Phylogeny of the Megascolecidae and Crassiclitellata (Annelida, Oligochaeta): combined versus partitioned analysis using nuclear (28S) and mitochondrial (12S, 16S) rDNA

Barrie G. M. JAMIESON
Department of Zoology and Entomology, University of Queensland,
Brisbane 4072, Queensland (Australia)
bjamieson@zen.uq.edu.au

Simon TILLIER
Annie TILLIER
Département Systématique et Évolution et Service de Systématique moléculaire,
Muséum national d'Histoire naturelle,
43 rue Cuvier, F-75231 Paris cedex 05 (France)

Jean-Lou JUSTINE
Laboratoire de Biologie parasitaire, Protistologie, Helminthologie,
Département Systématique et Évolution,
Muséum national d'Histoire naturelle,
61 rue Buffon, F-75321 Paris cedex 05 (France)
present address:
UR "Connaissance des Faunes et Flores Marines Tropicales",
Centre IRD de Nouméa, B.P. A5,
98848 Nouméa cedex (Nouvelle-Calédonie)

Edmund LING
Department of Zoology and Entomology, University of Queensland,
Brisbane 4072, Queensland (Australia)

Sam JAMES
Department of Life Sciences, Maharishi University of Management,
Fairfield, Iowa 52557 (USA)

Keith MCDONALD
Queensland Parks and Wildlife Service,
PO Box 834, Atherton 4883, Queensland (Australia)

Andrew F. HUGALL
Department of Zoology and Entomology, University of Queensland,
Brisbane 4072, Queensland (Australia)

Jamieson B. G. M., Tillier S., Tillier A., Justine J.-L., Ling E., James S., McDonald K. & Hugall A. F. 2002. — Phylogeny of the Megascolecidae and Crassiclitellata (Annelida, Oligochaeta): combined versus partitioned analysis using nuclear (28S) and mitochondrial (12S, 16S) rDNA. *Zoosystema* 24 (4) : 707-734.

ABSTRACT

Analysis of megascolecoid oligochaete (earthworms and allies) nuclear 28S rDNA and mitochondrial 12S and 16S rDNA using parsimony and likelihood, partition support and likelihood ratio tests, indicates that all higher, suprageneric, classifications within the Megascolecidae are incompatible with the molecular data. The two data-sets (nuclear and mitochondrial) may have historical or methodological incompatabilities therefore we explore the effect on measures of support and conflict at three levels: 1) separate analysis; 2) combining the data with single model; and 3) combining the relative support for competing topologies using separate models. Resolving power is identified *via* partition support, consensus and four competing likelihood ratio tests. Combined analysis identifies hidden support and conflict; more complex models reduce this conflict, possibly owing to removal of dynamic heterogeneity, and give a more resolved consensus. This is incompatible with morphological classifications, rejection of which varies among likelihood ratio tests. Both congruence and combined power support our conclusions: most of the groupings are based on homoplasies, for instance, multiple origin of racemose prostates or of « dichogastrin » meronephridia. The widely used classification of the non-ocnerodrilin Megascolecidae into three groups (Acanthodrilidae, with tubular prostates and holonephridia; Octochaetidae, with tubular prostates and meronephridia; and Megascolecidae, with racemose prostates) is not supported by molecular data. Monophyly of the Crassiclitellata Jamieson, 1988, oligochaetes with a multilayered clitellum, is confirmed. The results provide support for including the branchiobdellids and leeches in the Oligochaeta.

KEY WORDS
Annelida,
Oligochaeta,
Clitellata,
Crassiclitellata,
Megascolecidae,
molecular systematics,
maximum likelihood,
consensus,
parametric tests.

RÉSUMÉ

Phylogénie des Megascolecidae et Crassiclitellata (Annelida, Oligochaeta) : analyse combinée contre analyse partitionée, utilisant l'ADNr nucléaire (28S) et mitochondrial (12S, 16S).

