Molecular Phylogenetics and Evolution 106 (2017) 18–27

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Cryptic speciation and limited hybridization within *Lumbricus* earthworms (Clitellata: Lumbricidae)

Svante Martinsson *, Christer Erséus

Department of Biological & Environmental Sciences, University of Gothenburg, Box 463, SE-405 30 Göteborg, Sweden

ARTICLE INFO

Article history:
Received 8 July 2016
Revised 6 September 2016
Accepted 14 September 2016
Available online 14 September 2016

Keywords:
DNA-barcoding
Earthworms
Species delimitation
Cryptic species
Annelida

ABSTRACT

Cryptic mitochondrial (mt) lineages are known to exist in the earthworm morphospecies *Lumbricus rubellus* and *L. terrestris*. The latter was recently split into two species, *L. terrestris* and *L. herculeus*, based on large genetic distances and a statistical difference in body size. There is support for the separation of some lineages in *L. rubellus* into species, whereas other lineages, separated by similar mt genetic distances, have been found to be part of the same species. However, no study has evaluated the status of the cryptic mt lineages in *L. terrestris-L. herculeus* and *L. rubellus* using nuclear genes. We use a combination of methods to reveal extensive cryptic speciation and limited hybridization in *Lumbricus*, based on one nuclear (H3) and one mitochondrial (COI) marker. Using a Bayesian multi-locus species delimitation method, as well as single gene haplotype networks and gene trees, we delimit seven well supported cryptic species within the morphospecies *L. rubellus*, and confirm the split within the species-pair *L. terrestris-L. herculeus*. Limited hybridization was found between the most common species of *L. rubellus* (A) in northern Europe and two other species (B and H) in this complex, as well as between *L. terrestris* and *L. herculeus*. Deep mt divergence was found within *L. terrestris s.str.* but no support for further splitting of this taxon was found. Both *L. rubellus* and *L. terrestris* are well studied model organisms, and considering that cryptic species and hybridization were found within them, it is important that they are properly identified in future studies.

1. Introduction

Cryptic species, i.e. species that are morphologically so similar that they have been classified under the same nominal species, are common in many organism groups (e.g., Bickford et al., 2007; Pfenninger and Schwenk, 2007), including annelid worms (Erséus and Gustafsson, 2009; Nygren, 2014). Some of the most well-studied annelids are earthworms (Crassiclitellata), and cryptic lineages have been discovered in several earthworm families. The best studied family is Lumbricidae, with a mainly Palearctic distribution, and cryptic lineages have been reported in numerous lumbricid morphospecies (e.g., Fernández et al., 2011; Heethoff et al., 2004; King et al., 2008; Shekhovtsov et al., 2016, 2013). Cryptic lineages have also been found in the Mediterranean family Hormogastridae (e.g., Novo et al., 2012, 2010), in morphospecies of Megascolecidae from New Zealand (Buckley et al., 2011) and Taiwan (Chang et al., 2008), and in Glossoscolecidae from Brazil (Siqueira et al., 2013). This list of where cryptic lineages are found within Crassiclitellata should not been seen as complete, and the distribution of cryptic lineages probably reflects research intensity more than anything else.

The discovery of cryptic species was accelerated with the introduction of DNA-barcoding (Hebert et al., 2003), where a short standard marker is used to identify specimens to species. For animals, the standard barcoding marker is a fragment of the mitochondrial (mt) gene *Cytochrome Oxidase B subunit I* (COI). However, COI alone should not be used to delimit species, as deep mt divergence is found within many biological species (e.g., Achurra and Erséus, 2013; Dasmahapatra et al., 2010; Hogner et al., 2012; Martinsson et al., 2015; Webb et al., 2011), which is likely to result in oversplitting. It is also problematic to use COI alone for identification of animals, if hybridization is possible among the species studied, as it will identify the hybrids as belonging to the same species as their mother. It is estimated that at least 10% of all animal species can hybridize with at least one other species (Mallet, 2005). To correctly delimit species, DNA-barcoding must be supplemented by additional data, such as nuclear markers and morphology.

The development of multi species coalescent methods is not only beneficial for phylogeny estimations and species delimitations (Degnan and Rosenberg, 2009; Fujita et al., 2012), it also

* Corresponding author.
E-mail addresses: svante.martinsson@bioenv.gu.se (S. Martinsson), christer.erseus@bioenv.gu.se (C. Erséus).

http://dx.doi.org/10.1016/j.ympev.2016.09.011
1055-7903/

makes it possible to explore the causes for gene-tree conflicts. Setting aside errors in the phylogenetic estimation, the main causes are incomplete lineage sorting (ILS) and hybridization, but also paralogy will cause problems, if not detected. ILS occurs when an ancestral species, which has more than one allele lineage at a given locus, divides into two species. Both alleles can be kept in both of the new species, and when one of the species divides again the alleles may be fixed in different species. Depending on how the alleles become fixed in the species, the gene tree for a locus may differ from the species tree, this is incomplete lineage sorting (ILS) (Degnan and Rosenberg, 2009; Rogers and Gibbs, 2014). ILS is most prevalent when the time between speciation is short and the population sizes are large (Degnan and Rosenberg, 2009). Hybridization will also cause gene tree conflicts, but in this case the gene flow occur after the speciation event, which will lead to shorter genetic distances between taxa affected by hybridization, compared to if no hybridization was present. This difference between hybridization and ILS can be used to separate the two as causes of gene tree conflicts (Joly et al., 2009).

In the earthworm genus *Lumbricus* L., 1758 (Clitellata: Lumbricidae), cryptic mt lineages have been found in the two well-known morphospecies, *L. terrestris* L., 1758, (James et al., 2010), and *L. rubellus* Hoffmeister, 1843, (e.g., Donnelly et al., 2013; King et al., 2008). In *L. terrestris* two lineages were found and separated into two nominal species, *L. terrestris s.str.* and *L. herculeus* (Savigny, 1826), based on large genetic distances in COI and statistically significant differences in body size (James et al., 2010). However, it is known for other earthworm species that mt lineages can differ in body size and still be part of the same species (Martinsson et al., 2015; Shekhovtsov et al., 2014), and no study using nuclear data to verify the separation of *L. terrestris* and *L. herculeus* has been published. In *L. rubellus*, several well separated mt lineages have been found (e.g., Sechi, 2013). There is support from microsatellite (Donnelly et al., 2013), morphological (Donnelly et al., 2014) as well as behavioral (Jones et al., 2016) data that two *rubellus* lineages found in Great Britain are separate species. On the other hand, a recent study using RADseq data found no support for three Polish lineages of *L. rubellus* being reproductively isolated, although the mt genetic distances were similar to those found between the British lineages (Giska et al., 2015). However, it should be noted that the British and Polish worms studied had only one lineage in common. According to these previous studies there is still no nuclear support for the split between *L. terrestris* and *L. herculeus*, but there is some support (of varying kinds) for the existence of cryptic species within *L. rubellus*.

The lineages of *L. rubellus* have to some extent different geographic distributions. Lineages A and B (*sensu* Sechi, 2013) are the most common lineages found in Great Britain and Scandinavia (Sechi, 2013; C. Erséus, unpubl. data), and A is widespread in Europe. Lineage C and D are found in Central Europe; lineage E is widespread across Europe, and F is restricted to the Iberian peninsula, whereas G and H are found in Central Europe (Donnelly et al., 2014; Giska et al., 2015; Sechi, 2013). In the *L. herculeus-L. terrestris* complex, *L. terrestris* has a more northern distribution, and is found in most of Scandinavia, whereas *L. herculeus* is a more southern species found in continental Europe, and only in the southern most parts of Scandinavia (James et al., 2010). *Lumbricus rubellus* A (as *L. rubellus* L2) and *L. terrestris* are also found in N. America (Porco et al., 2013), and in this paper we also report *L. rubellus* A from New Zealand, and *L. rubellus* H2 and M from USA (present study).

