Identifying earthworms through DNA barcodes: Pitfalls and promise

Chih-Han Chang[a], Rodolphe Rougerie[b], Jiun-Hong Chen[c],*

[a]Department of Life Science, National Taiwan University, No. 1, Roosevelt Road, Section 4, Taipei 106, Taiwan
[b]Biodiversity Institute of Ontario, University of Guelph, 579 Gordon Street, Guelph, Ontario, N1G 2W1, Canada
[c]Institute of Zoology and Department of Life Science, National Taiwan University, No. 1, Roosevelt Road, Section 4, Taipei 106, Taiwan

Summary

This paper re-evaluated the use of DNA barcodes in earthworm species identification by re-analyzing sequence data for the mitochondrial cytochrome *c* oxidase subunit 1 (COI) gene. This analysis unveiled and confirmed taxonomic inconsistencies which significantly affect data interpretation. When considering synonymy and misidentification in published records, our results revealed no shared COI haplotypes between morphologically distinct species and higher interspecific than intraspecific divergence in most cases, with interspecific and intraspecific distances averaging 18.7% and 1.3% respectively. However, a few earthworm species endemic to Taiwan have deep intraspecific divergences which may represent potential cases of cryptic diversity, although incomplete lineage sorting cannot be ruled out without further study. We recognize the potential of DNA barcoding for earthworm taxonomy, but have identified several issues regarding the evolution of the COI gene in these organisms which remain to be further elucidated.
© 2008 Elsevier GmbH. All rights reserved.

Introduction

Following the suggestion that a standardized region of the mitochondrial genome can be used as a tag, or DNA barcode, for species identification and discrimination (Hebert et al. 2003a, b), much attention has focused on the use of genetic data as a complement to or surrogate for classical taxonomic studies. Although not explicitly called DNA barcoding in the case of earthworms, this approach has been used for several years to discriminate morphologically similar species (Chang and Chen 2005; Pérez-Losada et al. 2005; Chang et al. 2007)

*Corresponding author. Tel.: +886 2 33662502.
E-mail addresses: r91225025@ntu.edu.tw,
yuanpau@gmail.com (C.-H. Chang), rrougeri@uoguelph.ca
(R. Rougerie), chenjh@ntu.edu.tw (J.-H. Chen).

by using the DNA sequences of the mitochondrial cytochrome *c* oxidase subunit 1 (COI) (the same region of the gene that Hebert et al. (2003a, b) proposed as a DNA barcode for animals). Given the rapid development of DNA barcoding and the extent to which it accelerates the exploration of biodiversity, it is critical to properly assess how well the method works in earthworms. A recent attempt to do so (Huang et al. 2007) analyzed partial COI sequences from 86 specimens of 28 Chinese earthworm species. The authors tested if a 566 bp portion of the mitochondrial genome could serve as a DNA barcode which would reliably discriminate species in earthworms. They concluded that the ''*DNA barcode is a powerful tool for identifying species of earthworms and provides a useful complement to traditional morphological taxonomy*'' (Huang et al. 2007, their summary). Although we largely agree with this conclusion and the extended benefits that could be expected from the generalized use of this tool in earthworm taxonomy, we are concerned that their study may suffer from taxonomic and methodological inconsistencies which could compromise research conclusions.

In this study, we briefly discuss these problems in light of results presented by Huang et al. (2007) as well as previous systematic studies of earthworms based on COI sequences and morphology. By re-analyzing sequences from the Huang et al. (2007) study and by combining them with other published datasets (Chang and Chen 2005; Pérez-Losada et al. 2005; Chang et al. 2007), we re-evaluated the phylogenetic trees and genetic distances proposed by Huang et al. (2007) in order to produce a more accurate scenario for the potential use of DNA barcoding in earthworm taxonomy.

Materials and methods

The data of Huang et al. (2007) and three other published studies (Chang and Chen 2005; Pérez-Losada et al. 2005; Chang et al. 2007) were re-analyzed. Species names and synonyms followed the Chinese earthworm checklist and the *Pheretima* complex checklist compiled by Blakemore (2004, 2006a, b). Voucher specimens from studies by Chang and Chen (2005) and Chang et al. (2007) were identified by the senior author; these specimens reside at the Institute of Zoology, National Taiwan University in Taipei, Taiwan, where they are available for inspection. The results of these two studies are considered here as a reliable taxonomic reference associated with COI sequences of Chinese

and Taiwanese earthworms. Because these records include specimens from almost the entire known range of most target species, we expected them to provide a reliable estimate of intraspecific genetic variation.

Sequences of earthworms from published works were retrieved from GenBank (EF077528–EF077607 and DQ835672–DQ835677 for Huang et al. (2007), AY739309–AY739340 for Chang and Chen (2005), DQ224166–DQ224191 for Chang et al. (2007), AY874487–AY874523 for Pérez-Losada et al. (2005), U24570 for *Lumbricus terrestris*, and AF003256 for *Pontodrilus litoralis*). To re-evaluate the results of Huang et al. (2007), we used their dataset of 87 sequences of 566 bp, including 86 specimens of Chinese earthworms and one sequence of *L. terrestris* from GenBank. We combined sequences from all four studies to produce a total of 185 sequences that were aligned using the default settings of Clustal X 1.81 (Thompson et al. 1997). Because some sequences were shorter than those used in Huang et al. (2007), a homologous fragment of 535 bp was used in the analyses, corresponding to sites 88–622 of the gene in the earthworm *L. terrestris*. Data analyses were conducted using MEGA 3.1 (Kumar et al. 2004). A Neighbor Joining (NJ) optimal tree was built based on the three codon positions; we used the p-distance metric, considering transitions and transversions, a homogeneous pattern among lineages, and uniform rates of substitution among sites. Other models available in MEGA were considered as well, and we also tested the GTR+I+G model employed by Huang et al. (2007) using PAUP 4.0b10 (Swofford 2000). As no significant differences were observed, we only show the uncorrected p-distances which were also provided by these authors. A bootstrap analysis with 1000 pseudo-replicates was conducted to evaluate the robustness of clusters inferred from the NJ analysis. Furthermore, the number of changes in amino-acid composition, as inferred in MEGA, was used as an additional variable to explore intraspecific and interspecific variation of the COI fragment.

