/KT924196	SMNH149661
CE5155	1	SE: Bohuslän, Tjärnö	58.8869	11.1442	October 2008	C. Erséus	KT924112	KT924142	KT924195/KT924197	SMNH149662
CE5900	2	SE: Skåne, Vellinge	55.4797	12.9600	April 2009	C. Erséus	KT924106	KT924143	KT924198	SMNH149663
CE5982	2	SE: Västergötland, Nårunga	57.8586	12.7983	May 2009	C. Erséus	KT924107	KT924144	KT924199/KT924200	SMNH149664
CE6259	2	SE: Uppland, Åkersberga	59.4960	18.2743	June 2009	C. Erséus	KT924108	KT924145	KT924201/KT924202	SMNH149665
CE6268	2	SE: Uppland, Uppsala	59.850	17.632	June 2009	C. Erséus	KT924109	KT924146	KT924203	SMNH149666
CE6430	1	SE: Uppland, Östervåla	60.1897	17.0498	June 2009	C. Erséus	KT924110	KT924147	–	SMNH149667
CE6987	1	SE: Bohuslän, Smögen	58.3650	11.2250	August 2009	C. Erséus	KT924113	KT924148	KT924204	SMNH149668
CE9148	1	SE: Gotland, Etelhem	57.3309	18.5061	August 2010	C. Erséus	KT924114	KT924149	KT924205/KT924206	SMNH149669
CE9957	2	SE: Västergötland, Lerum	57.774	12.285	October 2010	C. Erséus	KT924111	KT924150	KT924207/KT924208	SMNH149670
CE10419	1	SE: Södermanland, Torshälla	59.4171	16.4739	October 2010	C. Erséus	JN261338	–	–	SMNH149671
CE11874	2	NO: Vest-Agder, Kristiansand	58.1622	7.9848	June 2011	C. Erséus	KT924083	KT924151	KT924209/KT924210	ZMBN106606
CE11908	2	NO: Vest-Agder, Frikstad,	58.1524	8.1395	June 2011	C. Erséus	KT924085	KT924152	KT924211	ZMBN106607
CE13319	2	NO: Östfold, Halden	59.1198	11.4020	September 2011	C. Erséus	KT924084	KT924153	KT924212	ZMBN106608
CE13532	2	SE: Västergötland, Kviberg	57.7439	12.0120	October 2011	C. Erséus	KT924093	KT924154	KT924213/KT924214	SMNH149672
CE13638	1	NO: Sør-Trøndelag, Trondheim	63.429	10.379	November 2011	E. Willassen, C. Erséus	KT924115	KT924155	KT924215/KT924216	ZMBN106609
CE14811	2	NO: Vestfold, Horten,	59.4347	10.4285	June 2012	C. Erséus	KT924086	KT924156	KT924217/KT924218	ZMBN106610
CE15977	2	SE: Västergötland, Göteborg	57.6853	11.9578	September 2012	C. Erséus	KT924094	KT924157	KT924219/KT924220	SMNH149673
CE16494	1	NO: Oslo, Oslo	59.921	10.771	October 2012	C. Erséus, S. Martinsson, Y. Liu	KT924087	KT924158	KT924221	ZMBN106611
CE16495	2	NO: Oslo, Oslo	59.921	10.771	October 2012	C. Erséus, S. Martinsson, Y. Liu	KT924088	KT924159	KT924222	ZMBN106612
CE16496	1	NO: Oslo, Oslo	59.921	10.771	October 2012	C. Erséus, S. Martinsson, Y. Liu	KT924089	KT924160	KT924223/KT924224	ZMBN106613
CE16497	2	NO: Oslo, Oslo	59.921	10.771	October 2012	C. Erséus, S. Martinsson, Y. Liu	KT924090	KT924161	KT924225/KT924226	ZMBN106614
CE16498	2	NO: Oslo, Oslo	59.921	10.771	October 2012	C. Erséus, S. Martinsson, Y. Liu	KT924091	KT924162	KT924227	ZMBN106615
CE16502	2	NO: Oslo, Oslo	59.921	10.771	October 2012	C. Erséus, S. Martinsson, Y. Liu	KT924092	KT924163	KT924228	ZMBN106616
CE19799	1	SE: Värmland, Karlstad	59.4118	13.4845	August 2013	C. Erséus	KT924096	KT924165	KT924231/KT924232	SMNH149674
CE19800	2	SE: Värmland, Karlstad	59.4118	13.4845	August 2013	C. Erséus	KT924097	KT924166	KT924233/KT924234	SMNH149692
CE16809	1	SE: Uppland, Sollentuna	59.4306	17.9875	November 2012	P. Hjelmstedt	KT924095	KT924164	KT924229/KT924230	SMNH149675
CE4972	*A. caliginosa*	SE: Bohuslän, Lysekil	58.4328	11.5800	September 2008	A. Ansebo	KT924116	KT924133	KT924184	SMNH149676

SE, Sweden; NO, Norway; DK, Denmark.