The aim of this study is to test the nuclear support (using the gene Histone 3) for cryptic lineages found in morphospecies of *Lumbricus* by COI DNA-barcoding. We do this by using both single gene trees and networks, as well as multi-locus methods. Finding nuclear support for these lineages will strengthen the hypothesis that they are reproductively isolated, and therefore should be treated as different species. We also examine and test for hybridization between the cryptic species delimited.

2. Material and methods

2.1. Specimens, DNA sequencing and assembly

In total 122 specimens representing five nominal *Lumbricus* species, viz. *L. castaneus* (Savigny, 1826), *L. festivus* (Savigny, 1826), *L. herculeus*, *L. rubellus* and *L. terrestris* were included in the study. The specimens were mainly from Norway and Sweden (Scandinavia), but also from local populations in the Czech Republic, Denmark, Germany, New Zealand and USA (Table S1). The Scandinavian specimens were selected from a much larger set of DNA-barcoded specimens (CE unpublished data), and specimens were selected to maximize the genetic variation among the samples/species, and some of the COI sequences for *L. terrestris* and *L. herculeus* are from James et al. (2010). For tree estimations, three out-group taxa, viz. *Aporrectodea caliginosa* (Savigny, 1826), *A. longa* (Ude, 1885) and *Dendrodrilus rubidus* (Savigny, 1826), were used. The sequences from the *Aporrectodea* species are from Martinsson et al. (2015).

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. The two genetic markers *Cytochrome Oxidase B subunit I* (COI) and *Histone 3* (H3) were amplified using the primers and programs described by Martinsson et al. (2015). The PCR-reactions were carried out as 25 µl reactions. To confirm amplification, the PCR products were run on a 1% agarose gel, and purified using 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). For mismatched specimens (see Sections 3.2 and 3.3) both COI and H3 were re-amplified and sequenced again to exclude possible mix-up during original amplification and sequencing.

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 the Geneious alignment with default settings in Geneious v. 7.1.8. In the H3 dataset several individuals showed clear sign of heterozygosity, i.e., showing distinct double peaks at certain positions in the chromatograms. Due to this we separated the H3 alleles using the PHASE algorithm (Stephens and Donnelly, 2003; Stephens et al., 2001) as implemented in DNAsp v.5.10 (Librado and 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 alleles was kept. All sequences are deposited in GenBank; see Table S1 for accession numbers.

2.2. Distance analysis and preliminary division of specimens into clusters

To divide the specimens into barcoding clusters, the uncorrected genetic p-distances were calculated for the COI dataset in MEGA 6 (Tamura et al., 2013). The specimens were divided based on the existence of a barcoding-gap, i.e., that the COI distances within the group are clearly smaller than the distances between the group and the closest other group. *Lumbricus rubellus* was divided into 8 mt lineages, *L. terrestris* into 2, and *L. castaneus*, *L. festivus* and *L. herculeus* formed one group each; in total 13 lineages. In order to ensure consistent naming of cryptic lineages, the lineages of *L. rubellus* were compared with a small dataset

obtained by researchers at the University of Cardiff (GenBank nos. KP642290-KP642109); this dataset was also used by Giska et al. (2015) for naming the lineages included in their study. Previously known COI lineages were named with letters based on this dataset, and new lineages were denoted with consecutive letters, whereas the lineages of *L. terrestris* were denoted with numbers (1–2). The mt lineages referred to in the present study are summarized in Table 1.

2.3. Haplotype networks

To visualize haplotype diversity, haplotype networks were constructed for both genes in PopART v1 (Leigh and Bryant, 2015) using statistical parsimony (Clement et al., 2002; Templeton et al., 1992). For the COI network, a trimmed alignment of 588 bp was used to ensure that the alignment had no missing data. For the H3 network, the full length alignment could be used.

2.4. Gene tree estimations

Individual gene trees were estimated using Bayesian Inference in MrBayes v3.2.6 (Ronquist et al., 2012). Model testing was conducted in MEGA 6.06 (Tamura et al., 2013): for the COI analysis the GTR + Γ substitution model was selected, for H3 the HKY + Γ model. The analyses were conducted in MrBayes with 4 MC^3, running for 20 million generations and sampling every 20,000th generation. An initial 'burn-in phase' of 25% was discarded. Resulting p-files were examined in Tracer v1.6 (Rambaut and Drummond, 2007) to evaluate convergence and to ensure sufficient burn-in for the trees.

2.5. Molecular species delimitation

To delimit species, a joint Bayesian multi-locus species delimitation and species tree estimation was conducted using the program BPP v3.1. The method uses the multispecies coalescent model to compare different models of species delimitation and species phylogeny in a Bayesian framework, accounting for incomplete lineage sorting due to ancestral polymorphism and gene tree-species tree conflicts (Rannala and Yang, 2013; Yang and Rannala, 2010, 2014). As the priors for the population size (θs) and the divergence time at the root (τ0) are known to have a considerable effect on the result of the analysis, and in particular so for θs (Leache and Fujita, 2010; Rannala, 2015), two analyses with different values assigned to these priors were performed. In both of them, the 13 mt lineages, divided based in the existence of a barcoding-gap (see Section 2.2) were used as the input species to be tested in the analyses. In analysis A the population size parameter (θs) was assigned the gamma prior G(2,100), with mean 2/100 = 0.02, and the divergence time at the root of the species tree (τ0) was assigned the gamma prior G(2,20). Analysis B had the population size parameters (θs) assigned the gamma prior G (2,400), with mean 2/400 = 0.005, and the divergence time at the root of the species tree (τ0) was assigned the gamma prior G (2,40). In both analyses the other divergence time parameters were assigned the Dirichlet prior (Yang and Rannala, 2010, equation 2). The analyses were both run three times to confirm consistency between runs. Lineages delimited with a posterior probability of >0.95 in all analyses are considered as well supported.

2.6. Species tree estimation and testing for incomplete lineage sorting vs. hybridization

To investigate causes of mitochondrial-nuclear discordance in the placement of four mismatching individuals, which were found in different lineages depending on the marker (see Sections 3.2 and 3.3), we used the statistical, posterior predictive checking method, based on genetic distances, developed by Joly et al. (2009). The method compares the pairwise genetic distances to gene trees simulated in species trees, to test the probability that the distances observed are caused by Incomplete Lineage Sorting (ISL) alone; i.e., if ILS can be excluded as the cause of discordance, hybridization is the main source of this kind of differences between trees. This method is implemented in the software JML (Joly, 2012), and uses the posterior distribution of species trees estimated in *BEAST (Heled and Drummond, 2010), as implemented in the BEAST software (Drummond and Rambaut, 2007; Drummond et al., 2012). Species trees were estimated both with and without the four mismatched specimens, and the same settings were used in both analyses. The sequences were divided into the species delimited using BPP (see Section 3.4). The substitution models used were GTR + Γ for COI and HKY + Γ for H3, and empirical base frequencies were used. The Yule process speciation prior, and the piecewise linear with constant root population size prior were used, and the population size (ploidy level) of COI was set to half of that of the H3. Uncorrelated lognormal clocks were used, the rate was fixed to 1 for COI and estimated for H3 with a uniform prior of 0–1 for the relax clock rate (ucld.mean). For species population mean and mean growth rate priors, an exponential distribution with mean 1 was used. For all other priors, default settings were used. The analyses were run for 100 million generations, sampling every 10,000 generation. Tracer v1.6 was used for examining effective sample size (ESS) for parameters and determining burn-in, and trees were summarized using TreeAnnotator v1.8, using the maximum clade credibility tree, discarding the first 10% as burn-in.