Results

The topology of the NJ tree inferred from the whole dataset (Figure 1) clearly illustrates the very strong signal of COI at the species level in earthworms. The clusters observed were all monophyletic, strongly supported (all plain lines on Figure 1 lead to nodes with more than 95% bootstrap support), and deeply divergent (see below and
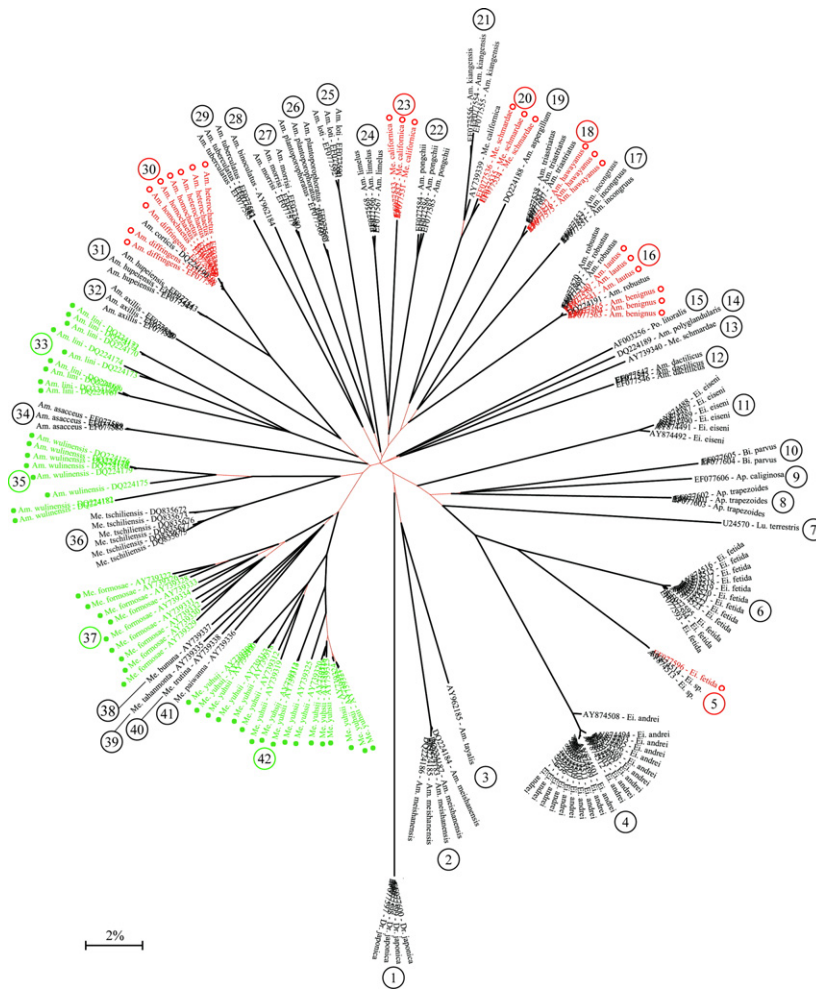
Figure 1. Neighbor Joining unrooted tree based on uncorrected p-distances for the 42 species analyzed after combination of partial COI sequences from Huang et al. (2007); Chang and Chen (2005); Chang et al. (2007) and Pérez-Losada et al. (2005). Plain black lines represent branches with bootstrap support over 95%; taxa in red (open circles) refer to inferred taxonomic inconsistencies in Huang et al. (2007) identifications, and taxa in green (closed circles) are those with high intraspecific divergences. Numbers refer to species names as follow: (1) *Drawida japonica*, (2) *Amynthas meishanensis*, (3) *Am. tayalis*, (4) *Eisenia andrei*, (5) *Eisenia* sp., (6) *E. fetida*, (7) *Lumbricus terrestris*, (8) *Aporectodea trapezoides*, (9) *Ap. Caliginosa*, (10) *Bimastos parvus*, (11) *E. eiseni*, (12) *Am. dactilicus*, (13) *Metaphire schmardae*, (14), *Am. polyglandularis*, (15) *Pontodrilus litoralis*, (16) *Am. robustus*, (17) *Am. incongruus,* (18) *Am. triastriatus*, (19) *Am. aspergillum*, (20) *M. californica*, (21) *Am. kiangensis*, (22) *Am. pongchii*, (23) *Metaphire* sp., (24) *Am. limelus*, (25) *Am. loti*, (26) *Am. plantoporophoratus*, (27) *Am. morrisi*, (28) *Am. binoculatus*, (29) *Am. tuberculatus*, (30) *Am. corticis*, (31) *Am. hubeiensis*, (32) *Am. axillis*, (33) *Am. lini*, (34) *Am. asacceus*, (35) *Am. wulinensis*, (36) *M. tschiliensis*, (37) *M. formosae*, (38) *M. bununa*, (39) *M. tahanmonta*, (40) *M. trutina*, (41) *M. paiwanna*, (42) *M. yuhsii*.

Table 1). The same conclusion, however, was not clear from the trees in Huang et al. (2007), because their figures did not show branch lengths proportional to genetic divergences. In several cases, the identifications provided by Huang et al. (Figure 1, red terminals) conflicted with those from the other datasets; these discrepancies are discussed later, but were, in the present study, considered as either overlooked synonymies or misidentifications. Ignoring these issues, the complete analysis included 42 recognized taxa, all forming monophyletic clusters.