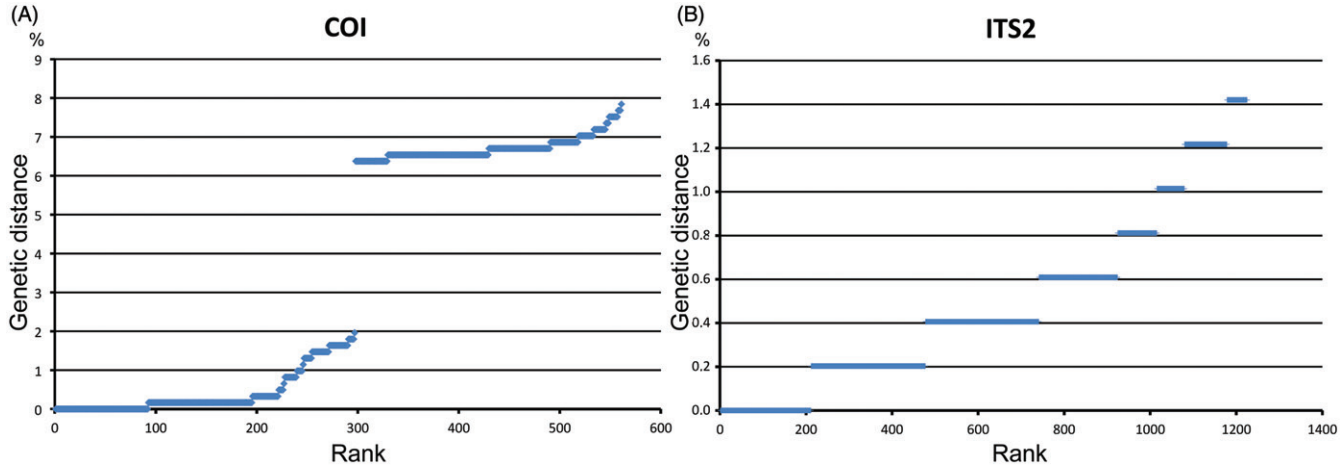

Figure 1. Ranked uncorrected genetic pairwise distances. (A) COI. (B) ITS2.

Coalescent analyses

Gene trees for COI and ITS2 were estimated under the null hypothesis that all specimens belong to a single species, using Bayesian coalescent analysis, as implemented in the BEAST package (Drummond & Rambaut, 2007; Drummond et al., 2012). Xml input files were created in BEAUTI v1.8.2 (Drummond et al., 2012), using the HKY + Γ+I model for both markers. The following settings were used for all analyses: base frequencies 'estimated' clock model 'lognormal relaxed clock (uncorrelated)'; tree prior 'coalescent/constant size'; UPGMA starting tree; constant.popsize 'lognormal: Log(Mean) = 0.0, Log(Stdev) = 1.0, offset = 0.0'. In the COI analysis, the Ucld.stdev was set to "normal" with a mean = 1.0, Stdev = 1.0. For all other priors, default settings were used. The analyses were run in BEAST v.1.8.2 (Drummond & Rambaut, 2007; Drummond et al., 2012). Analyses were run for 50 million generations, sampling every 5000th generation. Tracer v. 1.5 (Rambaut & Drummond, 2007) was used for examining the effective sample size (ESS) for parameters and determining the burn-in. Trees and posterior probabilities were summarized using TreeAnnotator v. 1.7.5 (Drummond & Rambaut, 2007) and showed on the Maximum clade credibility tree, discarding the first 10% as burn-in. The trees were drawn in FigTree v.1.3.1 (Rambaut, 2009) and further edited in Adobe Illustrator.

Molecular species delimitation

In order to test if the two haplogroups were significantly separated, the Maximum clade credibility tree for COI was imported into Geneious where the species delimitation plug-in (Masters et al., 2011) was used to calculate *P* (*Randomly Distinct*) ($P_{(RD)}$), which tests if random coalescent events could explain the observed distinctiveness of a group (Rodrigo et al., 2008), and *Rosenbergś* P_{AB}, which tests the probability for reciprocal monophyly of the clusters under random branching (Rosenberg, 2007). As groups corresponding to the two mitochondrial haplogroups were not recovered in the ITS2 tree (see Results section), the tests could not be performed on that tree. We also tested whether the specimens in the mitochondrial groups were significantly sorted, by calculating the genealogical sorting index (*gsi*) (Cummings et al., 2008), on a thinned section of trees from the posterior of the COI and ITS2 analyses in BEAST. The samples were thinned using LogCombiner v.1.8.2. (Drummond et al., 2012) sampling the trees every 500 000 generation, discarding the first 10% as burn-in, resulting in a sample of 90 trees from the posterior distribution for each gene. The *gsi* was calculated using an online web service (available at http://www.genealogicalsorting.org/). It was calculated for each haplogroup, using all 90 trees and tested by permutation, using 10 000 iterations.

