Taiwania, **50**(2): 71-79, 2005

NADH Dehydrogenase Subunit 1 Gene of the Earthworm *Amynthas gracilis* (Kinberg, 1867) (Oligochaeta: Megascolecidae), with the Discussion on Inferring the Megascolecid Phylogeny Using DNA Sequences

Chih-Han Chang[(1)], Shu-Chun Chuang[(2)], Yun-Ru Chen[(2)] and Jiun-Hong Chen[(1,2,3)]

(Manuscript received 18 January, 2005; accepted 25 March, 2005)

ABSTRACT: The complete sequence of the mitochondrial gene encoding NADH dehydrogenase subunit 1 (ND1) from a megascolecid earthworm, *Amynthas gracilis* (Kinberg, 1867), is published in this study. The gene is 923 bp in size, encoding a polypeptide of 307 amino acid residues. Its nucleotide sequence and amino acid sequence are 71.8% and 77.7% identical to those of the European earthworm *Lumbricus terrestris* ND1 gene, respectively. In addition, the GC contents among some published annelid ND1 full sequences were compared, and the relationships among these sequences were reconstructed using a phylogenetic analysis. Accordingly, we believe that a 3000-bp sequence comprising the ND1 and the CO1 genes and the genes used by Jamieson *et al.* (2002) can be a good candidate to be used in the phylogenetic analysis to unravel the phylogeny of the Megascolecidae.

KEY WORDS: Megascolecidae, *Amynthas gracilis*, ND1, phylogeny.

INTRODUCTION

NADH dehydrogenase, catalyzing the electron transfer from NADH within the respiratory chain in mitochondria, has been analyzed from various animal taxa (Boore and Brown, 1995; Leys *et al*., 2003; Li and Liao, 2003). In Annelida, the complete sequence of the first subunit, namely NADH dehydrogenase subunit 1 (ND1) of mitochondria, has only been published in one earthworm (*Lumbricus terrestris*), one leech (*Helobdella robusta*), and four polychaetes (*Platynereis dumerilii*, *Galathealinum brachiosum*, *Riftia pachyptila*, and *Clymenella torquata*) (Boore and Brown, 1995, 2000; Boore, 2001; Jennings and Halanych, 2005). Being native to Southeast and East Asia, Australia, South America, and Ethiopia, the Megascolecidae is the largest earthworm family with more than 1000 species. However, regardless of the importance and abundance of the Megascolecidae in the worldwide earthworm fauna, relatively few megascolecid gene sequences are available in GenBank: most of the commonly used genes including the ND1 in phylogenetic analyses have not been identified yet. In this paper, we reported the full-length DNA sequence of ND1 from the megascolecid earthworm *Amynthas gracilis* (Kinberg, 1867), a cosmopolitan species distributed from the Hawaiian Islands to East Asia. The DNA sequence and the deduced amino acid sequence were compared with those of the earthworm *L. terrestris* of the Lumbricidae. We also discussed the potential usage of ND1 in studying the phylogeny of the megascolecid earthworms.

1. Institute of Zoology, National Taiwan University, No. 1 Roosevelt Road, Sec. 4, Taipei 106, Taiwan.
2. Department of Life Science, National Taiwan University, No. 1 Roosevelt Road, Sec. 4, Taipei 106, Taiwan.
3. Corresponding author. Tel: 886-2-23630231 ext. 2354; Fax: 886-2-23658912; Email: chenjh@ntu.edu.tw.

MATERIALS AND METHODS

Sample collection, DNA extraction, PCR, and sequencing

The earthworm *A. gracilis* was collected from the agricultural land in the main campus of the National Taiwan University and anesthetized in a 10% ethanol solution for further DNA extraction. Muscle tissue from a fresh specimen was homogenized in liquid nitrogen and then digested in digestion buffer (10 mM Tris-HCl, 2 mM dihydrate EDTA, 10 mM NaCl, 10 mg/ml DTT, 1% SDS, and 0.4 mg/ml proteinase K) at 50 °C for 15 min. Total DNA was extracted from the digested tissue-buffer solution with a standard phenol/chloroform extraction method (Palumbi *et al.*, 1991).