Four taxa – *Amynthas lini, A. wulinensis, Metaphire formosae* and *M. yuhsii* – showed high intraspecific divergences (Figure 1, green terminals, numbered #33, #35, #37 and #42 respectively) ranging from 4.65% in *A. wulinensis* to 8.8% in *A. lini*. Table 1 summarizes divergences within and between species (presuming corrected identifications of the samples from Huang et al. (2007) for all taxa and all pairs of taxa. The mean intraspecific divergence was 1.30% [sd = 2.3] (0.45% [sd = 0.58] when we excluded those four taxa with high intraspecific

Table 1. Mean intra- (shaded boxes) and interspecific p-distances (given in percentage) for the 42 species analyzed after combination of partial COI sequences from Huang et al. (2007); Chang and Chen (2005); Chang et al. (2007) and Pérez-Losada et al. (2005)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[40]	[41]	[42]
[1] *Amynthas asacceus*	0.3																																									
[2] *Amynthas aspergillum*	19.6	-																																								
[3] *Amynthas axillis*	15.8	18.1	0																																							
[4] *Amynthas binoculatus*	17.4	19.6	19.0	-																																						
[5] *Amynthas corticis*	18.6	17.8	13.8	18.2	0.3																																					
[6] *Amynthas dactilicus*	18.1	18.5	16.4	18.6	17.4	0																																				
[7] *Amynthas triastriatus*	18.5	16.6	16.4	16.6	19.8	20.3	0.5																																			
[8] *Amynthas hupeiensis*	18.4	19.1	13.3	17.8	6.0	16.5	19.5	0.5																																		
[9] *Amynthas incongruus*	18.8	17.3	19.0	17.5	19.2	16.8	15.2	20.0	0																																	
[10] *Amynthas kiangensis*	20.9	17.6	20.9	22.0	21.2	19.7	18.8	20.8	17.9	0.9																																
[11] *Amynthas limelus*	18.3	18.4	16.9	18.3	15.2	15.7	18.1	15.9	17.3	19.7	0.6																															
[12] *Amynthas lini*	17.3	20.7	15.7	18.7	17.2	18.4	18.2	17.4	18.8	22.4	17.4	8.8																														
[13] *Amynthas loti*	19.6	20.1	20.2	19.1	18.5	17.8	19.9	19.7	20.0	20.2	17.1	20.3	0.5																													
[14] *Amynthas meishanensis*	19.0	21.7	19.7	17.3	18.3	20.8	19.4	18.2	19.8	22.0	16.9	20.0	19.8	0.3																												
[15] *Amynthas morrisi*	18.5	17.2	17.1	20.2	17.2	16.7	18.4	17.7	19.3	20.3	16.6	17.7	18.8	19.2	0.5																											
[16] *Amynthas plantoporophoratus*	16.8	17.7	16.5	16.7	15.9	16.1	16.6	15.9	17.7	18.7	17.0	18.4	16.5	17.9	14.4	0.6																										
[17] *Amynthas polyglandularis*	17.5	17.9	17.3	18.5	17.8	17.3	19.0	17.6	20.0	21.2	16.6	19.1	18.8	18.3	18.0	17.4	-																									
[18] *Amynthas pongchii*	17.6	16.0	15.8	18.5	16.7	16.0	17.7	16.4	17.3	18.7	14.5	16.7	17.6	16.9	15.3	16.1	1.6																									
[19] *Amynthas robustus*	15.7	16.1	16.4	17.0	16.1	18.2	14.0	17.1	15.7	18.3	14.0	16.8	19.4	19.4	18.1	14.9	19.5	15.9	0.2																							
[20] *Amynthas tayalis*	17.7	19.4	16.0	17.0	15.4	17.9	17.0	15.0	20.2	21.3	14.7	17.6	18.8	13.5	15.7	16.8	17.3	15.3	18.3	-																						
[21] *Amynthas tuberculatus*	19.8	19.3	17.6	15.9	17.1	17.6	19.3	18.0	20.2	20.7	17.9	19.3	18.4	18.3	17.1	17.9	18.5	17.0	19.6	16.3	0.1																					
[22] *Amynthas wulinensis*	17.8	18.8	16.1	16.9	17.0	16.1	18.0	16.6	18.8	20.0	18.0	17.1	21.0	18.4	17.8	16.8	18.4	16.3	16.1	18.7	4.7																					
[23] *Aporectodea caliginosa*	21.2	22.2	18.5	21.5	17.9	21.1	23.0	17.2	21.7	23.6	18.9	19.9	21.4	19.4	20.4	19.9	21.3	19.2	22.3	17.7	20.6	19.9	-																			
[24] *Aporectodea trapezoides*	18.3	22.2	18.6	19.8	19.4	19.2	19.2	18.9	19.6	21.1	18.7	20.5	20.3	20.0	20.7	19.2	20.3	19.4	19.0	19.6	20.7	19.7	17.0	0																		
[25] *Bimastos parvus*	20.2	22.2	17.7	21.5	19.9	20.7	22.2	19.9	22.2	22.0	20.5	19.2	21.1	20.0	21.4	20.3	22.2	22.5	20.0	20.3	21.7	21.0	16.6	15.8	0																	
[26] *Drawida japonica*	24.0	23.4	24.1	23.7	22.0	23.7	24.7	22.2	23.7	27.4	22.0	23.3	25.0	22.2	24.0	22.8	24.3	23.9	23.4	21.5	25.9	22.9	24.3	23.5	25.2	0																
[27] *Eisenia andrei*	20.2	20.6	20.3	22.5	19.6	22.1	21.4	17.9	22.5	23.1	22.0	22.0	23.2	19.8	22.5	21.3	22.1	22.6	20.9	21.0	22.0	20.9	20.1	17.9	18.2	25.8	0.3															