Table 1
Genetic distances (COI), in percent, given for within group comparisons as maximum pairwise distance, between groups comparisons are given as minimum pairwise distances. All values are uncorrected pairwise distances. Within group comparisons of groups consisting of singletons are not applicable (n/a).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.
1. *L. castaneus*	5.0												
2. *L. festivus*	18.8	1.4											
3. *L. herculeus*	18.1	15.2	4.3										
4. *L. rubellus* A	19.6	17.4	17.7	4.9									
5. *L. rubellus* B	16.5	16.2	17.0	12.5	2.6								
6. *L. rubellus* G	19.8	17.0	17.8	12.9	13.2	n/a							
7. *L. rubellus* H1	18.8	16.0	17.1	11.9	12.6	10.2	n/a						
8. *L. rubellus* H2	19.5	17.5	16.3	12.8	13.1	10.2	5.9	n/a					
9. *L. rubellus* J	17.9	15.2	17.3	16.4	13.8	14.1	14.7	14.4	0.8				
10. *L. rubellus* K	17.2	16.9	18.9	14.9	13.3	13.7	14.0	14.2	12.3	1.3			
11. *L. rubellus* M	19.6	15.7	18.4	16.4	14.1	16.0	13.7	14.4	9.0	13.1	n/a		
12. *L. terrestris* 1	19.0	16.0	14.2	15.4	15.0	16.4	14.9	13.9	16.0	16.1	17.5	6.2	
13. *L. terrestris* 2	19.4	14.5	15.5	19.1	17.2	18.1	16.0	16.3	18.8	15.8	18.3	14.1	0.2

For each of the two species tree analyses, we compared the genetic distances from 1000 gene trees, simulated under species trees from the posterior distribution of the *BEAST analysis, to the pairwise genetic distances in both the H3 and COI datasets, with the mismatching specimens placed according to their COI as well as to their H3 sequences. That is, in total eight analyses were run. The mean clock rates and heredity scalar, for each gene from the *BEAST analyses were used, i.e. the effective population size for the mitochondrial COI was set to half of that of the nuclear H3 and the mutation rate of H3 was set to about 0.2 of that of COI. The default significance level was set to $P \leqslant 0.1$ for the analyses. If the specimens are of hybrid origin we expect significantly shorter distances than those to be expected by ILS alone in the analyses where the mismatched specimen is placed according to the gene that is not analyzed, i.e. we expect significant distances for a mismatched specimen placed according to its COI sequence when we compare the pairwise distances of H3 to the simulated distances, and vice versa. We also expect that we will find more significant results when we use the species tree analysis without the mismatched specimens as input, compared to when we use the species tree analysis with them as input. As the inclusion of hybrids will affect the species tree estimations, i.e., the trees estimated are likely to find the species involved in the hybridization as more closely related and with smaller genetic distances between them compared to trees estimated without hybrids.

3. Results

3.1. DNA sequencing, assembly and distance analysis

All 122 included specimens were successfully sequenced for both markers, and after phasing the H3 dataset included 172 sequences. The COI alignment was 658 bp long and the H3 alignment 306 bp long. The maximal intra-group pairwise distances varied between 0.2% in *L. terrestris* 2–6.2% in *L. terrestris* 1, and the minimum inter-group distances varied between 5.9% for *L. rubellus* H1 and H2 to 19.8% for *L. castaneus* and *L. rubellus* G. Three groups, *L. rubellus* G, H1 and H2, only comprise one specimen each, and thus no intra-group comparisons could be done for them (Table 1).

3.2. Haplotype networks

To visualize haplotype diversity and distribution in the two markers (COI and H3) we constructed haplotype networks. In the COI network (Fig. 1A) all lineages form distinct, well separated haplotype groups. However, in the H3 network (Fig. 1B), no separation between *L. terrestris* 1 and 2 exists, four haplotypes are shared between them, and all haplotypes are intermixed with each other. Moreover, one individual of *L. terrestris* 1 (CE4550) shares its (homozygous) H3 haplotype with *L. herculeus*, which otherwise was found as a distinct group. In the group consisting of H3 haplotypes of *L. rubellus* lineage A are also the H3 sequences from 2 individuals of *L. rubellus* B (CE8089 and CE11832) and the only individual of *L. rubellus* H2 (CE3584). With the exception of these four mismatching specimens, and the complete mixing of *L. terrestris* 1 and 2 haplotypes, all lineages form distinct haplotype groups, while the mismatched specimens are signs of possible hybridization.

3.3. Gene tree estimations

To test the support for monophyly of the mt groups we estimated single gene genealogies using Bayesian Inference. The estimated genealogies for both markers showed good convergence and high ESS (estimated sample size) values.

In the COI tree (Fig. 2A) all groups with more than one sequence are monophyletic with maximum support, and groups with only one sequence were always found as distinct lineages, but deeper branches often have low support. The genus *Lumbricus* as such was found monophyletic, but only with some support (PP = 0.90); *L. terrestris* 2 and *L. festivus* form a clade (PP = 0.99); *L. rubellus* J, M and K was found together (PP = 0.99) and in this clade *L. rubellus* J and M are sister-groups (PP = 1); the remaining lineages of *L. rubellus* form another group (PP = 0.96), in this group *L. rubellus* B is sister to the remaining lineages (PP 0.96), and there is support (PP = 1) for a sister-group relationship between *L. rubellus* H1 and H2.

In the H3 tree (Fig. 2B) *Lumbricus* was found monophyletic with maximum support. *Lumbricus castaneus* was found as sister-group to the remaining *Lumbricus* (PP = 0.98), which splits into one clade consisting of *L. terrestris*, *L. herculeus* and *L. festivus* (PP = 1), and one consisting of all lineages of *L. rubellus* (PP = 0.85). As also seen in the haplotype network, sequences of *L. terrestris* 1 and 2 are completely mixed, *L. herculeus* is monophyletic, with the exception of the specimen of *L. terrestris* 1 (CE4550) that was found within *L. herculeus* (PP = 0.96). Even if this specimen is not considered *L. terrestris* was found paraphyletic with respect to *L. herculeus*. In the clade consisting of all *L. rubellus* lineages, *L. rubellus* A is a basal, poorly resolved and paraphyletic group, also including the two specimens of *L. rubellus* B (CE8089 and CE11832), and the specimen of *L. rubellus* H2 (CE3584); the four mismatched specimens that are placed together with a species different from the one suggested by COI, are indicated in red in Fig. 2B. The remaining lineages of *L. rubellus* were found in a weakly supported (PP = 0.76) clade nested inside *L. rubellus* A. This clade forms a polytomy consisting of four groups: *L. rubellus* B (excluding the two specimens found within *L. rubellus* A); *L. rubellus* J; *L. rubellus* H1; and a clade consisting of *L. rubellus* G, K and M. Within the last-mentioned clade *L. rubellus* M is monophyletic with maximum support, whereas *L. rubellus* K is paraphyletic; *L. rubellus* G is a single specimen only.