Multilocus species delimitation

For this analysis, we complemented our three datasets with sequences from a specimen of *Aporrectodea caliginosa* (CE4972) to function as an out-group, as well as a positive control; unfortunately, we have no genetic data for the assumed sister-species to *A. longa*, i.e., *A. giardi*. Bayesian species delimitation was conducted using the program BPP v3.1 (Yang, 2015). The method uses the multispecies coalescent model to compare different models of species delimitation in a Bayesian framework, accounting for incomplete lineage sorting due to ancestral polymorphism and gene tree-species tree conflicts (Rannala & Yang, 2013; Yang & Rannala, 2010, 2014). We analyzed the data using a fixed species tree [((*longa1*, *longa2*), *caliginosa*))], using species delimitation algorithm 0 (Yang & Rannala, 2010, Equations (3) and (4)) with ε = 2. The population size parameters (θs) were assigned the gamma prior G (2, 100), with mean 2/100 = 0.02. The divergence time at the root of the species tree (τ0) was assigned the gamma prior G (2, 50), while the other divergence time parameters were assigned the Dirichlet prior (Yang & Rannala, 2010, Equation (2)). The analysis was run for 200 000 generation with a burn-in of 10 000 generations and a sample frequency of 5. The analysis was run three times to confirm consistency between runs.

Morphological study

Two measurements of size, number of segments, and body length (of preserved specimens) were compared between the

two groups. Only fully mature specimens with well-developed clitellum were included; therefore, only five individuals of cluster 1 and 12 individuals from cluster 2 were measured. The characters were studied on specimens preserved in 95% ethanol using a dissection microscope. The two characters were statistically analyzed using a two sample *t*-test, assuming unequal variance, performed in Microsoft Excel 2010. The result is visualized with univariate boxplots (as recommended by Weissgerber et al., 2015).

Results

DNA sequencing

For all 34 specimens, COI was successfully sequenced, for H3, 33 specimens, and for ITS2, 32 specimens were successfully sequenced (Table 1). After phasing, the ITS2 dataset consists of 50 sequences. After trimming, the COI dataset consists of 612 base pairs (bp), whereof 59 positions are variable; the H3 dataset consists of 333 bp, whereof none is variable; the ITS2 dataset consists of 493 bp, whereof eight positions are variable. As only a single haplotype was found for H3, this gene was only included in the mutilocus species delimitation analysis.

Distance analyses

For *COI*, the maximum pairwise distance within *A. longa* was 7.8%. The distances within the two groups were between 0.0 and 2.0%, and the distances between them were between 6.4 and 7.8% (Figure 1A), meaning that a barcoding gap was observed between 2.0 and 6.4% pairwise distances. The ABGD analysis for *COI* found two groups, corresponding to the two haplogroups. For ITS2, the maximum pairwise distance was 1.4%. and there was a continuous variation between 0 and 1.4% (Figure 1B), but the ABGD analysis found only a single group.

Haplotype networks

The COI haplotype network (Figure 2A) consists of 12 haplotypes, and the two haplogroups are well separated, with group 1 containing three haplotypes and group 2 containing nine haplotypes. The ITS2 haplotype network (Figure 2B) consists of 15 haplotypes, whereof six are shared between specimens of the two COI haplogroups, three are unique for group 1 and six are unique for group 2.

Bayesian coalescent analyses

In both the COI and ITS2 analyses, the effective sample size (ESS) was large for most parameters. The COI tree (Figure 3A) shows the two haplogroups as reciprocally monophyletic; however, only group 1 is well supported (pp. 0.95), whereas group 2 has low support (pp. 0.52). In the ITS2 tree (Figure 3B), the two groups suggested by COI were not found monophyletic, all supported clades include sequences from both groups, and in many cases, the two ITS2 alleles from the same individual were found in different clades as well.

Species delimitation tests

The result from the $P_{(RD)}$ species delimitation test of the two groups in the COI tree was not significant, with a p value of 0.10 for group 1 and a p value of 0.13 four group 2. However, the *Rosenbergś* P_{AB} species delimitation test gave a result that was highly significant with a value of $1.1E^{-10}$. The species delimitation tests were not performed on the ITS2 tree because the two groups suggested by COI were not found monophyletic, and the test could, therefore, not be performed on these groups. The gsi_T (the weighted combined *gsi* for all trees) for COI was 0.9931 for group 1 and 0.9375 for group 2, both groups were highly significantly sorted ($p = 1E^{-04}$ for both groups). For ITS2, the gsi_T, for group 1, was 0.126 and for group 2, it was 0.0477, none of the groups were significantly sorted ($p = 0.8121$ and $p = 0.9614$, respectively).

Multilocus species delimitation

All three runs of the BPP analysis had maximum support (PP = 1) for *A. longa* being a species separate from *A. caliginosa*. The analysis preferred the two species model, in which *A. longa* forms only one species, over the three species model in which the two haplogroups of *A. longa* are treated as two different species, with varying support (PP = 0.60–0.66) between runs.