Chez les oligochètes mégascolécoides, une analyse de l'ADNr nucléaire (28S) et mitochondrial (12S et 16S), utilisant la parcimonie et la vraisemblance, le soutien des partitions et les tests de taux de vraisemblance, indique que toutes les classifications supragénériques des Megascolecidae sont incompatibles avec les données moléculaires. Les deux jeux de données, nucléaire et mitochondrial, peuvent présenter des incompatibilités historiques ou méthodologiques et nous explorons donc l'effet du soutien et du conflit sur les mesures, à trois niveaux : 1) analyses séparées ; 2) combinaison des données avec un seul modèle ; et 3) combinaison des soutiens pour des topologies en compétition en utilisant des modèles séparés. La puissance de résolution est identifiée par le soutien des partitions, le consensus, et quatre tests de taux de vraisemblance en compétition. L'analyse combinée identifie le soutien caché et le conflit ; des modèles plus complexes réduisent ce conflit, probablement parce qu'ils suppriment une hétérogénéité dynamique et produisent un consensus mieux résolu. Ceci est incompatible avec les classifications morphologiques, qui sont plus ou moins rejetées en fonction du test de taux de vraisemblance utilisé. La congruence et le soutien de pouvoir combiné soutiennent nos conclusions : la

MOTS CLÉS
Annelida,
Oligochaeta,
Clitellata,
Crassiclitellata,
Megascolecidae,
systématique moléculaire,
maximum de vraisemblance,
consensus,
tests paramétriques.

plupart des groupements ont été basés sur des homoplasies, par exemple l'origine multiple des prostates racémeuses et de la méronéphridie de « type Dichogastrinae ». La classification largement utilisée des Megascolescidae non-Ocnerodrelinae en trois groupes (Acanthodrilidae, avec prostate tubulaire et holonéphridie ; Octochaetidae, avec prostate tubulaire et méronéphridie ; Megascolescidae, avec prostate racémeuse) n'est pas soutenue par les données moléculaires. La monophylie des Crassiclitellata Jamieson, 1988, oligochètes avec un clitellum à plusieurs couches, est confirmée. Les résultats fournissent des arguments pour l'inclusion des branchiobdellides et des hirudinées dans les oligochètes.

INTRODUCTION

General

The Clitellata Michaelsen, 1919 are annelids which include the oligochaetes (earthworms and their allies), branchiobdellids (ectoparasites of freshwater crayfish) and leeches. They are defined by the possession of a modification of the epidermis, the clitellum, which is located at least partly behind the female pores and which secretes a cocoon in which the eggs are laid. They were renamed Euclitellata by Jamieson (1983) because a clitellum also occurs in questid polychaetes though there anterior to the female pores (Jamieson 1983). Authorships of taxa indicated in Table 1 will not be repeated in the text except where required for clarity.

Molecular studies have confirmed monophyly of the Clitellata, using 18S rRNA (Winnepenninckx *et al*. 1998; Martin *et al*. 2000; Martin 2001; Rota *et al*. 2001), cytochrome oxidase I (COI) (Siddall & Burreson 1998; Nylander *et al*. 1999) or elongation factor 1-alpha (Kojima 1998); see also a review by McHugh (2000). An apparent exception to clitellate monophyly in a parsimony analysis of 18S rRNA was rejected as being due to spurious attraction between two apparently polychaete sequences and the branchiobdellids (Martin 2001). With regard to the position of the Clitellata within the Annelida, molecular analysis has indicated that (eu)clitellates form a clade within the Polychaeta and that polychaetes are a paraphyletic or polyphyletic group (Kojima 1998; McHugh 2000; Martin 2001; Rota *et al*. 2001). Relationships of oligochaetes, branchiobdellids and leeches, within the Clitellata, have been more elusive of definition.