3.4. Molecular species delimitation

We tested the support for mt lineages (see Section 2.2) and delimited species by performing multi-locus species delimitation analyses. In two different analyses the influence of the population size and root divergence priors was tested. The runs in each analysis gave consistent results and the means of the posterior probability values of the three runs are considered here. In analysis A, all species except *L. rubellus* H1 and *L. rubellus* H2 were delimited with a posterior probability (PP) of <0.95, and in analysis B, *L. terrestris* 1 and *L. terrestris* 2 were not strongly supported (PP 0.736) either. Both analyses found the same species tree as the most probable, but with low support ($PP \leqslant 0.1$), and a summary of the support of all species hypotheses is mapped on this tree (Fig. 3). Based on these results we consider *L. castaneus*, *L. festivus*, *L. herculeus*, *L. rubellus* A, *L. rubellus* B, *L. rubellus* G, *L. rubellus* J, *L. rubellus* K, *and L. rubellus* M well delimited and they can be seen as different species, but we do not have support to split *L. rubellus* H1 and H2 or *L. terrestris* 1 and 2 from each other. This gives us 11 well supported species in our dataset.

3.5. Species tree estimation and testing for incomplete lineage sorting vs. hybridization

To find the relationships between the species, and to explore the effect of the mismatched specimens on the phylogeny we estimated species trees, using the species delimited in the analyses above as input. Two species tree estimations were performed,

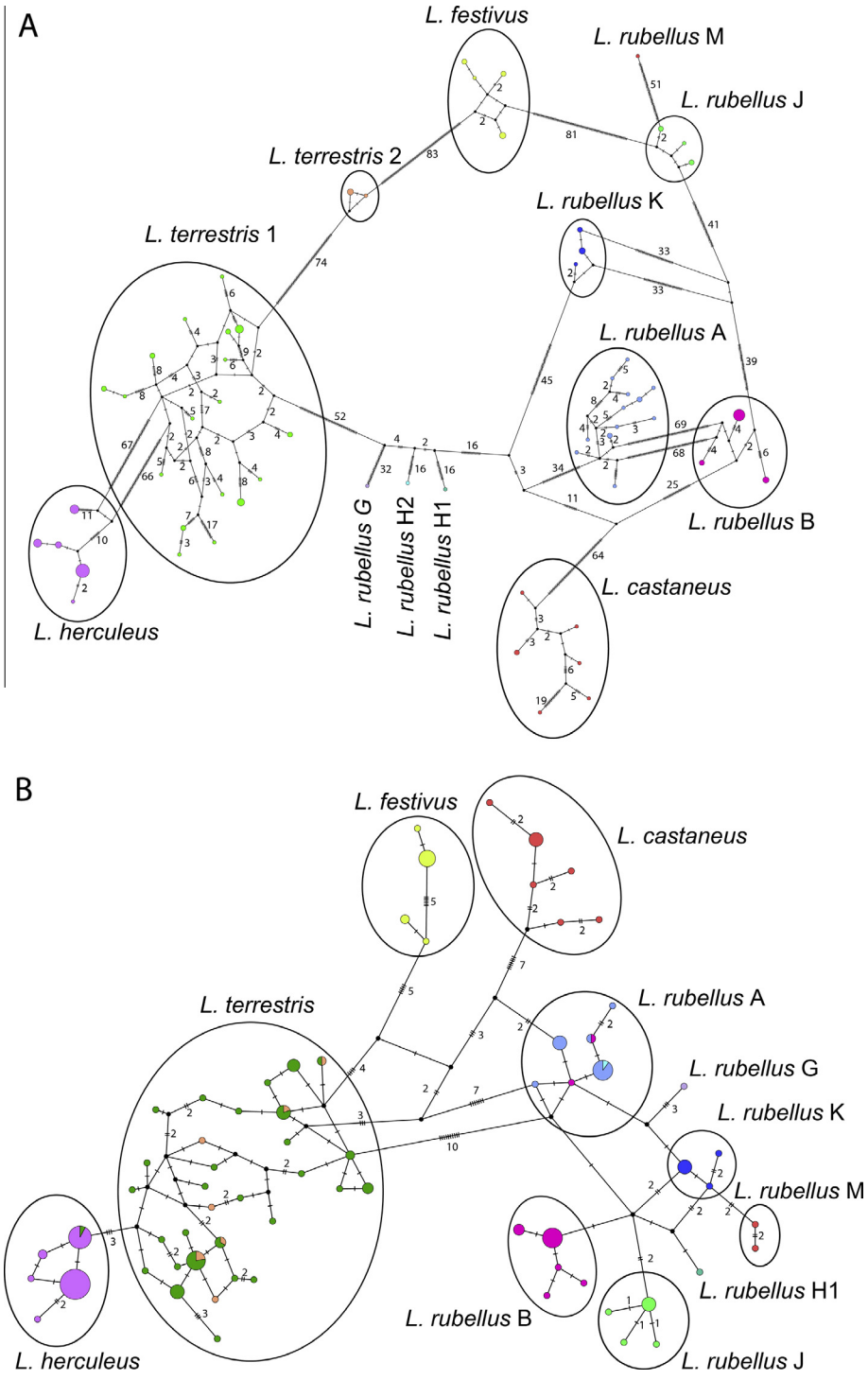

Fig. 1. Statistical parsimony haplotype networks. The size of the circles is relative to the number of sequences sharing that haplotype, the colors correspond to mt lineages, and the number are the number of substitutions between haplotypes. A. COI network. B. H3 network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

one with all specimens (A) and one without the mismatched specimens (B). Trees from the posterior distribution of these two analyses were used as input for testing if the four mismatched specimen are of hybrid origin, or if they can be explained by incomplete lineage sorting (ILS) alone. The two maximum clade credibility species trees (Fig. 4A and B) from the *BEAST analyses are similar: in both trees, *Lumbricus* is monophyletic with low support, and *L. castaneus* is sister to the remaining *Lumbricus* species, also with low support. The remaining species are divided into two clades, both with maximum support; one consists of *L. festivus*, *L. herculeus*, and *L. terrestris*, the other clade is the species within *L. rubellus s.lat*. In the first clade *L. terrestris* and *L. herculeus* are found as sister-group with no to low support. The *L. rubellus* clade is divided into two groups, one well supported in both analyses, consisting of *L. rubellus* J, K and M, and the other low to well supported group consisting of *L. rubellus* A, B, G and H. In the analysis with the mismatched specimens placed according to their COI sequences (Fig. 4A), *L. rubellus* A and B are sister-groups and sister to them