Morphological study

The length of the specimens of COI haplogroup 1 was 11.5–14.5 cm (mean 12.74), and of those of haplogroup 2 was 10.5–14.5 cm (mean 12.67) (Figure 4A). There is no significant difference between the two groups ($p = 0.9159$). The number of segments varied between 193 and 208 (mean 201.2) in haplogroup 1 and between 152 and 214 (mean 182.2) in haplogroup 2 (Figure 4B). The difference between the two groups is significant ($p = 0.0044$)

Discussion

In this study, we have found that the two mt-haplogroups within *Aporrectodea longa* are well separated in COI, with the existence of a barcoding-gap of 4.4% uncorrected *p*-distance. Group 1 is well supported in the gene-tree, and group 2 is recovered but with low support. In the statistical tests performed, we found that the two groups are significantly sorted (*gsi*) and that their reciprocal monophyly is not due to random branching (*Rosenberg's* P_{AB}), however, we cannot reject random coalescent as the cause for the observed distinctness of the two groups ($P_{(RD)}$). Finding support for the existence of two groups in COI is not surprising as it was the observation of this separation that initiated this study. The barcoding-gap seen in COI is more than twice the variation within the clusters, however, it is still small compared with what is often found between species of clitellates (Gustafsson et al., 2009; James et al., 2010; Kvist et al., 2010; Martinsson & Erséus, 2014; Matamoros et al., 2012). In ITS2, the two groups were not found separated by any of the methods and tests used, and we have no support for separation or sorting in this nuclear marker. The other nuclear marker used in this study, H3, showed no

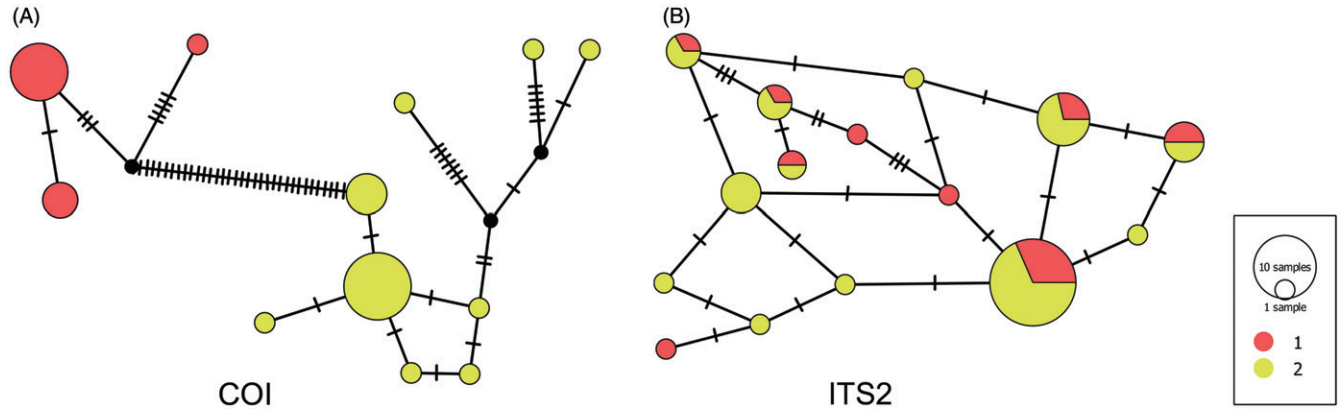

Figure 2. Median-joining parsimony haplotype networks. (A) COI. (B) ITS2. Color corresponds to COI haplogroup, hatch marks show number of mutations.

variation at all and could, therefore, not be analyzed alone. In the Multilocus species delimitation analysis (BPP), we found that the model treating *A. longa* as a single species was preferred over one splitting the two haplogroups into two species. Based on the results mentioned above, we conclude that we have no genetic support for proposing a split of *A. longa* into two species.

The difference between the markers could be due to the faster mutation rate and a faster lineage sorting of the mitochondrial genome compared with the nuclear one (Brown et al., 1979; Neigel & Avise, 1986). However, we did not see any sorting in the ITS2 trees, and if the two groups suggested by COI were the result of a recent speciation event we would still expect to see some degree of sorting in ITS2 even if not complete. The two statistical tests performed to test if chance can be rejected as the cause of the observed tree-structure, $P_{(RD)}$ and P_{AB} gave conflicting result, P_{AB} gave high significance for that the two group's reciprocal monophyly are not caused by random branching, whereas $P_{(RD)}$ did not reject the possibility that the distinctness of the two groups is due to random coalescence. As noted in at least one other study on clitellates (Martinsson et al., 2013), the *Rosenberg's* P_{AB} statistic seems to be more liberal than $P_{(RD)}$.

We find a statistical difference in one of the two measurements of size used, the number of segments, group 1 having a higher mean than group 2, but both the highest and lowest values are found in group 2, which thus is more variable then group 1. As the sample sizes are small and the measurements overlap, it is hard to know if the statistical difference is biologically significant. In another earthworm species, *Octolasion tyrtaeum* (Savigny, 1826), two morphs differing in size, was also found to differ in mt-DNA (Heethoff et al., 2004). However, a more recent study has found that the mt-DNA clades in this taxon are not always correlated with size (Shekhovtsov et al., 2014), and it is possible that this would be the case also for *A. longa* if a larger sample from a broader geographical area were analyzed. In another case, statistical differences in body size was found between *Lumbricus terrestris* and *L. herculeus*, that forms a cryptic species pair, but in that case, the genetic differences are much larger, with a mean COI distance between them of over 17% (James et al., 2010), than in our case.