Paraphyly of the Oligochaeta with leeches and/or branchiobdellids lying within the oligochaete clade, has long been suspected on morphological grounds (Michaelsen 1928-1932; Brinkhurst & Nemec 1986; Brinkhurst & Gelder 1989; Jamieson *et al*. 1987; Jamieson 1988; Purschke *et al*. 1993) and is being increasingly confirmed from molecular analyses. Siddall & Burreson (1998) and Siddall *et al*. (2001), using a combination of nuclear 18S and mitochondrial COI sequences, found support for the argument that leeches and branchiobdellids, with the leech-like fish parasite *Acanthobdella* Grube, 1851, form a monophylum within the Oligochaeta close to the aquatic oligochaete family Lumbriculidae. Derivation of leeches from a lumbriculid-like ancestor had been suggested on morphological grounds by Michaelsen (1928-1932) and Brinkhurst & Nemec (1986). Martin (2001), using complete 18S rRNA gene sequences, and taking secondary structure into account, also included Euhirudinea Lukin, 1956 (true leeches) and Acanthobdellida Livanow, 1905 in the Oligochaeta. He suggested the Branchiobdellida Holt, 1963 *via* the Lumbriculidae as a possible link between the two assemblages; the exact position of Hirudinea and Branchiobdellida within oligochaetes remained unresolved. An extremely ancient radiation of polychaetes and emergence of (eu)clitellates was proposed. Rota *et al*. (2001), for the same gene, with and without

consideration of secondary structure, but omitting lumbriculids, confirmed the sister-group relationship between *Acanthobdella* and Hirudinea but found the branchiobdellids to be a more distant clade, paired with two polychaetes; the many polychaetes included appeared highly polyphyletic.

CRASSICLITELLATE AND MEGASCOLECID HYPOTHESES

Oligochaetes *s.s.* are marine, freshwater and terrestrial. Unlike leeches they do not include parasitic species though two ectocommensal species on earthworms are known. With the exception of some earthworm-like genera, aquatic oligochaetes are usually small and are loosely termed "microdriles". They are characterized by a plesiomorphic type of clitellum in which, like the epidermis from which it is derived, there is only a single layer of cells. Its simple structure and limited ability to secrete nutrients into the cocoon correlate with the production of small numbers of large, yolky eggs. A major evolutionary innovation in earthworms (loosely termed "megadriles"), in contrast, has been the development of a clitellum consisting of several layers of cells with the ability to secrete large proteinaceous reserves into the cocoon. Correlated with this, the eggs possess little yolk, are therefore small and produced in large numbers (Jamieson 1992). In a morphocladistic analysis (Jamieson 1988), all families with multilayered clitella were found to form a single clade, named the Crassiclitellata Jamieson, 1988. Thus acquisition of a multilayered clitellum was deduced to be a monophyletic event. However, Omodeo (2000) implied that a multilayered clitellum has arisen more than once when he derived the Eudrilidae (with multilayered clitellum) from the Alluroididae Michaelsen, 1900 (with clitellum consisting of a single layer of cells) independently of other earthworms, a familial relationship not supported morphocladistically. In the present study we test the monophyly of the Crassiclitellata using molecular data.

It is not the aim of this paper to make a detailed examination of higher-level relationships within the Crassiclitellata. However, some analysis is made of division of the Crassiclitellata on morphocladistic evidence (Jamieson 1988) into two groups the Aquamegadrili Jamieson, 1988 and Terrimegadrili Jamieson, 1988. Aquamegadrili have an aquatic or semi-aquatic mode of life, and consisted of the families Sparganophilidae (Holarctic), Biwadrilidae Jamieson, 1971 (Japan), Almidae (mostly warm tropics but including *Criodrilus* Hoffmeister, 1845, in the Mediterranean region, etc.) and Lutodrilidae (Southern Neartic). It is not unlikely that the aquamegadrile families, irrespective of mono- or polyphyly of the group, have always had an aquatic or amphibious existence. The remainder of the Crassiclitellata were predominantly terrestrial, hence the term Terrimegadrili. These consisted of the superfamilies Ocnerodriloidea Beddard, 1891, Eudriloidea Claus, 1880, Lumbricoidea Claus, 1876, and Megascolecoidea Rosa, 1891. The validity of recognizing the Aquamegadrili and Terrimegadrili is tested here from molecular data.