[28] *Eisenia eiseni*	18.7	20.7	17.7	20.1	18.6	20.2	19.8	18.7	21.0	20.8	19.3	19.0	18.6	19.6	16.6	17.5	20.5	19.5	18.8	17.1	19.7	19.6	16.5	18.0	16.1	23.5	21.1	0.3														
[29] *Eisenia fetida*	21.2	21.7	21.0	19.3	18.6	21.4	22.0	18.6	20.9	23.8	18.4	19.4	21.8	19.4	20.8	21.0	21.3	21.0	21.5	20.7	19.1	20.5	18.3	17.8	19.7	24.9	15.0	18.3	0.2													
[30] *Eisenia* sp.	19.5	19.9	20.5	20.8	19.5	20.9	22.7	20.5	22.3	24.0	18.7	19.5	22.4	20.3	22.0	21.6	19.9	19.8	19.0	20.7	21.5	21.2	18.3	15.9	17.0	25.1	14.7	18.1	11.5	0.1												
[31] *Lumbricus terrestris*	20.1	22.8	21.5	21.3	21.8	20.5	23.7	21.9	21.3	22.9	20.2	22.1	22.4	20.7	20.9	21.9	21.7	21.8	20.9	20.0	18.8	25.4	19.3	20.2	18.5	18.2	-															
[32] *Metaphire bununa*	16.2	18.6	17.1	18.3	17.4	17.1	18.5	15.8	19.6	19.2	16.4	17.2	19.4	17.6	18.1	16.2	16.0	14.3	17.0	16.8	19.5	16.0	19.6	19.4	21.9	25.4	20.2	18.5	20.3	20.1	21.1	-										
[33] *Metaphire californica*	19.4	15.2	18.6	20.0	18.9	17.6	17.5	19.0	16.6	4.8	18.3	20.2	18.7	20.1	18.6	16.1	19.5	16.2	15.7	19.7	18.3	18.6	21.1	19.8	19.3	25.1	21.9	18.7	23.2	22.1	20.8	16.7	1.6									
[34] *Metaphire formosae*	16.4	18.0	16.7	16.3	16.8	16.9	17.4	16.2	17.8	18.3	16.0	17.4	18.7	18.7	17.6	16.4	17.1	16.0	15.0	16.2	18.3	16.6	21.2	19.4	21.6	24.7	20.3	20.1	20.1	19.8	21.8	11.9	16.7	6.8								
[35] *Metaphire paiwanna*	15.8	17.5	15.8	17.1	15.8	17.5	17.1	15.4	18.5	17.8	15.6	15.9	18.5	18.3	17.4	14.7	17.3	14.8	13.7	15.3	18.1	16.9	20.7	19.0	20.7	23.7	21.4	17.6	20.5	21.6	22.4	10.2	16.3	10.5	-							
[36] *Metaphire schmardae*	17.8	21.9	17.3	17.7	18.6	16.6	19.4	17.9	18.5	23.9	16.1	18.1	20.6	17.1	18.8	17.6	17.7	17.4	18.2	17.7	19.8	17.1	20.2	18.3	20.3	22.6	21.0	20.9	19.7	19.2	21.1	16.4	21.2	18.4	19.6	-						
[37] *Metaphire* sp.	17.9	17.1	18.1	17.3	17.3	17.3	17.5	17.3	17.0	18.3	13.3	18.8	17.5	18.6	18.4	15.3	18.5	14.4	14.7	18.6	18.8	17.0	20.7	20.7	20.7	21.7	21.4	19.9	20.3	20.8	22.8	15.8	16.0	15.5	13.9	18.6	0					
[38] *Metaphire tahanmonta*	17.3	19.4	17.5	17.1	18.2	17.7	19.4	18.1	19.4	19.0	16.7	18.0	17.5	19.4	18.5	17.6	16.2	15.6	16.4	17.7	19.7	18.3	20.7	17.5	20.5	25.4	21.0	19.2	20.1	19.9	23.0	10.7	17.0	11.8	10.5	19.2	17.0	-				
[39] *Metaphire trutina*	17.3	19.2	15.4	19.0	16.4	17.1	18.3	15.3	18.8	18.8	16.6	16.8	18.6	17.1	17.3	17.5	14.8	14.8	15.4	20.2	16.9	19.4	18.5	19.4	24.5	20.2	16.0	18.9	18.1	22.0	10.9	16.7	10.9	9.4	18.3	17.3	10.9	-				
[40] *Metaphire tschiliensis*	15.6	17.6	15.9	16.4	14.3	15.9	16.3	13.8	17.2	18.6	14.7	15.2	18.6	14.0	16.6	14.8	17.7	13.6	15.4	20.2	16.9	21.0	23.4	20.1	17.4	20.5	21.3	20.8	10.4	16.6	12.9	12.5	16.9	14.9	13.4	13.3	2.4					
[41] *Metaphire yuhsii*	15.8	18.1	15.6	17.0	16.1	16.6	18.1	15.5	18.1	18.0	15.1	16.5	18.4	17.0	16.5	15.9	16.0	15.1	16.4	15.2	17.2	16.5	19.7	18.4	20.4	24.7	20.0	17.9	18.6	19.7	20.7	10.9	16.3	11.4	11.2	17.4	16.4	11.1	12.2	12.1	5.9	
[42] *Pontodrilus litoralis*	18.6	19.2	19.0	17.9	17.7	18.5	19.0	18.5	18.3	21.7	16.6	19.4	20.2	17.0	16.6	17.5	17.0	18.0	17.5	17.1	19.8	17.9	18.6	18.1	19.8	22.8	19.3	18.9	20.1	17.3	19.8	17.1	18.6	16.7	17.3	17.7	16.2	18.6	16.0	17.1	16.9	-

variation). By contrast, the mean interspecific distance was 18.66% [sd = 2.5], ranging from a low of 4.77% between *M. californica* and *A. kiangensis* to a high of 27.37% between *Drawida japonica* and *A. kiangensis* (22.4%, 25.4% and 25.8% if the within-genus distances in *Amynthas*, *Metaphire* and *Eisenia*, respectively, were considered). Taking into account all intra- and interspecific divergences (Figure 2), the relationship between species distinction and COI divergence was clear—a vast predominance of cases where intraspecific variation was low and interspecific variation high. In terms of amino acid changes, most of the nucleotide substitutions within species were synonymous. Nevertheless, non-synonymous changes occurred within nine taxa (Table 2; Figure 3); these changes were congruent with the topology of the nucleotide-based NJ tree (i.e. matching exclusive groupings) in three cases (*A. wulinensis*, *A. corticis* and *M. tschiliensis*; Figure 3, grey boxes).