Specimens from both COI haplogroups co-occur at five localities (Table 1) and in three of these cases, specimens from group 1 and group 2 do share ITS2 haplotypes. This can be an indication that they are part of the same local population, and that the whole Scandinavian metapopulation of *A. longa* is panmictic, i.e., all individuals reproduce freely with each other.

There are at least three possible explanations for the observed genetic pattern. It can be explained as (1) a case of despeciation (Turner, 2002), where two separated, divergent lineages after secondary contact start to interbreed to such extent that the species boundaries between them are broken up and they form a single metapopulation. This has been suggested as explanation for sympatric, divergent, interbreeding lineages within bird species (Hogner et al., 2012; Webb et al., 2011), and it has also been suggested as an explanation for mt-divergence in some clitellates (De Wit & Erséus, 2010; Martinsson et al., 2013). (2) A second explanation would be introgression, where due to limited hybridization, a part of one population's genome gets incorporated in another distinct population. Introgression has been shown to explain mt-paraphyly in several animal groups, including, e.g., ducks and chipmunks (Good et al., 2008; Peters et al., 2007). In our case, a distinct mitochondrial lineage originated elsewhere could have been introgressed into the Scandinavian population. About 5% of the Scandinavian population of *A. longa* are from COI haplogroup 1, while the large majority is from haplogroup 2 (CE unpublished data), and it therefore seems reasonable to assume that haplogroup 1 has introgressed into haplogroup 2. Searches in GenBank and BOLD also give much fewer records of haplogroup 1 than of haplogroup 2, and it is not possible to speculate on the origin of this group. According to searches in GenBank and BOLD, group 2 seems to be widespread in Western Europe with records from Sweden, Norway, Denmark, Germany, United Kingdom, and France and also introduced to Canada, whereas group 1 is so far found in Scandinavia and Great Britain, and also introduced to Canada. (3) Finally, the great COI divergence in NW European *A. longa* may be due to ancient mitochondrial polymorphism, where the two haplogroups have arisen and remained within the species, whereas intermediate haplogroups have disappeared due to random genetic drift (Kimura, 1955). The Pleistocene glaciations probably caused massive extinction among European earthworms,