The Lumbricoidea as redefined by Jamieson (1978, 1988) included the Lumbricidae (Holarctic), Komarekionidae (Nearctic), Glossoscolecidae (Neotropical), Microchaetidae (Ethiopian, South of the Kalahari), Hormogastridae (western Palaearctic, Tyrrhenian), and Ailoscolecidae Bouché, 1969 (Palaearctic). Whether the Kynotidae Jamieson, 1971 (Malagasy) should be assigned to the Aquamegadrili or Terrimegadrili was uncertain. Some further families have been added more recently (see Omodeo 2000). The families Megascolecidae, Ocnerodrilidae and Eudrilidae had been tentatively included in the superfamily Megascolecoidea by Jamieson (1978, 1980). However, in the cladistic analysis (Jamieson 1988), the Eudrilidae (superfamily Eudriloidea) and especially the Ocnerodrilidae (superfamily Ocnerodriloidea) occupied a basal position relative to the other terrimegadrile families. These classifications are evaluated below and, to anticipate, are all called into question.

The largest, most speciose, earthworm family is the Megascolecidae for which a Pangean origin has been suggested (Jamieson 1981). They are

native in the Nearctic, Ethiopian, Oriental, Australian, eastern Palaearctic (China, Japan, Korea) and southern Neotropical regions, with Central America. In South America North of the Juramento-Salado River, megascolecids are replaced by the large family Glossoscolecidae. In the Ethiopian region, particularly in Tropical West and East Africa, the family Eudrilidae greatly outnumbers it in genera. Currently recognized subfamilies of the Megascolecidae are the Acanthodrilinae Vejdowsky, 1884 and Megascolecinae Rosa, 1891, with or without the Ocnerodrilinae Beddard, 1891. Indigenous acanthodriles are predominant in the earthworm faunas of the southern and eastern portions of North America, Mexico, Guatemala, southern South America, South Africa, New Zealand, New Caledonia, and parts of Australia. The native range of the Ocnerodrilidae includes the warmer parts of North and South America, the Dominican Republic, Africa, India and Burma (Jamieson 1981).

Several, sometimes widely divergent, classifications of these megascolecoid earthworms have been proposed since publication of Stephenson's monograph of the Oligochaeta (1930) (see Michaelsen 1933; Pickford 1937; Omodeo 1958; Lee 1959; Gates 1959; Sims 1966, 1967; Jamieson 1971a-c, 1978, 1988). Particular attention will be paid in the present study to the system of Gates (1959), supported by Sims (1966, 1967), and that of Jamieson (1971a-c) as both of these systems are widely used. Detailed discussion of these alternative classifications may be found in the 1971 papers.

The procedures for analysing the molecular data for representatives of the above taxa will now be considered.

Combining data, likelihood models and hypothesis testing

Higher-level systematics is turning to congruence and combined power in multiple data-set analyses as the most convincing source of phylogenetic signal. This raises issues of how to combine the data and how to identify possible conflict. Krajewski *et al.* (1999) summed up the issues involved with integrating different data in terms of 1) historical heterogeneity (the data have different histories) versus 2) dynamic heterogeneity (methods for extracting the phylogenetic signal are incompatible). As they are independent genomes, the nuclear 28S rDNA and the mtDNA data-sets are biologically independent sources of information (i.e. they could have historically different genealogies) and so allow congruence as evidence. They may also have very different apparent sequence evolution patterns, being "dynamically" heterogeneous. To date most analyses have combined data (e.g., Flook *et al.* 1999; Crandall *et al.* 2000). However, lumping heterogeneous data produces a patently artifactual model, compromising parameter estimation and interpretation of levels of support (Goldman 1993; Yang 1996).