Discussion

Taxonomy

The accuracy of species names is the basis of taxonomic and systematic studies. Certainly, the correct use of names is crucial for testing the ability of DNA barcoding as a tool to discriminate species. This fact might sound obvious, but several studies aiming to test DNA barcoding efficiency have neglected this basic requirement and used,

for example, sequences downloaded from GenBank (Memon et al. 2006; Meier et al. 2006). The reliability of the taxonomic information attached to the records in this repository is inconsistent (Harris 2003; Vilgalys 2003) and also suffers from the usual lack of any linkage to accessible vouchers. This drawback led NCBI to adopt a 'DNA barcode' standard tag, identifying those records in GenBank which comply with a number of requisites and in particular the indication of the location of voucher specimens to make them accessible for further examination when necessary.

The study by Huang et al. (2007) included 28 different species of earthworms, but the identification of several species appear problematic. Several

Table 2. Intraspecific changes in amino-acid composition; values for *Amynthas kiangensis* are given within brackets as they are likely caused by sequencing error (see text)

	Protein haplotypes	Amino-acid changes
Amynthas kiangensis	(3)	(4)
Amynthas lini	2	1
Amynthas wulinensis	2[a]	3
Amynthas corticis	2[a]	1
Amynthas hupeiensis	3	1
Amynthas pongchii	3	1
Metaphire tschiliensis	2[a]	1
Metaphire formosae	3	1
Metaphire yuhsii	2	1

[a]Referring to haplotypes matching exclusive clusters of specimens (see also Figure 3)
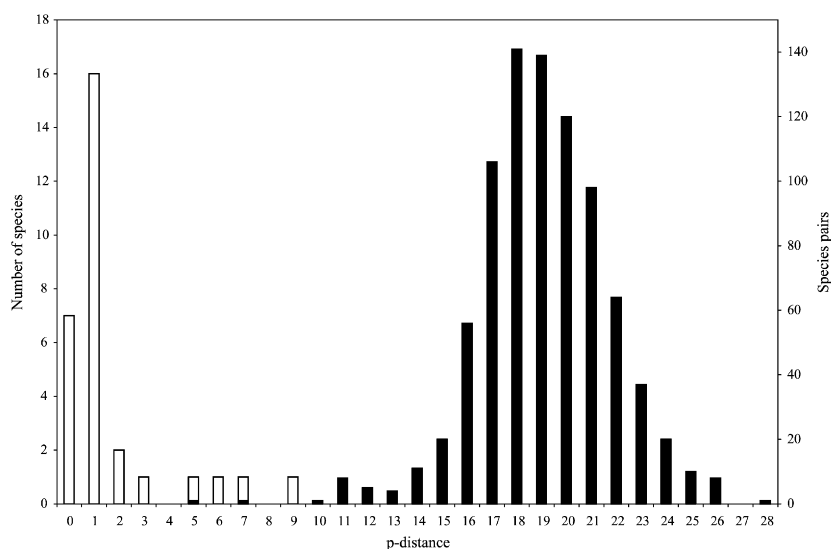
Figure 2. Distribution of intraspecific mean p-distance values (white bars-left *Y*-axis) across the 42 species considered in the combined dataset, and interspecific mean p-distance values (black bars-right *Y*-axis) across the 861 pairwise comparisons for those 42 species (values along the *X*-axis represent the higher value for each interval).
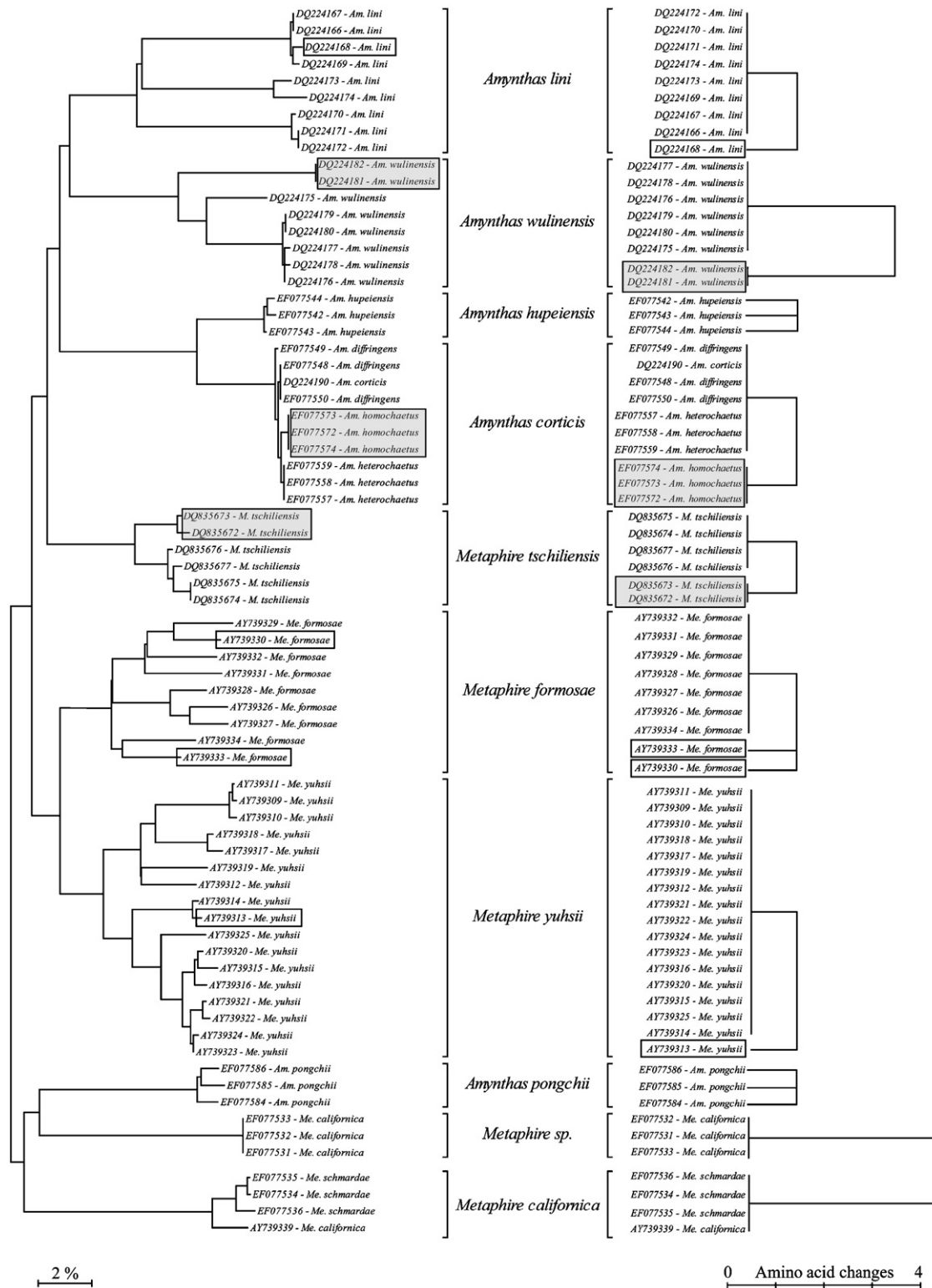
Figure 3. Neighbor Joining trees inferred from nucleotide-based uncorrected p-distance (left) and number of differences in amino acids composition (right) for selected taxa having intraspecific protein haplotypes (the amino acid changes scale is not cumulative and the trees were built independently for each species); boxes highlight identical specimens in both trees, with grey boxes highlighting cases where the topologies are congruent.