Four oligonucleotide primers designed in this study were used for amplification of the ND1 fragment, including 16SND1 (5'-CTA CAT GAG CTG AGT TCA GAC CG-3'), ND1R (5'-TTT ACA AAT AGT TTA ATG GC-3'), ND1F (5'-TTT ACA TCC GTA TTA CTA AGA TT-3'), and ND3ND1 (5'-TGA GAA TGG GAT CCG TGC-3'). The primers 16SND1 and ND3ND1 were designed with referred to the published 16S rRNA (for 16SND1) and ND3 (for ND3ND1) sequences of the earthworm *L. terrestris* and the polychaete *P. dumerilii* (under GenBank accession nos. AF178678 and U24570 for each species, respectively); the primers ND1F and ND1R were designed with referred to the published *L. terrestris* ND1 sequence and our preliminary result of the partial *A. gracilis* ND1 sequence (unpublished data). The primer pair 16SND1 and ND1R amplified the 3' of 16S rRNA, tRNA-Leu, tRNA-Ala, tRNA-Ser, tRNA-Leu, and the 5' of ND1; the other pair ND1F and ND3ND1 amplified the 3' of ND1, tRNA-Ile, tRNA-Lys, and the 5' of ND3. A 1551-bp mitochondrial DNA fragment composed of two overlapping fragments was amplified. PCR amplifications were carried out in 50-μl total volumes using one cycle of 94 °C for 1 min, 35 cycles of denaturation for 40 s at 94 °C, annealing for 30 s at 50 °C, and extension for 1 min at 72 °C, and one cycle at 72 °C for 10 min for flanking primers 16SND1 and ND1R; one cycle at 94 °C for 1 min, 35 cycles of denaturation for 1 min at 94 °C, annealing for 45 s at 46 °C, and extension for 2 min at 72 °C, and one cycle at 72 °C for 10 min were used for flanking primers ND1F and ND3ND1. PCR products were sequenced in both directions using the same primers as for the PCR amplification. Sequencing was performed with the Dyenamic ET dye terminator cycle sequencing kit (Amersham Biosciences). Products were analyzed using a MegaBACE 500 automated sequencer (Amersham Biosciences).

Sequence analyses

The 923-bp full sequence of the *A. gracilis* ND1 gene was obtained by aligning the 1551-bp sequence with the ND1 full sequence of *L. terrestris* using the default settings of Clustal W 1.82 (Thompson *et al.*, 1994); this sequence was then translated to the amino acid sequence using the standard invertebrate mtDNA translation table (Boore and Brown, 1995). The obtained sequences, both nucleotide and amino acid, were compared with those of *L. terrestris* to confirm their identities and are available in GenBank (under accession no. AY805986).

The published ND1 full sequences of annelid species, including two polychaetes *P. dumerilii* and *G. brachiosum* (pogonophoran), one earthworm *L. terrestris*, and one leech *H. robusta*, were retrieved from GenBank (under accession nos. AF178678, AF178679, U24570, and AF178680 for each species respectively) and used together with the *A. gracilis* ND1 full sequence in the analyses. The GC contents of these sequences were compared and the

relationships of them were reconstructed. Because the DNA sequences were difficult to align, the deduced amino acid sequences were used in the phylogenetic analysis. In addition, two sequences of mollusks, one from the gastropod *Pupa strigosa* and the other from the cephalopod *Loligo bleekeri*, retrieved from GenBank (under accession numbers AB028237 and AB029616 respectively) were used as outgroups. The sequence alignment was performed using the default settings of Clustal W (Thompson *et al.*, 1994). Phylogenetic analyses were carried out using neighbor-joining (NJ) analyses using the Poisson correction implemented in MEGA 2.1 (Kumar *et al*., 2001). Gaps in the amino acid sequences were treated as missing data. Bootstrapping of 1000 pseudo-replicates was used to examine the robustness of lineages.

Reanalysis of published megascolecid 16S rRNA sequences

The aligned magascolecid 16S rRNA sequences reported by Jamieson *et al.* (2002) were acquired from http://www.ncbi.nlm.nih.gov/entrez/batchseq.cgi?db=popset&view=ps&val=19851832. Neighbor joining (NJ) analyses were performed using MEGA 2.1 with Kimura's two-parameter model (Kimura, 1980). Gaps in the DNA sequences were treated as missing data. Bootstrapping of 1000 pseudo-replicates was used to examine the robustness of clades and their phylogenetic relationships.