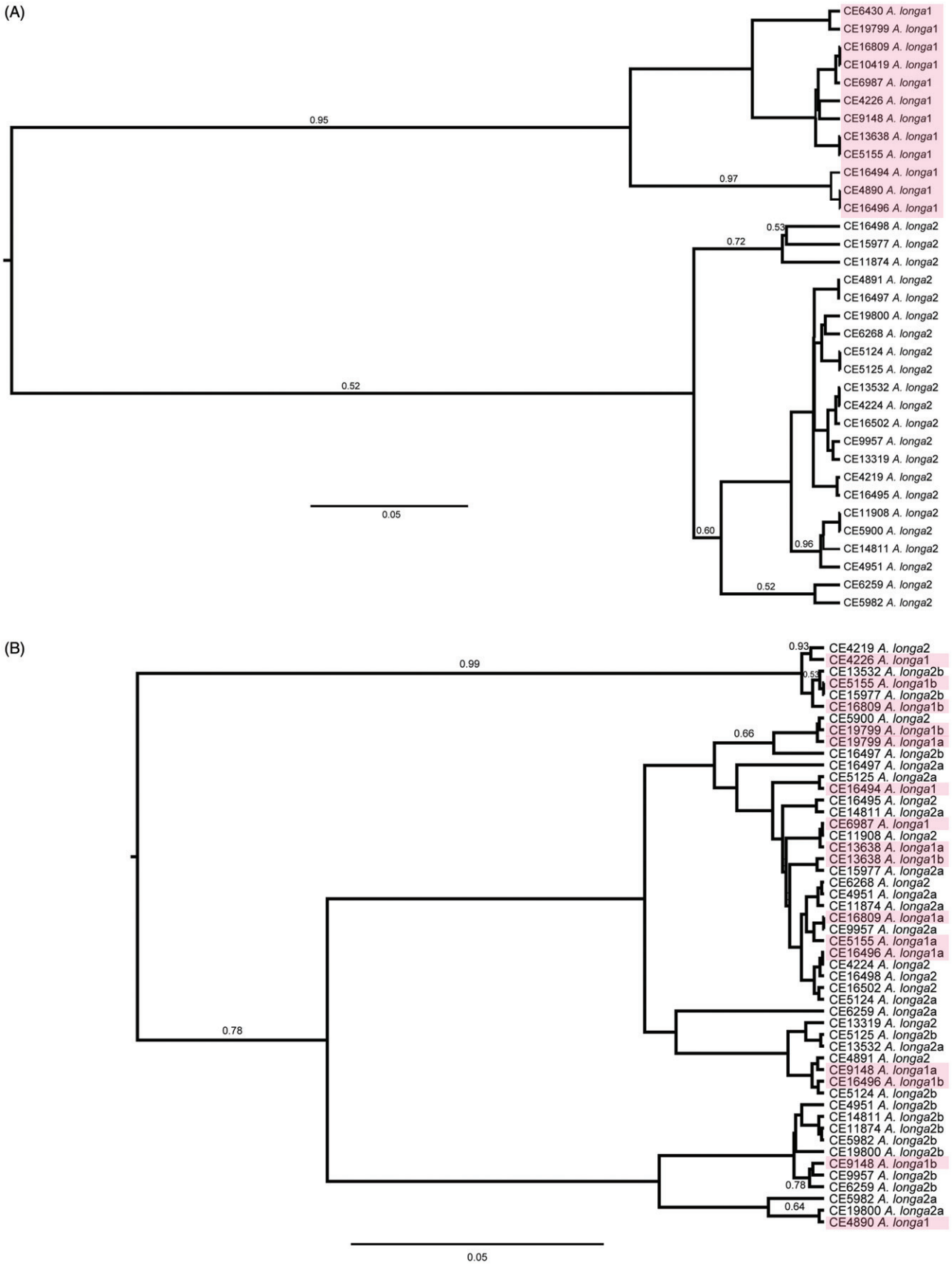

Figure 3. Bayesian coalescent gene trees. (A) COI. (B) ITS2. Numbers at branches are posterior probabilities (pp); only values above 0.5 are shown. The first number in the terminal taxa are specimens ID nos., the last number corresponds to the COI haplogroup, and the 'a' or 'b' last in some terminal taxa indicate different, phased, haplotypes from the same specimen. Scale bar shows expected number of substitutions per site.

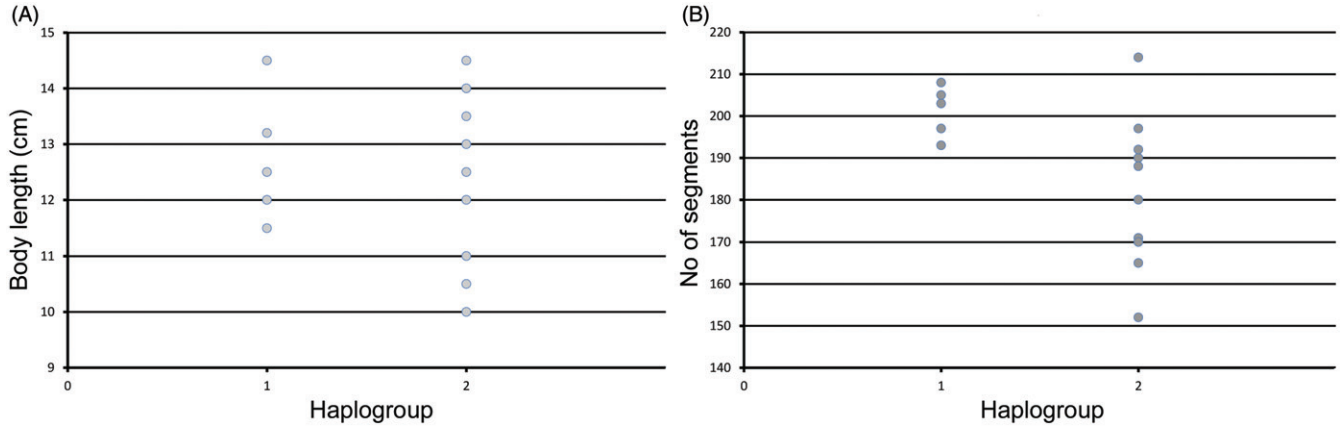

Figure 4. Body size. (A) Body length. (B) Number of segments.

and the surviving species went through bottlenecks, reducing the genetic variation within them (Amos & Harwood, 1998).

Conclusions

This study again highlights the problems with using DNA-barcoding alone for species discovery and delimitation. If we would have used only COI data and analyzed them with distance methods, it is likely that we would have concluded that *A. longa* is in fact two species. However, by incorporating additional data and analyses, we found no support for cryptic speciation within this species. Previous studies on clitellates have also questioned the uncritical use of COI alone for species recognition (Achurra & Erséus, 2013; Martinsson et al., 2013). Nevertheless, it is still important to point out that when a good reference library with well-delimited species exists, DNA-barcoding is a useful tool for specimen identification, not the least with in Clitellata, where cryptic species are common (Erséus & Gustafsson, 2009). However, for species delimitation as such, we strongly recommend that mitochondrial markers are always used in combination with other data sources.

Acknowledgements

The authors thank Tomas Cedhagen, Endre Willassen, Yingkui Liu, Per Hjelmstedt, and Anna Ansebo for assistance with fieldwork, and Anna Ansebo, Emelie Lindquist, Marcus Svensson, Per Hjelmstedt, and Yingkui Liu for help in the lab. The authors also thank Bernhard Pfeil, who provided useful advice and comments.