Table 3. Summary of the taxonomic issues reported for taxon names used by Huang et al. (2007)

Name used by Huang et al. (2007)	Taxonomic statement
Amynthas homochaetus (Tsai, Shen & Tsai, 2002)	Misspelling and wrong author name for *A. homosetus* (Chen, 1938); misidentification of *A. corticis*[a]
A. plantoporophoratus (Tsai, 1964)	Wrong author name for *A. plantoporophoratus* (Thai, 1984)
A. diffringens	Junior synonym of *A. corticis*[b]
A. heterochaetus	Junior synonym of *A. corticis*[b]
A. hawayanus	Junior synonym of *A. gracilis*[b]
A. lautus	Junior synonym of *A. robustus*[b]
A. benignus	Misidentification of *A. robustus*[a]
Metaphire schmardae	Misidentification of *M. californica*[a]
M. californica	Misidentification of an unknown *Metaphire* species[a]
Eisenia fetida #EF077596	Unknown *Eisenia* species[a]

[a]See discussion.
[b]See Blakemore (2004, 2006a, b).

inconsistencies and doubtful identifications became apparent, both from the available taxonomic information on these organisms and from combining their sequences with other data. From the 28 names used by Huang et al. (2007), 10 need reconsideration (Table 3). For instance, specimens identified by Huang et al. (2007) as *Amynthas heterochaetus*, *A. diffringens* and *A. homochaetus* belong to a single cluster (Figure 1, #30; mean p-distance = 0.3%) and were grouped with another specimen of *A. corticis* from a previously published dataset (Chang and Chen 2005). Significantly, both *A. heterochaetus* and *A. diffringens* were concluded to be junior synonyms of *A. corticis* by Blakemore (2002, 2004). Hence, it is reasonable to assume that all the specimens within this cluster belong to a single species, *A. corticis*, and further, that the specimens of *A. homochaetus* in Huang et al. (2007) are possibly misidentified *A. corticis*. Interestingly, however, the three specimens of *A. homochaetus* differed from all others in this cluster (Figures 1 and 3) by a single amino-acid substitution (an asparagine for a serine) in position 18 of the COI fragment (a single nucleotide substitution in position 53 of the gene fragment). Although this difference cannot be considered as indicating species distinctness, it does suggest the need for

a re-examination of these specimens. *A. robustus* seems to represent another case of misidentification or use of synonyms by Huang et al. (2007); their specimens of *A. lautus* and *A. benignus* cluster very tightly with specimens of *A. robustus* from the same study as well as from the dataset of Chang et al. (2007) (Figure 1, #16, mean p-distance = 0.17%). *A. lautus* has long been considered a junior synonym of *A. robustus* (Blakemore 2002, 2004), but as the validity of *A. benignus* has not yet been questioned by taxonomists, the results for that species either reveal a misidentification by Huang et al. (2007), a new synonymy, or less likely, the first case of different earthworm species sharing a common COI haplotype. A careful re-examination of these specimens is required to resolve this matter.