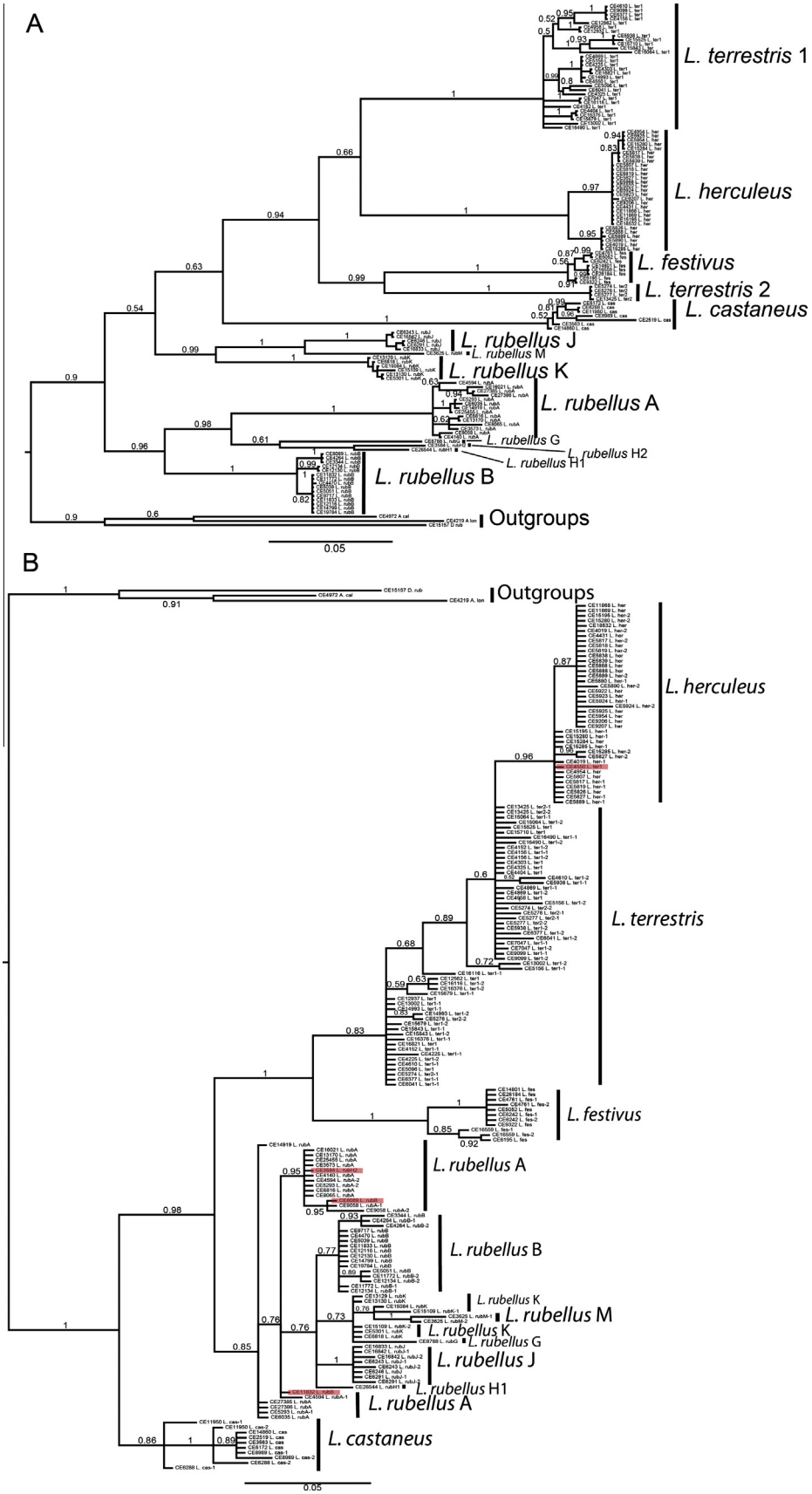

Fig. 2. Single gene trees estimated using Bayesian inference. Scale is given as expected number of substitutions/base. A. COI tree. B. H3 tree, specimens marked in red are placed in together with another species than in the COI tree. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

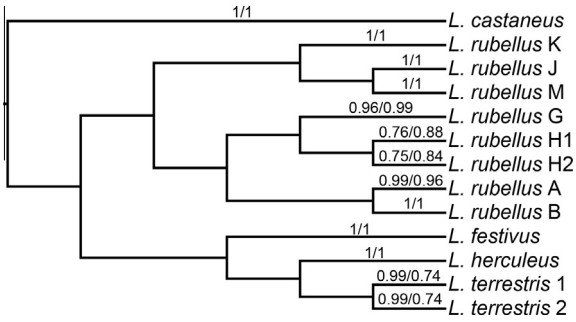

Fig. 3. Species tree estimated in BPP, with posterior probabilities (PP) for species delimitations given as mean of analysis A/analysis B, species with both PP > 0.95 are considered as well supported species, whereas we do not consider species with PP < 0.95 supported and therefore consider *L. rubellus* H1 and H2 to constitute one species, the same for *L. terrestris* 1 and 2, they are also considered as one single species.

is *L. rubellus* H, whereas *L. rubellus* G is sister to all these three. However, the relationship between these four species has low or no support. In the species tree with the mismatched specimens removed (Fig. 4B), the branching order is reversed, with *L. rubellus* G and H as sisters, followed by *L. rubellus* A, and as sister to these three *L. rubellus* B. The support for these relationships is higher than in the other species tree, but still low.

The results of the posterior predictive checking analyses are summarized in Table 2. When the pairwise distances in the H3 and COI datasets were tested against the distribution of pairwise distances in the simulated gene trees, we found significantly smaller distances than expected under the assumption of incomplete lineage sorting (ILS) alone in the following cases: (1) for H3 distances between CE4550 (a *L. terrestris* according to its COI) and specimens of *L. herculeus* in both analyses where CE4550 was placed in *L. terrestris*, and (2) for COI distances between CE4550 and specimens of *L. terrestris* in both analyses where CE4550 was placed in *L. herculeus* (where it belongs as suggested by H3). (3) For H3 distances between CE3584 (a *L. rubellus* H according to its COI) and specimens of *L. rubellus* A in both analyses where CE3584 was placed in *L. rubellus* H (as suggested by its COI sequence). (4) For H3 distances between both CE8089 and CE11832 (both *L. rubellus* B according to COI) and specimens of *L. rubellus* A in the analysis based on species trees estimated without the mismatched specimens included. (5) For COI distances between both CE8089 and CE11832 (both *L. rubellus* A according to H3) and specimens of *L. rubellus* B in the analysis based on species trees estimated without the mismatched specimens included. To summarize, for CE4550 we found significant results in all four analysis were we expected it, if the specimen is of hybrid origin, and for the other the specimens (CE3584, CE8089 and CE11832) we found significant results in two of the four analyses where we expected it. For CE8089 and CE11832 we found significantly short distances only when using the species trees estimated without the mismatched specimens included. However, for specimen CE3584 we only found significant H3 distances when it was placed according to COI, regardless of which species trees were used (Table 2). In all analyses comparing the simulated distances to the H3 dataset many other comparisons, not involving these four mismatched specimens, gave significant results, indicating a poor fit of this data set to the assumption of ILS only. The results from the JML analyses are what can be expected if the four mismatched specimens are of hybrid origin (see Section 2.6), and therefore they support the hypothesis of a hybrid origin for these four specimens.

4. Discussion

Our results confirm the split between *L. terrestris* and *L. herculeus* suggested by James et al. (2010), and the separation of *L. rubellus* A and B into different species, as suggested in studies on British *L. rubellus* (Donnelly et al., 2013, 2014; Jones et al., 2016). Unfortunately, of the Polish mt lineages studied by Giska et al. (2015) we only had material of *L. rubellus* A available for this study, making it impossible to compare our result with theirs. However, we have support for the existence of five additional cryptic species of *L. rubellus s.l.*

In general, studies on hybridization in earthworms are rare, but hybridization between lineages of *Allolobophora chlorotica* has been found (Dupont et al., 2011, 2016). This is the first study that explores hybridization between *Lumbricus* species, and we found evidence of limited hybridization between *L. terrestris* and *L. herculeus*, as well as between some species within *L. rubellus s.lat.* However, the analyses do not tell us anything about the age of the hybridization events. It is noteworthy that all cases of hybridization observed by us seem to be unidirectional, i.e., the hybrids always contain COI sequences of one of its parent species but not the other, whereas they contain the H3 species from the other parent species. In our case, all mismatched worms in the *L. rubellus* complex clustered with *L. rubellus* A using the nuclear data, and with either *L. rubellus* B or H using COI. A similar pattern was also observed by Dupont et al. (2016) in their study of the *A. chlorotica* complex, and it has also been showed experimentally that hybrids between the two color forms of that species are male sterile (Lowe and Butt, 2008), i.e., the male sexual system in the worms is non-functional, while the female system remains functional. As the observed hybridization within the *L. rubellus* complex is unidirectional, it is possible that *L. rubellus* A-B/H crosses are also male sterile. However, this needs to be tested in breeding experiments. It is also worth noting that in all *Lumbricus* species, as far as it is known, the chromosome number is always the same (i.e., 2n = 36) and no polyploid specimens have been reported (Sims