RESULTS AND DISCUSSION

The coding region of *A. gracilis* ND1 gene is 923 bp in size, encoding 307 amino acids, and the TAA stop codon is completed by the addition of 3' A residues to the mRNA (Boore and Brown, 1995) (Fig. 1). The nucleotide sequence and amino acid sequence of *A. gracilis* were 71.8% and 77.7% identical to those of *L. terrestris,* respectively. Two hundred and sixty variable sites were found between *L. terrestris* and *A. gracilis*, including 131 transitions and 129 transversions.

The nucleotide compositions of the annelid taxa are listed in Table 1. Among invertebrate mitochondrial genomes, guanine (G) and cytosine (C) are the two less-frequently found nucleotides, and their GC contents range from 23.3% to 40.4% (Wolstenholme, 1992). The GC contents of the ND1 genes vary among annelid taxa: *P. dumerilii* has a GC content of 34.1%; *G. brachiosum*, 21.2%; *L. terrestris*, 39.1%; *H. robusta*, 31.1%. The GC content of the *A. gracilis* ND1 gene, 33.1%, is slightly lower than that of the earthworm *L. terrestris*, but is higher than that of the pogonophoran *G. brachiosum*. The phylogenetic relationship of the annelid ND1 genes based on amino acid sequences (Fig. 2) revealed the commonly accepted relationships among annelid taxa, with both earthworms (*L. terrestris* and *A. gracilis*) and clitellate (earthworms and *H. robusta*) being monophyletic groups (with bootstrap values of 100 and 99 for each taxon respectively) and polychaetes (*P. dumerilii* and *G. brachiosum*) as their sister group.

In phylogenetic studies, due to the variable secondary structures of rRNA, the ambiguously aligned regions are encountered more often in the aligned rRNA gene sequences than in the aligned protein-coding gene sequences within the mitochondrial genomes. Meanwhile, the elimination of ambiguously aligned regions shortens the available sequences for analysis (Blaxter, 2004). In addition, aligning of rRNA gene sequences needs referring to correct secondary structures, which may not always be available. Therefore, within mitochondrial genomes, DNA sequences of protein-coding genes are usually more convenient than those of rRNA genes for phylogenetic tree reconstruction. Many mitochondrial protein