A few other cases of misidentification (Table 3) are likely to have arisen from the combination of datasets. Firstly, specimens identified by Huang et al. (2007) as *Metaphire schmardae* could not be classified with *M. schmardae* from Chang and Chen (2005), but rather with *M. californica* (Figure 3). Secondly, specimens identified as *M. californica* are distant (p-distance = 15.96%, four amino-acid substitutions) from specimens of this species in Chang and Chen (2005) and are thought to belong to an undetermined species of this genus, named herein as *Metaphire* sp. (Figures 1 and 3). Moreover, Blakemore (2006a, b) recently hypothesized that *A. kiangensis* is a junior synonym of *M. californica*. Our results suggest that these two species are closely allied, as they displayed the lowest interspecific p-distance (4.77%) which was far below the mean interspecific value of 18.66%. However, this case needs further examination. Furthermore, we found that the sequences of *A. kiangensis* obtained by Huang et al. (2007) are questionable because 10 of the 12 amino-acid substitutions separating this species from *M. californica* occured among the first 13 amino-acid positions. This result suggests a sequencing error, and by removing this doubtful portion of the gene up to nucleotide position 42, the p-distance between the two taxa dropped to 2.38%, and only two amino-acid differences remained. Thirdly, specimens identified as *A. triastriatus* and *A. hawayanus* (a junior synonym of *A. gracilis*) appeared to be conspecific (mean p-distance = 0.45%; Figure 1, #18). As it is impossible to determine if these specimens actually belong to *A. triastriatus* or to *A. gracilis*, we retained the identification of *A. triastriatus* proposed by Huang et al. (2007). Fourthly, the divergent specimen of *Eisenia fetida* reported in Huang et al. (2007) was also strongly distinct from *E. fetida* specimens from Pérez-Losada et al. (2005) (mean

p-distance = 11.52%; Figure 1, #5 and #6); it was grouped with two specimens of an unknown *Eisenia* species already mentioned by Pérez-Losada et al. (2005) (mean p-distance = 0.13%), and both these authors and Huang et al. (2007) suggested it represent an unrecognized sibling taxon.

Sequence divergence in earthworms

In light of these taxonomic problems, the statements by Huang et al. (2007) about the genetic variation within and between species of earthworms need reconsideration. It is surprising that these authors reported an interspecific distance higher than 15.8% in all cases, given the cases of overlooked synonymies mentioned previously. When including Huang et al. (2007)'s sequences of *Amynthas lautus*, *A. benignus*, *A. robustus*, *A. homosetus*, *A. heterochaetus*, and *A. diffringens* in the analysis as different species, the interspecific distance was as low as 0.19% (between *A. benignus* and *A. lautus*, and between *A. benignus* and *A. robustus*). These results conflict with Huang et al. (2007)'s conclusion that intra- and interspecific distances in earthworms show no overlap and they ''*were able to successfully discriminate known species of Chinese earthworms using COI sequences in all cases*''. By combining the dataset of Huang et al. (2007) with other published datasets, and by re-considering the taxonomic identifications proposed by these authors, we have provided a more accurate estimation of the extent to which the COI DNA barcode enables species discrimination in earthworms. Our results (Figures 1 and 2) show that the vast majority of species analyzed to date display a bimodal pattern in which interspecific distance is much higher than intraspecific divergence. This is in agreement with Huang et al. (2007)'s conclusion regarding the general applicability and strong potential of DNA barcoding as a new tool in earthworm taxonomy.

However, our work has revealed a few cases, some already mentioned in the literature (Chang and Chen 2005; Chang et al. 2007), in which intraspecific variability is significantly higher than usual. The deep intraspecific divergence in the Taiwanese endemics *A. lini*, *A. wulinensis*, *Metaphire formosae* and *M. yuhsii* requires further study. It is noteworthy that these four cases (as well as six other Taiwanese taxa, unpublished data) represent taxa for which the most extensive geographical sampling was available. Although we may discover similar cases in other species with further sampling, these deep divergences are likely

related to the history of earthworm diversification and speciation in Taiwan. Indeed, the phylogeographic study by Chang and Chen (2005) suggested that the genetic diversity observed within the sexually reproducing species *M. formosae* and *M. yuhsii* (Figure 1, #37 and #42 respectively) originated either after the formation of some volcano groups in the northern part of the island, or with the formation of the ancient Tamsui River. Both events were considered as possible barriers to gene flow for these organisms, suggesting cryptic diversity is a possible interpretation of the observed genetic pattern rather than incomplete lineage sorting. Nevertheless, applying a taxonomic framework to such complex cases of recent diversification is challenging, and only a thorough phylogeographic study including additional independent markers (e.g. a nuclear gene in the absence of morphological variations) can rule out the occurrence of ancestral polymorphism and assess the level of gene flow between the different populations as represented to date by their different COI haplotypes. Differences in the amino acid sequence of COI between individuals considered as belonging to the same species are uncommon, suggesting a strong selective constraint on the enzyme composition and structure. In a few instances however, intraspecific non-synonymous mutations were observed (Figure 3; Table 2). Whereas these replacements are likely to represent individual polymorphism when they are linked to shallow genetic divergence and not to any geographic structure of the population (as in *Amynthas lini*, see Figure 3), their existence suggests strong and persistent genetic segregation when they are multiple and associated to deep genetic divergence. For instance, in *A. wulinensis*, two individuals (specimens #DQ224181 and #DQ224182) were 8% divergent from other conspecifics and differ from them by three amino acid substitutions; this suggests a long genetic isolation permitting fixation of these non-synonymous mutations. Further studies are needed to assess if that isolation persists today or if the observed polymorphism is ancestral; the current geographic isolation of this population (Chang et al. 2007) favors the first hypothesis.