Real conflict (e.g., introgression, ancestral polymorphism, paralogous loci) may be overlooked by tests of total character support (such as likelihood ratio tests). On the other hand lumping individual differences in consensus may overlook underlying similarities. Partition support analysis (Baker & DeSalle 1997) allows such hidden support and conflict to be identified. Hence, for likelihood analyses in addition to combined data methods we have chosen to combine the data by adding the likelihoods of each partition optimised individually (Edwards 1972; Adachi & Hasegawa 1992; Huelsenbeck & Bull 1996; Yang 1996; Wilgenbusch & De Queiroz 2000). We then compare individual contributions with partition support, extended to likelihood (Lee & Hugall in press). Summing the likelihoods of hypotheses (the "support surface") from different data-sets is the general likelihood framework for combining disparate sources of information, which in the case of phylogenetics could include non-sequence data, given appropriate models.

Here we apply this approach to a phylogenetic study of nuclear and mitochondrial DNA sequences in clitellates, allowing a revision of the higher classification of the earthworm family Megascolecidae and permitting examination of the validity of the Crassiclitellata and the relationships of these within the Clitellata.

TABLE 1. — Taxa sequenced for this study. Classification according to Jamieson (1971a-c, 1988). * *Lumbricus terrestris* Linnaeus, 1758 GenBank accession. †, for Figure 4 analysis, 16S is from *Digaster lingi*. Specimens collected by authors except where indicated.

Taxon	Location	28S	12S	16S
OUTGROUP				
Hirudinea Lamarck, 1818				
Haemadipsidae Blanchard, 1893				
Haemadipsidae Gen. sp. (terrestrial)	Mt Glorious, Queensland	AY101557		
Haemadipsidae Gen. sp. (freshwater)	North Stradbroke Island, Queensland	AY101558		
Branchiobdellidae Odier, 1823				
Cambarincola pamelae Holt, 1984	USA (R. O. Brinkhurst)	AF406601		
Xironodrilus formosus Ellis, 1919	USA (R. O. Brinkhurst)	AF406600		
Oligochaeta Grube, 1850				
Lumbriculidae Vejdovsky, 1884				
Lamprodrilus Michaelsen, 1901 sp.	Lake Baikal (P. Martin)	AF406592		
Lumbriculus variegatus (Müller, 1774)	Cultured, Brisbane	AF406594		
Rhynchelmis brachycephala Michaelsen, 1891	Lake Baikal (P. Martin)	AF406593		
Tenagodrilus musculus Eckroth & Brinkhurst, 1996	USA (R. O. Brinkhurst)	AF406591		
Enchytraeidae Vejdovsky, 1879				
Achaeta bohemica Vejdovsky, 1879	Arezzo, Italy (E. Rota)	AF406595		
Enchytraeus albidus Henle, 1837	Brisbane, Queensland	AF406597		
Fridericia bisetosa (Levinsen, 1884)	Arezzo, Italy (E. Rota)	AF406596		
Tubificidae Vejdovsky, 1884				
Branchiura sowerbyi Beddard, 1892	Cultured, Australia	AY101559		
Naididae Ehrenberg, 1831				
Dero Oken, 1815 (*Aulophorus* Schmarda, 1861) sp.	Cultured, Queensland	AF406598		
Haplotaxidae Michaelsen, 1900				
Haplotaxis Hoffmeister, 1843 sp.	Logan River USA (R. O. Brinkhurst)	AF406599		
Haplotaxis gordioides (Hartmann, 1821)	Trezzotinella, Italy (E. Rota)	AF406602		
Sparganophilidae Michaelsen, 1918				
Sparganophilus tamesis Benham, 1892	Missouri, USA	AY101566		
Lutodrilidae McMahan, 1976				
Lutodrilus multivesiculatus McMahan, 1976	Louisiana, USA	AY101567		
Almidae Duboscq, 1902				
Criodrilus lacuum Hoffmeister, 1845	Algeria (P. Omodeo)	AY048492	AY101545	
Ocnerodrilidae Beddard, 1891				
Eukerria saltensis (Beddard, 1895)	Cultured, Brisbane, Queensland	AY048496	AY101546	AF406590
Eudrilidae Claus, 1880				
Eudrilus eugeniae (Kinberg, 1867)	Cultured, Brisbane, Queensland	AY101568	AY048471	
Komarekionidae Gates, 1974				
Komarekiona eatoni Gates, 1974	Kentucky, USA	AY101569		
Microchaetidae Michaelsen, 1900				
Microchaetidae Gen. sp.	Grahamstown, South Africa (A. Hodgson)	AY101570		
Glossoscolecidae Michaelsen, 1900				
Glossoscolecidae Gen. sp.	Prov. Pichincha, Ecuador	AY048507		
Pontoscolex corethrurus (Müller, 1856)	North Queensland	AY101571		
Lumbricidae Claus, 1876				
Eisenia fetida (Savigny, 1826)	Paris, France	AY048508		
Lumbricidae Gen. sp.	Samford, Queensland	AY048498	AY048472	U24570*
Hormogastridae Michaelsen, 1900 (As Hormogastrinae)				
Hormogaster redii Rosa, 1887	Molara Is., Sardinia, Italy (P. Omodeo)	AY048506		