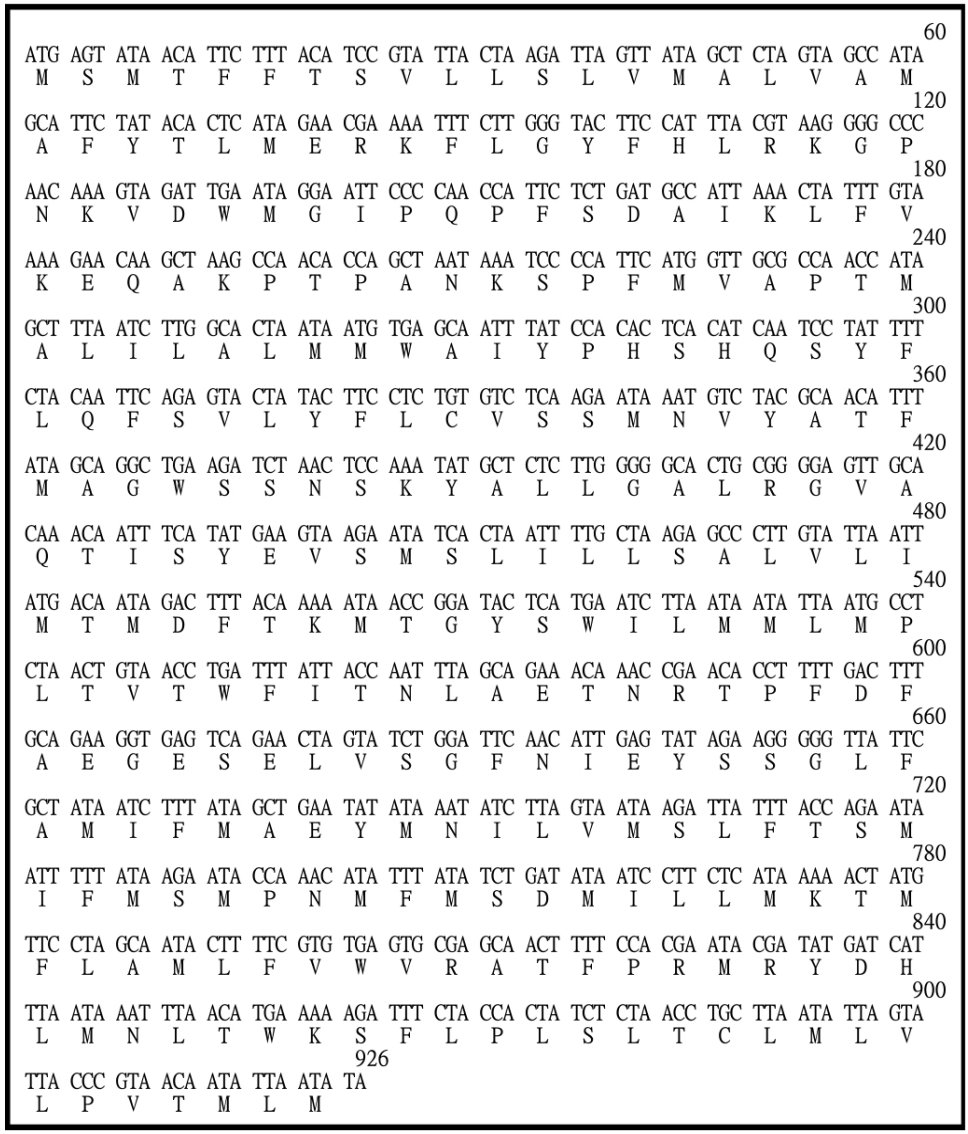

Fig. 1. The nucleotide and deduced amino acid sequences of NADH dehydrogenase subunit 1 of the earthworm *Amynthas gracilis*. Nucleotide sequence numbers are shown at the right. The deduced amino acid residues are numbered beginning with the first Met residue and are shown under the sequence. The TAA stop codon is completed by the addition of 3' A residues to the mRNA (Boore and Brown, 1995).

coding genes, such as CO1 (cytochrome *c* oxidase subunit 1), CO2, ND1, ND2, ND5, ND6, ATPase 6, and Cytb (cytochrome *b*), have been extensively used in invertebrate phylogenetic studies (Fukami *et al*., 2000; Leys *et al*., 2003; Sanchez *et al*., 2003; Heethoff *et al*., 2004; Kato and Yagi, 2004; Pop *et al*., 2003). So far, only CO1 and CO2 have been used to study earthworm phylogeny (Pop *et al*., 2003; Heethoff *et al*., 2004; Chang and Chen, 2005), and only CO1 has been applied to the phylogenetic study of the megascolecid earthworms (Chang and Chen, 2005). On the other hand, the ND1 sequences have been used in phylogenetic studies of Arthropoda (Leys *et al*., 2003), Mollusca (Serb *et al*., 2003), Platyhelminthes (Li and Liao, 2003), and Nematoda (van Herwerden *et al*., 2000), but not in those of Annelida yet. Since using the ND1 genes is getting popular in invertebrate phylogenetics, we believe that this gene should be useful for the study of earthworm phylogeny.

Table 1. Comparison of nucleotide compositions among the ND1 genes of annelid taxa.

species	A (%)	C (%)	G (%)	T (%)
Lunbricus terrestris	27.9	22.3	16.8	33.0
Amynthas gracilis	35.0	19.2	13.9	31.9
Helobdella robusta	33.5	18.9	12.2	35.4
Platynereis dumerilii	32.8	20.5	13.6	33.1
Galathealinum brachiosum	31.2	12.9	8.3	47.6

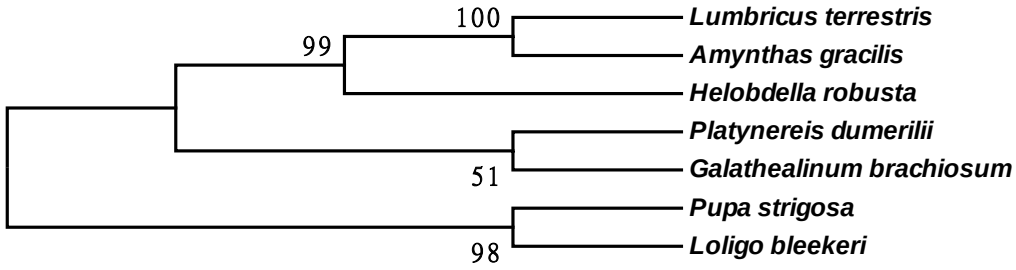

Fig. 2. The cladogram of the relationships among published full-length amino acid sequences of the annelid ND1 genes. The tree was constructed using Kimura's two-parameter model using Poisson correction. Two mollusks, *Pupa strigosa* and *Loligo bleekeri*, are used as the outgroups in this analysis. The numbers around nodes are bootstrap values.