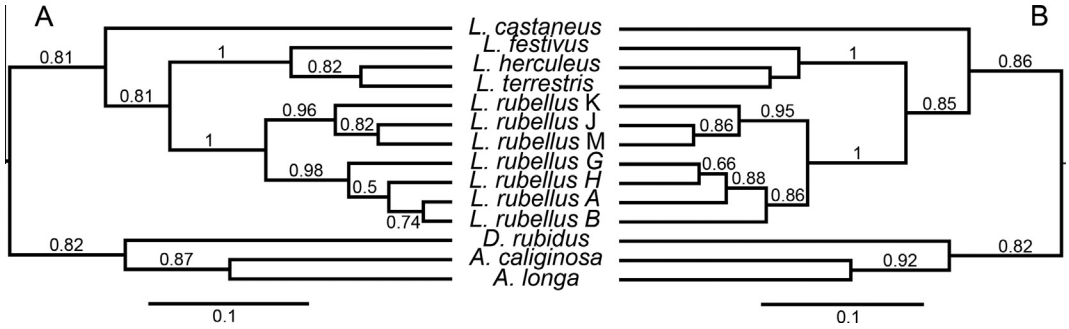

Fig. 4. Species trees estimated using *BEAST. A. Species tree including the mismatched, hybrid specimens. B. Species tree excluding the mismatched specimens. Only posterior probability >0.5 are shown as support values.

Table 2
Results from posterior predictive checking analyses performed in JML. Mismatched specimens are compared to specimens of the species that either their COI or their H3 sequence places them in. P-values for the probability that observed pairwise genetic distances between specimens are explained by incomplete lineage sorting (ILS) alone are given for each comparison and analysis. If the null-hypothesis of ILS can be rejected, the alternative hypothesis of hybridization is accepted. Only P-values <0.1 are given.

	Species trees with mismatched specimens				Species trees without mismatched specimens			
	Mismatched specimens placed according to COI		Mismatched specimens placed according to H3		Mismatched specimens placed according to COI		Mismatched specimens placed according to H3	
Comparisons	COI dist.	H3 dist.	COI dist.	H3 dist.	COI dist.	H3 dist.	COI dist.	H3 dist.
CE4550 – *L. herculeus*	–	0.024	–	–	–	0.002	–	–
CE4550 – *L. terrestris*	–	–	0.012	–	–	–	0.003	–
CE3584 – *L. rubellus* A	–	0.025	–	–	–	0.005	–	–
CE3584 – *L. rubellus* H	–	–	–	–	–	–	–	–
CE8089 – *L. rubellus* A	–	–	–	–	–	0.003	–	–
CE8089 – *L. rubellus* B	–	–	–	–	–	–	0.004	–
CE11832 – *L. rubellus* A	–	–	–	–	–	0.01	–	–
CE11382 – *L. rubellus* B	–	–	–	–	–	–	0.004	–

and Gerard, 1985; Csaba Csuzdi, pers. comm.). The mismatched *L. terrestris*/*herculeus* specimen (CE4550) was found in the Swedish province of Jämtland, which is about 500 km north of the northernmost record of (other) *L. herculeus*, which in Scandinavia are mainly found in Denmark, and the south of Sweden and Norway (James et al., 2010; C. Erséus, unpublished records). It seems likely that the record of an *L. herculeus* H3 haplotype that far north is due to human introduction, either of the specimen itself or of *L. herculeus* specimens that have hybridized with local *L. terrestris*.

In *L. terrestris* there are two well separated mt lineages (*terrestris* 1 and 2), but these were completely mixed in our H3 haplotype network (Fig. 1B) and H3 tree (Fig. 2B). Similar results were previously reported for *Aporrectodea longa* (Martinsson et al., 2015), but in that case the COI distances between the two groups were not as large as in *L. terrestris* (approx. 7% in *A. longa* vs. 14% in *L. terrestris*). The COI distances between the *L. terrestris* 1 and 2 are more in line with what is seen between other *Lumbricus* species, and also what was observed by Giska et al. (2015) for three lineages of *L. rubellus*, that were separated by approximately 13–17% pairwise differences. Giska et al. (2015), however, did not find, using RADseq data, any support for their lineages to be different species. *Lumbricus terrestris* now is yet another example of deep intraspecific coalescence in COI (with up to 14% distances), and as a general conclusion, it cannot be automatically assumed that COI lineages which are that far apart, represent different species, without testing this with nuclear data. This illustrates the big problem with a DNA-barcoding gap as the sole criterion for species delimitation in Clitellata, which has also been noted in other studies (e.g., Achurra and Erséus, 2013; Martinsson et al., 2013, 2015). Deep intra-specific splits between different lineages have been observed in other organisms as well (e.g., Hogner et al., 2012; Munoz et al., 2011; Webb et al., 2011).

When using the program BPP for species delimitation, as we have done in this study, it is recommended to run the analyses several times with different values of the priors for the population size (θs) and the divergence time at the root (τ0) (Rannala, 2015). These priors can drastically affect the results, especially the θ prior (Leache and Fujita, 2010; Rannala, 2015). A small θ generally increases the probability for splitting (Rannala, 2015). However, we observed an opposite result. When θ was reduced the probability for a split between *L. terrestris* 1 and 2 was lowered (PP = 0.74 compared to PP = 0.99). We noted small increases in PP for some species, mainly *L. rubellus* H1 and H2, but these changes were not as dramatic as for *L. terrestris* 1 and 2.

Lumbricus rubellus and *L. terrestris* are both well-studied model organisms (Giska et al., 2015; James et al., 2010), e.g., with *L. rubellus* commonly used in ecotoxicological research. Several other clitellate species are also popular in experimental lab work (Halanych and Borda, 2009), and many morphospecies used are complexes of cryptic species, and in other cases described species are misidentified and either used alone under the wrong name or as a mixture with the species it has been identified to (e.g., Gustafsson et al., 2009; Martinsson and Erséus, 2014; Römbke et al., 2016; Siddall et al., 2007). The existence of cryptic lineages among such model morphospecies is problematic, as they may differ in both physiological and ecological traits (Feckler et al., 2013, 2014). For *L. rubellus* it has been shown that two lineages vary in their methylation pattern in response to arsenic pollution (Kille et al., 2013). It is important to better define the species boundaries in taxa used as models, and then to explore the possible ecological and physiological differences between the cryptic taxa revealed. Needless to say, in all these cases the species studied should be identified by molecular methods, e.g., DNA-barcoding. However, if the cryptic species are known to hybridize it is problematic to use DNA-barcoding based on COI alone, as it makes it impossible to identify hybrids.

To conclude, using nuclear DNA we found good support for the existence of seven cryptic species within the morphospecies *Lumbricus rubellus*, and confirmation of the separation of *L. terrestris* and *L. herculeus*. We also found evidence for limited hybridization between the latter two species and within the *L. rubellus* species complex.

Acknowledgements

We are indebted to Thomas Cedhagen, Trond Knapp Haraldsen, Arne Larsson, Urban Olsson, and Ted von Proschwitz, for collection and donation of specimens; to Anna Ansebo, Pierre De Wit, Per Hjelmstedt, Samuel W. James, Magnus Johansson, Mårten Klinth, Emelie Lindquist, Maria Lindström, Yingkui Liu, Mark J. Wetzel, Endre Willassen and Bronwyn W. Williams, for assistance in the field and/or the molecular lab; to Csaba Csuzdi for comments regarding chromosome numbers in *Lumbricus* spp; to Kerryn Elliott for comments that improved the language; and to the Swedish Research Council, the Swedish Taxonomy Initiative (ArtDatabanken, Uppsala), the Norwegian Taxonomy Initiative (Artsdatabanken, Trondheim), the Adlerbert Foundation, and the Royal Society of Arts and Sciences in Gothenburg, for financial support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.09.011.