Declaration of interest

The authors report that they have no conflict of interest. The authors alone are responsible for the content and writing of the paper. The research received support from the Swedish Taxonomy Initiative and the Norwegian Taxonomy Initiative to C. E., and from the Adelbertska research foundation to S. M. and C. E.

References

Achurra A, Erséus C. (2013). DNA barcoding and species delimitation: The *Stylodrilus heringianus* case (Annelida: Clitellata: Lumbriculidae). Invertebr Syst 27:118–28.

Amos W, Harwood J. (1998). Factors affecting levels of genetic diversity in natural populations. Philos Trans R Soc Lond, B, Biol Sci 353:177–86.

Bandelt HJ, Forster P, Röhl A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, et al. (2012). The effect of geographical scale of sampling on DNA barcoding. Syst Biol 61:851–69.

Blouin MS. (2002). Molecular prospecting for cryptic species of nematodes: Mitochondrial DNA versus internal transcribed spacer. Int J Parasitol 32:527–31.

Brown GG, Barois I, Lavelle P. (2000). Regulation of soil organic matter dynamics and microbial activity in the drilosphere and the role of interactions with other edaphic functional domains. Eur J Soil Biol 36:177–98.

Brown WM, George M, Wilson Jr AC. (1979). Rapid evolution of animal mitochondrial DNA. PNAS 76:1967–71.

Cummings MP, Neel MC, Shaw KL. (2008). A genealogical approach to quantifying lineage divergence. Evolution 62:2411–22.

Dasmahapatra KK, Elias M, Hill RI, Hoffman JI, Mallet J. (2010). Mitochondrial DNA barcoding detects some species that are real, and some that are not. Mol Ecol Resour 10:264–73.

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. J Zool Syst Evol Res 48:285–93.

Dominguez J, Aira M, Breinholt JW, Stojanovic M, James SW, Perez-Losada M. (2015). Underground evolution: New roots for the old tree of lumbricid earthworms. Mol Phylogenet Evol 83:7–19.

Drummond AJ, Rambaut A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.

Drummond AJ, Suchard MA, Xie D, Rambaut A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–73.

Ekrem T, Stur E, Hebert PDN. (2010). Females do count: Documenting Chironomidae (Diptera) species diversity using DNA barcoding. Org Divers Evol 10:397–408.

Ekrem T, Willassen E, Stur E. (2007). A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol Phylogenet Evol 43:530–42.

Erséus C, Gustafsson D. (2009). Cryptic speciation in clitellate model organisms. In: Shain DH, editor. Annelids in modern biology. Hoboken, NJ: John Wiley & Sons, Inc.

Fernandez R, Almodovar A, Novo M, Simancas B, Diazcosin DJ. (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 E 64:368–79.

Good JM, Hird S, Reid N, Demboski JR, Steppan SJ, Martin-Nims TR, Sullivan J. (2008). Ancient hybridization and mitochondrial capture between two species of chipmunks. Mol Ecol 17:1313–27.

Gustafsson DR, Price DA, Erséus C. (2009). Genetic variation in the popular lab worm *Lumbriculus variegatus* (Annelida: Clitellata: Lumbriculidae) reveals cryptic speciation. Mol Phylogenet Evol 51:182–9.

Hebert PD, Cywinska A, Ball SL, Dewaard JR. (2003). Biological identifications through DNA barcodes. Proc Biol Sci 270:313–21.

Heethoff M, Etzold K, Scheu S. (2004). Mitochondrial COII sequences indicate that the parthenogenetic earthworm *Octolasion tyrtaeum* (Savigny 1826) constitutes of two lineages differing in body size and genotype. Pedobiologia 48:9–13.

Hogner S, Laskemoen T, Lifjeld JT, Porkert J, Kleven O, Albayrak T, Kabasakal B, Johnsen A. (2012). Deep sympatric mitochondrial divergence without reproductive isolation in the common redstart *Phoenicurus phoenicurus*. Ecol Evol 2:2974–88.

James SW, Porco D, Decaens 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.

Katoh K, Misawa K, Kuma K, Miyata T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–66.

Kimura M. (1955). Random genetic drift in multi-allelic locus. Evolution 9:419–35.

King RA, Tibble AL, Symondson WO. (2008). Opening a can of worms: Unprecedented sympatric cryptic diversity within British lumbricid earthworms. Mol Ecol 17:4684–98.

Kvist S. (2013). Barcoding in the dark? A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Mol Phylogenet Evol 69:39–45.

Kvist S, Sarkar IN, Erséus C. (2010). Genetic variation and phylogeny of the cosmopolitan marine genus *Tubificoides* (Annelida: Clitellata: Naididae: Tubificinae). Mol Phylogenet Evol 57:687–702.

Leigh JW, Bryant D. (2015). Popart: Full-feature software for haplotype network construction. Method Ecol Evol 6:1110–16.

Librado P, Rozas J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–2.

Martinsson S, Achurra A, Svensson M, Erséus C. (2013). Integrative taxonomy of the freshwater worm *Rhyacodrilus falciformis* s.l. (Clitellata: Naididae), with the description of a new species. Zool Scripta 42:612–22.