The systematics and phylogeny of the Megascolecidae have created much debate over the past 70 years. Recently, a 1372-bp DNA sequence including nuclear 28S rRNA and mitochondrial 12S and 16S rRNA gene sequences obtained from 32 in-group species was used for a phylogenetic study of the megascolecid earthworms (Jamieson *et al.,* 2002). However, although the monophyly of the Megascolecidae was supported, no acceptable intergeneric phylogenetic structure for the Megascolecidae was revealed. In addition, the inferred phylogeny reconstructed using maximum likelihood and maximum parsimony analyses was illustrated as consensus cladograms, from which the relative branch lengths are not available. Therefore, we reanalyzed the 16S rRNA data reported in that study using a different approach, the neighbor-joining method (Fig. 3). In this new analysis, the inferred phylogenetic tree is characterized by short interior branches and long external ones, with the basal relationships among species of the Megascolecidae poorly resolved. This topology suggests that a rapid and explosive radiation occurred early in the history of these taxa (Maekawa *et al.*, 2001; Su *et al.*, 2001). During that short period, only a few substitutions accumulated and hence only a few phylogenetic informative sites in the DNA sequences are available, leading to the unresolved basal relationships with short interior branches. Apparently, the performance may be improved by adding taxa or characters (sequences). Adding taxa is undoubtedly necessary because only one or two species for most genera were included in the analyses. These sample sizes are so small that the species can not represent for the whole genera that they belonging to, especially when the monophyly of some genera is doubtful. On the other hand, in the discussion on adding taxa or characters in phylogenetic analyses, Graybeal (1998) stated that taxa were added specifically to break up long branches, and it would be entirely misleading to claim that more taxa always improve phylogenetic accuracy. In addition, including longer sequences will increase the number of phylogenetic informative sites, which provides more information of substitutions occurred during that short period of radiation. Consequently, adding more taxa in analyses may be useful to improve accuracy, and including more phylogenetic informative characters, i.e., using longer sequences, can solve the basal relationships. Both are necessary.