Conclusion

Our results generally support the statement by Huang et al. (2007) that the use of DNA barcoding in earthworm taxonomy would provide a very significant advance toward both a better understanding

and a faster description of the diversity of this group. The two species *Eisenia fetida* and *E. andrei*, for example, represent a famous documented case of species having been widely confounded in the literature. That error led to erroneous identification of model organisms used in ecotoxicological studies and organic waste management (Dominguez et al. 2005; Pérez-Losada et al. 2005; Chang et al. 2007). Interestingly, distinction between these two species was assessed unambiguously from both a remarkable experimental study of reproductive isolation by Dominguez et al. (2005) and DNA studies by Pérez-Losada et al. (2005) (the p-distance between these two species as calculated in the present paper was as high as 15%). This result places DNA barcoding as a very promising surrogate to the long and difficult experimental breeding often necessary to assess the status of sibling taxa. Moreover, it has to be added that DNA barcoding can be applied to any lifestage of these organisms, and as such represents a considerable promise for earthworm taxonomy since identification of juveniles from morphology is currently impossible in most cases. At a time where environmental sequencing (Hajibabaei et al. 2007) lets one envision large-scale surveys of the fauna in any substrate sample, the assessment of the potential of a DNA identification tag for earthworms, a group dominating the soil biota in biomass and ecological services (Lavelle et al. 2006), is fundamental. By critically analyzing the results and conclusions of Huang et al. (2007), we aimed to improve the general understanding of COI divergence patterns in earthworms, so that the potential use of a portion of this gene as a DNA barcode can be objectively assessed, considering both its effectiveness and its limitations, and thus identifying the areas that remain to be explored. We demonstrated that the large ''barcode gap'' suggested by Huang et al. (2007) could not be generalized in earthworms, and did not even apply to their own dataset. However, by combining available evidence and by considering both sequence distances and reciprocal monophyly of taxa, our results also clearly demonstrated that the genetic analysis matches the species delineations inferred from traditional approaches in all cases, although evidence of overlooked species was revealed. We hope these results will orientate future research on earthworm diversity toward answering some pending questions relative to the use of DNA barcodes. Firstly, do deep intraspecific divergences such as the ones in *Metaphire formosae*, *M. yuhsii* and *Amynthas lini* represent ancestral polymorphisms or geographic differentiation of isolated populations? Secondly, are there cases in earthworms where different species show shallow divergence, as perhaps in *M. tschiliensis* or *A. corticis*? We conclude that further exploration of mitochondrial diversity in earthworms will lead to major improvements in our understanding of the evolutionary pathways and rates of the mitochondrial genome. Earthworms are highly diverse in term of life history, and it can be expected that investigations of COI variation patterns within and between endemic and invasive species, as well as between sexually or parthenogenetically reproducing populations, will reveal how population dynamic parameters affect the evolution of the mitochondrial genome and the informativeness of its genes as indicators of species distinctness.

Acknowledgements

References

Blakemore, R.J., 2002. Cosmopolitan earthworms–an eco-taxonomic guide to the Peregrine species of the World. (First CD Edition). VermEcology, PO BOX 414 Kippax, ACT 2615, Australia.

Blakemore, R.J., 2004. Checklist of pheretimoid earthworms after Sims & Easton, 1972. In: Moreno, A., Borges, S. (Eds.), Avances en Taxonomia de Lombrices de Tierra/Advances in Earthworm Taxonomy (Annelida: Oligochaeta). Editorial Complutense. Universidad Complutense, Madrid, Spain, pp. 126–154.

Blakemore, R.J., 2006a. A series of searchable texts on earthworm biodiversity, ecology and systematics from various regions of the world–2nd Edition (2006). In: Kaneko, N., Ito, M.T., (Eds.), COE Soil Ecology Research Group, Yokohama National University, Japan. CD-ROM.

Blakemore, R.J., 2006b. A series of searchable texts on earthworm biodiversity, ecology and systematics from various regions of the world-2nd Edition Supplement. In: Kaneko, N., Ito., M.T., (Eds.), COE Soil Ecology Research Group, Yokohama National University, Japan. CD-ROM. Online: ⟨http://www.bio-eco.eis. ynu.ac.jp/eng/database/earthworm/⟩.

Chang, C.-H., Chen, J.-H., 2005. Taxonomic status and intraspecific phylogeography of two sibling species of *Metaphire* (Oligochaeta: Megascolecidae) in Taiwan. Pedobiologia 49, 591–600.

Chang, C.-H., Lin, Y.-H., Chen, I.-H., Chuang, S.-C., Chen, J.-H., 2007. Taxonomic re-evaluation of the Taiwanese montane earthworm *Amynthas wulinensis* Tsai, Shen & Tsai, 2001 (Oligochaeta: Megascolecidae):

polytypic species or species complex? Org. Divers. Evol. 7, 231–240.

Dominguez, J., Velando, A., Ferreiro, A., 2005. Are *Eisenia fetida* (Savigny, 1826) and *Eisenia andrei* (Oligochaeta, Lumbricidae) different biological species? Pedobiologia 49, 81–87.

Hajibabaei, M., Singer, G., Clare, E., Hebert, P., 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. BMC Biology 5.

Harris, J.D., 2003. Can you bank on GenBank? Trends Ecol. Evol. 18, 317–319.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. B. 270, 313–321.

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003b. Barcoding anomal life: cytochrome c oxidase subunit 1divergences among closely related species. Proc. R. Soc. B. 270 (Suppl.), S96–S99.

Huang, J., Sun, Z.J., Tang, G.L., Su, Z.Y., 2007. Identifying earthworms through DNA barcodes. Pedobiologia 51, 301–309.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics 5, 150–163.

Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., Rossi, J.P., 2006. Soil invertebrates and ecosystem services. Eur. J. Soil Biol. 42, S3–S15.

Meier, R., Shiyang, K., Vaidya, G., Ng, P.K.L., 2006. DNA barcoding and taxonomy of Diptera: a tale of high intraspecific variability and low identification success. Systematic Biol. 55, 715–728.

Memon, N., Meier, R., Manan, A., Feng-Yi, Su.K., 2006. On the use of DNA sequences for determining the species limits of a polymorphic new species in the stink bug genus *Halys* (Heteroptera: Pentatomidae) from Pakistan. Syst. Entomol. 31, 703–710.

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

Swofford, D.L., 2000. PAUP 4.0: Phylogenetic Analysis Using Parsimony (and Other Methods). Sinauer, Sunderland.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Vilgalys, R., 2003. Taxonomic misidentification in public DNA databases. New Phytol. 160, 4–5.