Martinsson S, Cui Y, Martin PJ, Pinder A, Quinlan K, Wetzel MJ, Erséus C. (2015). DNA-barcoding of invasive European earthworms (Clitellata: Lumbricidae) in south-western Australia. Biol Invasions 17:2527–32.

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

Masters BC, Fan V, Ross HA. (2011). Species delimitation – A geneious plugin for the exploration of species boundaries. Mol Ecol Resour 11:154–7.

Matamoros L, Rota E, Erséus C. (2012). Cryptic diversity among the achaetous *Marionina* (Annelida, Clitellata, Enchytraeidae). Syst Biodivers 10:509–25.

Neigel JE, Avise JC. (1986). Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Nevo E, Karlin S, editors. Evolutionary processes and theory. New York: Academic Press.

Nygren A, Eklöf J, Pleijel F. (2009). Arctic-boreal sibling species of *Paranaitis* (Polychaeta, Phyllodocidae). Mar Biol Res 5:315–27.

Padula V, Araújo AK, Matthews-Cascon H, Schrödl M. (2014). Is the Mediterranean nudibranch *Cratena peregrina* (Gmelin, 1791) present on the Brazilian coast? Integrative species delimitation and description of *Cratena minor* n. sp. J Mollus Stud 80:575–84.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE. (2007). Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between Gadwall and Falcated Duck (*Anas spp.*). Evolution 61:1992–2006.

Porco D, Decaëns T, Deharveng L, James S, Skarżyński D, Erséus C, Butt K, et al. (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 Invasions 15:899–910.

Puillandre N, Lambert A, Brouillet S, Achaz G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. Mol Ecol 21:1864–77.

Rambaut A. (2009). FigTreev1.3.1. Available at: http://tree.bio.ed.ac.uk/software/figtree/ (last accessed August 21, 2012).

Rambaut A, Drummond AJ. (2007). Tracerv1.5. Available at: http://beast.bio.ed.ac.uk/Tracer (last accessed April 05, 2013).

Rannala B, Yang Z. (2013). Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194:245–53.

Richard B, Decaens T, Rougerie R, James SW, Porco D, Hebert PD. (2010). Re-integrating earthworm juveniles into soil biodiversity studies: Species identification through DNA barcoding. Mol Ecol Resour 10:606–14.

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? Philos Trans R Soc Lond, B, Biol Sci 363:3893–902.

Rosenberg NA. (2007). Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution 61:317–23.

Rossi N, Mantelatto FL. (2013). Molecular analysis of the freshwater prawn *Macrobrachium olfersii* (Decapoda, Palaemonidae) supports the existence of a single species throughout its distribution. PLoS One 8:e54698.

Römbke J, Aira M, Backeljau T, Breugelmans K, Domínguez J, Funke E, Graf N, et al. (2015). DNA barcoding of earthworms (*Eisenia fetida/andrei* complex) from 28 ecotoxicological test laboratories. Appl Soil Ecol in press.

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, Fungal barcoding Consortium. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 109:6241–6.

Shekhovtsov SV, Golovanova EV, Peltek SE. (2014). Genetic diversity of the earthworm *Octolasion tyrtaeum* (Lumbricidae, Annelida). Pedobiologia 57:245–50.

Sims RW, Gerard BM. (1985). Earthworms: Keys and notes for the identification and study of the species. London: Brill.

Stephens M, Donnelly P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–9.

Stephens M, Smith NJ, Donnelly P. (2001). A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–89.

Stur E, Ekrem T. (2011). Exploring unknown life stages of *Arctic Tanytarsini* (Diptera: Chironomidae) with DNA barcoding. Zootaxa 2743:27–39.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–9.

Turner GF. (2002). Parallel speciation, despeciation and respeciation: Implications for species definition. Fish Fish 3:225–9.

Ude H. (1885). Über die Rückenporen der terricolen Oligochaeten, nebst Beiträgen zur Histologie des Leibesschlauches und zur Systematik der Lumbriciden. Z Wiss Zool 43:87–143.

Waugh J. (2007). DNA barcoding in animal species: Progress, potential and pitfalls. Bioessays 29:188–97.

Webb WC, Marzluff JM, Omland KE. (2011). Random interbreeding between cryptic lineages of the Common Raven: Evidence for speciation in reverse. Mol Ecol 20:2390–402.

Weissgerber TL, Milic NM, Winham SJ, Garovic VD. (2015). Beyond bar and line graphs: Time for a new data presentation paradigm. PLoS Biol 13:e1002128.

Werle E, Schneider C, Renner M, Volker M, Fiehn W. (1994). Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Res 22:4354–5.

Yang Z. (2015). A tutorial of BPP for species tree estimation and species delimitation. Curr Zool 61:854–65.

Yang Z, Rannala B. (2010). Bayesian species delimitation using multilocus sequence data. Proc Natl Acad Sci USA 107:9264–9.

Yang Z, Rannala B. (2014). Unguided species delimitation using DNA sequence data from multiple Loci. Mol Biol Evol 31:3125–35.

Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang Y, et al. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS One 5: e13102