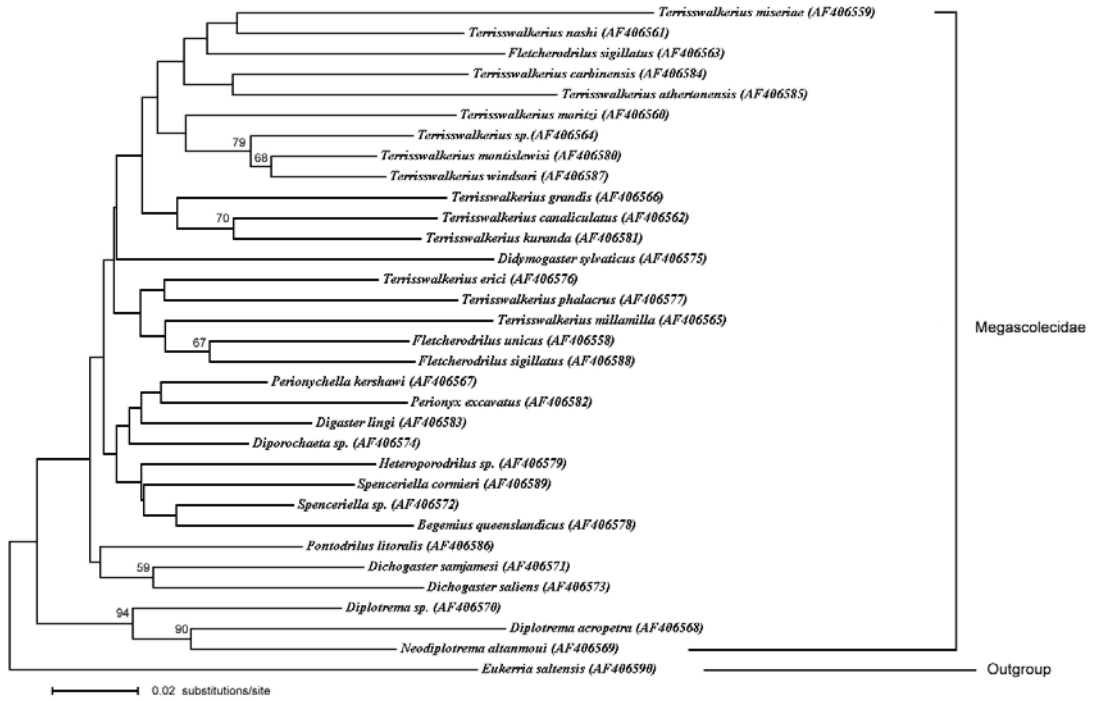

Fig. 3. Neighbor-joining tree of the Megascolecidae reconstructed using the aligned 16S rRNA sequences reported by Jamieson *et al.* (2002). Bootstrap values ≥50 are shown around nodes and the GenBank accession numbers are shown after the species names. The tree has short interior branches and long external ones with unresolved basal relationships, suggesting the occurrence of a rapid and explosive radiation early in the history of these taxa.

For the purpose of including longer sequences, given a total length of about 1600 bp, the combined data set of the complete ND1 sequences and the partial CO1 sequences amplified by the primer pair, LCO1490 and HCO2198 (Folmer *et al.*, 1994), can be a great candidate to be combined with the previous 1372-bp sequence reported by Jamieson *et al.* (2002). The combination results in longer sequences of about 3000 bp. We strongly believe that these sequences can offer a new chance to unravel the debate concerning megascolecid phylogeny. The report of the megascolecid ND1 gene in this study shall help to progress this purpose in the future.

LITERATURE CITED

Blaxter, M. L. 2004. The promise of a DNA taxonomy. Phil. Trans. R. Soc. Lond. B **359**: 669-679.

Boore, J. L. 2001. Complete mitochondria genome sequence of polychaete Annelid *Platynerelis dumerilii*. Mol. Biol. Evol. **18**: 1413-1416.

Boore, J. L. and W. M. Brown. 1995. Complete sequence of the mitochondria DNA of the Annelid worm *Lumbricius terrestris*. Genetics **140**: 305-319.

Boore, J. L. and W. M. Brown. 2000. Mitochondrial genomes of *Galathealinum, Helobdella,* and *Platynereis*: sequence and gene arrangement comparisons indicate that pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol. Biol. Evol. **17**: 87-106.

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

Folmer, O., M. Back, W. Hoeh, R. Lutz and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotech **3**: 294-299.

Fukami, H., M. Omori and M. Hatta. 2000. Phylogenetic relationships in the coral family Acroporidae, reassessed by inference from mitochondrial genes. Zool. Sci. (Tokyo) **17**: 689-696.

Heethoff, M., K. Etzold and S. Scheu. 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.

Graybeal, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. **47**: 9-17.

Jamieson, B. G. M., S. Tillier, A. Tillier, J. L. Justine, E. Ling, S. James, K. McDonald and A. F. Hugall. 2002. Phylogeny of the Megascolecidae and Crassiclitellata (Annelida, Oligochaeta): combined versus partitioned analysis using nuclear (28S) and mitochondrial (12S, 16S) rDNA. Zoosystema **24**: 707-734.

Jennings, R. M. and K. M. Halanych. 2005. Mitochondrial Genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): Evidence for Conserved Gene Order in Annelida. Mol. Biol. Evol. **22**: 210-222.

Kato, Y. and T. Yagi. 2004. Biogeography of the subspecies of *Parides* (*Byasa*) *alcinous* (Lepidoptera: Papilionidae) based on a phylogenetic analysis of mitochondrial ND5 sequences. Syst. Entomol. **29**: 1-9.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequence. J. Mol. Evol. **16**: 111-120.

Kumar, S., K. Tamura, I. B. Jakobsen and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics **17**: 1244-1245.

Leys, R., C. H. S. Watts, S. J. B. Cooper and W. F. Humphreys. 2003. Evolution of subterranean diving beetles (Coleoptera: Dytiscidae: Hydroporini, Bidessini) in the arid zone of Australia. Evolution **57**: 2819-2834.