References

Achurra, A., Erséus, C., 2013. DNA barcoding and species delimitation: the *Stylodrilus heringianus* case (Annelida: Clitellata: Lumbriculidae). Invertebr. Syst. 27, 118–128. http://dx.doi.org/10.1071/Is12049.

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 Ecol. Evol. 22, 148–155. http://dx.doi.org/10.1016/j.tree.2006.11.004.

Buckley, T.R., James, S., Allwood, J., Bartlam, S., Howitt, R., Prada, D., 2011. Phylogenetic analysis of New Zealand earthworms (Oligochaeta: Megascolecidae) reveals ancient clades and cryptic taxonomic diversity. Mol. Phylogenet. Evol. 58, 85–96. http://dx.doi.org/10.1016/j.ympev.2010.09.024.

Chang, C.H., Lin, S.M., Chen, J.H., 2008. Molecular systematics and phylogeography of the gigantic earthworms of the *Metaphire formosae* species group (Clitellata, Megascolecidae). Mol. Phylogenet. Evol. 49, 958–968. http://dx.doi.org/10.1016/j.ympev.2008.08.025.

Clement, M., Snell, Q., Walke, P., Posada, D., Crandall, K., 2002. TCS: estimating gene genealogies. Proc. 16th Int. Parallel Distrib. Process Symp., vol. 2, p. 184. http://dx.doi.org/10.1109/IPDPS.2002.1016585.

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. Mol. Ecol. Resour. 10, 264–273. http://dx.doi.org/10.1111/j.1755-0998.2009.02763.x.

Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340. http://dx.doi.org/10.1016/j.tree.2009.01.009.

Donnelly, R.K., Harper, G.L., Morgan, A.J., Orozco-Terwengel, P., Pinto-Juma, G.A., Bruford, M.W., 2013. Nuclear DNA recapitulates the cryptic mitochondrial lineages of *Lumbricus rubellus* and suggests the existence of cryptic species in an ecotoxological soil sentinel. Biol. J. Linn. Soc. 110, 780–795. http://dx.doi.org/10.1111/bij.12171.

Donnelly, R.K., Harper, G.L., Morgan, A.J., Pinto-Juma, G.A., Bruford, M.W., 2014. Mitochondrial DNA and morphological variation in the sentinel earthworm species *Lumbricus rubellus*. Eur. J. Soil Biol. 64, 23–29. http://dx.doi.org/10.1016/j.ejsobi.2014.07.002.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. http://dx.doi.org/10.1186/1471-2148-7-214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. http://dx.doi.org/10.1093/molbev/mss075.

Dupont, L., Lazrek, F., Porco, D., King, R.A., Rougerie, R., Symondson, W.O.C., Livet, A., Richard, B., Decaëns, T., Butt, K.R., Mathieu, J., 2011. New insight into the genetic structure of the *Allolobophora chlorotica* aggregate in Europe using microsatellite and mitochondrial data. Pedobiologia 54, 217–224. http://dx.doi.org/10.1016/j.pedobi.2011.03.004.

Dupont, L., Porco, D., Symondson, W.O., Roy, V., 2016. Hybridization relics complicate barcode-based identification of species in earthworms. Mol. Ecol. Resour. 16, 883–894. http://dx.doi.org/10.1111/1755-0998.12517.

Erséus, C., Gustafsson, D., 2009. Cryptic speciation in clitellate model organisms. In: Shain, D.H. (Ed.), Annelids in Modern Biology. John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 31–46.

Feckler, A., Schulz, R., Bundschuh, M., 2013. Cryptic lineages-same but different? Integr. Environ. Assess. Manage. 9, 172–173. http://dx.doi.org/10.1002/ieam.1370.

Feckler, A., Zubrod, J.P., Thielsch, A., Schwenk, K., Schulz, R., Bundschuh, M., Frid, C., 2014. Cryptic species diversity: an overlooked factor in environmental management? J. Appl. Ecol. 51, 958–967. http://dx.doi.org/10.1111/1365-2664.12246.

Fernández, R., Almodóvar, A., Novo, M., Gutiérrez, M., Díaz Cosín, D.J., 2011. A vagrant clone in a peregrine species: phylogeography, high clonal diversity and geographical distribution in the earthworm *Aporrectodea trapezoides* (Dugès, 1828). Soil Biol. Biochem. 43, 2085–2093. http://dx.doi.org/10.1016/j.soilbio.2011.06.007.

Fujita, M.K., Leache, A.D., Burbrink, F.T., McGuire, J.A., Moritz, C., 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27, 480–488. http://dx.doi.org/10.1016/j.tree.2012.04.012.

Giska, I., Sechi, P., Babik, W., 2015. Deeply divergent sympatric mitochondrial lineages of the earthworm *Lumbricus rubellus* are not reproductively isolated. BMC Evol. Biol. 15, 217. http://dx.doi.org/10.1186/s12862-015-0488-9.

Gustafsson, D.R., Price, D.A., Erséus, C., 2009. Genetic variation in the popular lab worm *Lumbriculus variegatus* (Annelida: Clitellata: Lumbriculidae) reveals cryptic speciation. Mol. Phylogenet. Evol. 51, 182–189. http://dx.doi.org/10.1016/j.ympev.2008.12.016.

Halanych, K.M., Borda, E., 2009. Developing models for Lophotrochozoan and Annelid biology. In: Shain, D.H. (Ed.), Annelids in Modern Biology. John Wiley & Sons, Hoboken, New Jersey, pp. 1–12.

Hebert, P.D., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. Biol. Sci. Ser. B 270, 313–321. http://dx.doi.org/10.1098/rspb.2002.2218.

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. http://dx.doi.org/10.1016/j.pedobi.2003.04.001.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580. http://dx.doi.org/10.1093/molbev/msp274.

Hoffmeister, W., 1843. Beitrag zor Kenntnis deutscher Landanneliden. Archiv. Naturgesch. 91, 183–198.

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*. Ecol. Evol. 2, 2974–2988. http://dx.doi.org/10.1002/ece3.398.

James, S.W., 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. http://dx.doi.org/10.1371/journal.pone.0015629.

Joly, S., 2012. JML: testing hybridization from species trees. Mol. Ecol. Resour. 12, 179–184. http://dx.doi.org/10.1111/j.1755-0998.2011.03065.x.

Joly, S., McLenachan, P.A., Lockhart, P.J., 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. Am. Nat. 174, E54–E70. http://dx.doi.org/10.1086/600082.

Jones, G.L., Wills, A., Morgan, A.J., Thomas, R.J., Kille, P., Novo, M., 2016. The worm has turned: behavioural drivers of reproductive isolation between cryptic lineages. Soil Biol. Biochem. 98, 11–17. http://dx.doi.org/10.1016/j.soilbio.2016.03.015.

Kille, P., Andre, J., Anderson, C., Ang, H.N., Bruford, M.W., Bundy, J.G., Donnelly, R., Hodson, M.E., Juma, G., Lahive, E., Morgan, A.J., Stürzenbaum, S.R., Spurgeon, D.J., 2013. DNA sequence variation and methylation in an arsenic tolerant earthworm population. Soil Biol. Biochem. 57, 524–532. http://dx.doi.org/10.1016/j.soilbio.2012.10.014.

King, R.A., Tibble, A.L., Symondson, W.O., 2008. Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. Mol. Ecol. 17, 4684–4698. http://dx.doi.org/10.1111/j.1365-294X.2008.03931.x.