Li, J. and X. Liao. 2003. The taxonomic status of *Digramma* (Pseudophyllidea: Ligulidae) inferred from DNA sequences. J. Parasitol. **89**: 792-799.

Mawkawa, K., M. Kon, K. Araya and T. Matsumoto. 2001. Phylogeny and biogeography of wood-feeding cockroaches, genus *Salganea* Stal (Blaberidae: Panesthiinae), in Southeast Asia based on mitochondrial DNA sequences. J. Mol. Evol. **53**: 651-659.

Palumbi, S., R. A. Martin, S. Romano, W. O. McMillan, L. Stice and G. Grabowski. 1991. The simple fool's guide to PCR, Version 2. University of Hawaii Zoology Department, Honolulu, HI, USA. 46 pp.

Pop, A. A., M. Wink and V. V. Pop. 2003. Use of 18S, 16S rDNA and cytochrome c oxidase sequences in earthworm taxonomy (Oligochaeta, Lumbricidae). Pedobiologia **47**: 428-433.

Sanchez, J. A., C. S. McFadden, S. C. France and H. R. Lasker. 2003. Molecular phylogenetic analyses of shallow-water Caribbean octocorals. Mar. Biol. (Berlin) **142**: 975-987.

Serb, J. M., J. E. Buhay and C. Lydeard. 2003. Molecular systematics of the North American freshwater bivalve genus *Quadrula* (Unionidae: Ambleminae) based on mitochondrial ND1 sequences. Mol. Phylogenet. Evol. **28**: 1-11.

Su, Z., Y. Imura and S. Osawa. 2001. Evolutionary discontinuity of the Carabine ground beetles. J. Mol. Evol. **53**: 517-529.

Thompson, J. D., D. G. Higgins and P. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequencing weighting, position-specific gap penalties and weight matrix choice. Nucleic Acid Res. **22**: 4673-4680.

van Herwerden, L., D. Blair and T. Agatsuma. 2000. Multiple lineages of the mitochondrial gene NADH dehydrogenase subunit 1 (ND1) in parasitic helminths: implications for molecular evolutionary studies of facultatively anaerobic eukaryotes. J. Mol. Evol. **51**: 339-352.

Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cyt. **141**: 173-216.

巨蚓科蚯蚓纖細遠環蚓 *Amynthas gracilis* (Kinberg, 1867)的 NADH 去氫酶次單元 1（NADH dehydrogenase subunit 1, ND1）基因序列解碼，並討論利用 DNA 序列重建巨蚓科的親緣關係

張智涵[1]、莊淑君[2]、陳韻如[2]、陳俊宏[1,2,3]

(收稿日期：2005 年 1 月 18 日；接受日期：2005 年 3 月 25 日)

摘　　要

本研究將一種巨蚓科蚯蚓 *Amynthas gracilis* (Kinberg, 1867)的粒線體 NADH dehydrogenase subunit 1 (ND1)基因的完整 DNA 序列解出。此基因的 DNA 序列長度為 923 bp，轉譯形成 307 個胺基酸序列。將 *A. gracilis* 的 DNA 序列及胺基酸序列與正蚓科蚯蚓 *Lumbricus terrestris* 相比較，發現它們的相似性分別達 71.8% 及 77.7%。除此之外，此研究亦從基因銀行調出其它來自環節動物的 ND1 基因完整 DNA 序列，除了比較這些序列的 GC 含量外，亦利用譜系分析建構出這些序列的關係。比較前人在不同類群無脊椎動物的研究後，我們認為一段包含 ND1、CO1 與其它已發表基因在內，總長度超過 3000 bp 的 DNA 序列，將有助於瞭解目前不甚清楚的巨蚓科親緣關係。

關鍵詞: 巨蚓科、纖細遠環蚓、NADH 去氫酶次單元 1、親緣關係。

1. 國立台灣大學動物學研究所，台北市 106 羅斯福路 4 段 1 號，台灣。
2. 國立台灣大學生命科學系，台北市 106 羅斯福路 4 段 1 號，台灣。
3. 通信作者。Tel: 886-2-23630231 ext. 2354; Fax: 886-2-23658912; Email: chenjh@ntu.edu.tw。