Leache, A.D., Fujita, M.K., 2010. Bayesian species delimitation in West African forest geckos (*Hemidactylus fasciatus*). Proc. R. Soc. Biol. Sci. Ser. B 277, 3071–3077. http://dx.doi.org/10.1098/rspb.2010.0662.

Leigh, J.W., Bryant, D., 2015. Popart: full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116. http://dx.doi.org/10.1111/2041-210X.12410.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. http://dx.doi.org/10.1093/bioinformatics/btp187.

Linnaeus, C., 1758. Systema Naturae, 10th ed. Salvi, Holmiae.

Lowe, C.N., Butt, K.R., 2008. *Allolobophora chlorotica* (Savigny, 1826): evidence for classification as two separate species. Pedobiologia 52, 81–84. http://dx.doi.org/10.1016/j.pedobi.2008.04.001.

Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229–237. http://dx.doi.org/10.1016/j.tree.2005.02.010.

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. Scr. 42, 612–622. http://dx.doi.org/10.1111/zsc.12032.

Martinsson, S., Erséus, C., 2014. Cryptic diversity in the well-studied terrestrial worm *Cognettia sphagnetorum* (Clitellata: Enchytraeidae). Pedobiologia 57, 27–35. http://dx.doi.org/10.1016/j.pedobi.2013.09.006.

Martinsson, S., Rhodén, C., Erséus, C., 2015. Barcoding gap, but no support for cryptic speciation in the earthworm *Aporrectodea longa* (Clitellata: Lumbricidae). Mitochondr. DNA, 1–9. http://dx.doi.org/10.3109/19401736.2015.1115487.

Munoz, A.G., Baxter, S.W., Linares, M., Jiggins, C.D., 2011. Deep mitochondrial divergence within a Heliconius butterfly species is not explained by cryptic speciation or endosymbiotic bacteria. BMC Evol. Biol. 11, 358. http://dx.doi.org/10.1186/1471-2148-11-358.

Novo, M., Almodovar, A., Fernandez, R., Trigo, D., Diaz-Cosin, D.J., Giribet, G., 2012. Appearances can be deceptive: different diversification patterns within a group of Mediterranean earthworms (Oligochaeta, Hormogastridae). Mol. Ecol. 21, 3776–3793. http://dx.doi.org/10.1111/j.1365-294X.2012.05648.x.

Novo, M., Almodovar, A., Fernandez, R., Trigo, D., Diaz Cosin, D.J., 2010. Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. Mol. Phylogenet. Evol. 56, 507–512. http://dx.doi.org/10.1016/j.ympev.2010.04.010.

Nygren, A., 2014. Cryptic polychaete diversity: a review. Zool. Scr. 43, 172–183. http://dx.doi.org/10.1111/zsc.12044.

Pfenninger, M., Schwenk, K., 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol. Biol. 7, 121. http://dx.doi.org/10.1186/1471-2148-7-121.

Porco, D., Decaëns, T., Deharveng, L., James, S., Skarżyński, D., Erséus, C., Butt, K., Richard, B., Hebert, P.N., 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. Biol. Invasions 15, 899–910. http://dx.doi.org/10.1007/s10530-012-0338-2.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.5 Available from <http://beast.bio.ed.ac.uk/Tracer>.

Rannala, B., 2015. The art and science of species delimitation. Curr. Zool. 61, 846–853.

Rannala, B., Yang, Z., 2013. Improved reversible jump algorithms for Bayesian species delimitation. Genetics 194, 245–253. http://dx.doi.org/10.1534/genetics.112.149039.

Rogers, J., Gibbs, R.A., 2014. Comparative primate genomics: emerging patterns of genome content and dynamics. Nat. Rev. Genet. 15, 347–359. http://dx.doi.org/10.1038/nrg3707.

Römbke, J., Aira, M., Backeljau, T., Breugelmans, K., Domínguez, J., Funke, E., Graf, N., Hajibabaei, M., Pérez-Losada, M., Porto, P.G., Schmelz, R.M., Vierna, J., Vizcaíno, A., Pfenninger, M., 2016. DNA barcoding of earthworms (*Eisenia fetida/andrei* complex) from 28 ecotoxicological test laboratories. Appl. Soil Ecol. 104, 3–11. http://dx.doi.org/10.1016/j.apsoil.2015.02.010.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. http://dx.doi.org/10.1093/sysbio/sys029.

Savigny, J.C., 1826. Analyses de travaux de l'Académie Royale des Sciences pendant l'année 1821, partie physique. Zool. Mém. Acad. Sci. Inst. France (Hist.) 5, 176–184.

Sechi, P., 2013. An Evolutionary History of The Peregrine Epigeic Earthworm *Lumbricus rubellus* Cardiff School of Biosciences. Cardiff University Unpublished, p. 193.

Shekhovtsov, S.V., Berman, D.I., Bazarova, N.E., Bulakhova, N.A., Porco, D., Peltek, S. E., 2016. Cryptic genetic lineages in *Eisenia nordenskioldi pallida* (Oligochaeta, Lumbricidae). Eur. J. Soil Biol. 75, 151–156. http://dx.doi.org/10.1016/j.ejsobi.2016.06.004.

Shekhovtsov, S.V., Golovanova, E.V., Peltek, S.E., 2013. Cryptic diversity within the Nordenskiold's earthworm, *Eisenia nordenskioldi* subsp. *nordenskioldi* (Lumbricidae, Annelida). Eur. J. Soil Biol. 58, 13–18. http://dx.doi.org/10.1016/j.ejsobi.2013.05.004.

Shekhovtsov, S.V., Golovanova, E.V., Peltek, S.E., 2014. Genetic diversity of the earthworm *Octolasion tyrtaeum* (Lumbricidae, Annelida). Pedobiologia 57, 245–250. http://dx.doi.org/10.1016/j.pedobi.2014.09.002.

Siddall, M.E., Trontelj, P., Utevsky, S.Y., Nkamany, M., Macdonald, K.S., 2007. Diverse molecular data demonstrate that commercially available medicinal leeches are not *Hirudo medicinalis*. Proc. Biol. Sci. R. Soc. 274, 1481–1487. http://dx.doi.org/10.1098/rspb.2007.0248.

Sims, R.W., Gerard, B.M., 1985. Earthworms: Keys and Notes for the Identification and Study of the Species. Brill, London.

Siqueira, F.d.F., Sandes, S.H.d.C., Drumond, M.A., Campos, S.H., Martins, R.P., Fonseca, C.G.d., Carvalho, M.R.S., 2013. Genetic diversity and population genetic structure in giant earthworm *Rhinodrilus alatus* (Annelida: Clitellata: Glossoscolecidae). Pedobiologia 56, 15–21. http://dx.doi.org/10.1016/j.pedobi.2012.08.006.

Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73, 1162–1169. http://dx.doi.org/10.1086/379378.

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989. http://dx.doi.org/10.1086/319501.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. http://dx.doi.org/10.1093/molbev/mst197.

Templeton, A.R., Crandall, K.A., Sing, C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132, 619–633.

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.

Webb, W.C., Marzluff, J.M., Omland, K.E., 2011. Random interbreeding between cryptic lineages of the Common Raven: evidence for speciation in reverse. Mol. Ecol. 20, 2390–2402. http://dx.doi.org/10.1111/j.1365-294X.2011.05095.x.

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, 4354–4355.

Yang, Z., Rannala, B., 2010. Bayesian species delimitation using multilocus sequence data. PNAS 107, 9264–9269. http://dx.doi.org/10.1073/pnas.0913022107.

Yang, Z., Rannala, B., 2014. Unguided species delimitation using DNA sequence data from multiple Loci. Mol. Biol. Evol. 31, 3125–3135. http://dx.doi.org/10.1093/molbev/msu279.