Taxon	Location	28S	12S	16S
INGROUP				
Oligochaeta Grube, 1850				
Megascolecidae Rosa, 1891				
Acanthodrilinae Vejdovsky, 1884				
Diplocardia longiseta Murchie, 1965	Iowa, USA	AY101572		
Diplotrema acropetra Jamieson, 1997	Rocky Peak, Starke Station, Queensland	AY101573	AY048469	AF406568
Diplotrema Spencer, 1900 sp.	Souita Falls, Atherton Tableland, Queensland	AY048477	AY048466	AF406570
Neodiplotrema altanmoui Jamieson, 1997	Altanmoui Range, Queensland	AY101574	AY048468	AF406569
Rhododrilus glandifera Jamieson, 1995	Wooroonooran National Park, Queensland		AY048467	
Megascolecinae Rosa, 1891				
Perionychini Jamieson, 1971				
cf. *Argilophilus* Eisen, 1893 sp.	Oregon, USA	AY101575		
Diporochaeta Beddard, 1890 sp.	Near Cradle Mt, Tasmania	AY048479	AY048460	AF406574
Fletcherodrilus unicus (Fletcher, 1889)	Brisbane, Queensland	AY048474	AY101547	
Fletcherodrilus unicus (Fletcher, 1889)	Broken River, Eungella, Queensland	AY101565	AY048423	AF406558
Fletcherodrilus fasciatus (Fletcher, 1890)	Lamington National Park, Queensland	AY048503	AY101548	
Fletcherodrilus sigillatus (Michaelsen, 1916)	Pelling's, Atherton Tableland, Queensland	AY048473	AY048425	AF406588
Heteroporodrilus Jamieson, 1970 sp.	Brisbane, Queensland	AY048497	AY101553	AF406579
Diporochaeta cf. *kershawi* (Jamieson, 1974)	Tasmania (R. Blakemore)	AY048484	AY048461	AF406567
Perionyx excavatus Perrier, 1872	Cultured, Brisbane, Queensland	AY048499	AY048456	AF406582
Pontodrilus litoralis (Grube, 1855)	Bush Bay, Western Australia (M. Harvey)	AY101576	AY048463	AF406586
Terrisswalkerius athertonensis (Michaelsen, 1916)	Carbine Tableland, Queensland	AY048504	AY048453	AF406585
Terrisswalkerius grandis (Spencer, 1900)	Lamb Range, Queensland	AY048495	AY101551	AF406566
Terrisswalkerius kuranda (Jamieson, 1976)	Ebony Creek Rd, Queensland	AY048478	AY048432	AF406581
Terrisswalkerius millaamillaa (Jamieson, 1976)	Souita Falls, Atherton Tableland, Queensland	AY048476	AY048452	AF406565
Terrisswalkerius phalacrus (Michaelsen, 1916)	Souita Falls, Queensland	AY048490	AY048450	AF406577
Terrisswalkerius phalacrus (Michaelsen, 1916)	Kennedy Falls, Queensland	AY101556	AY048449	
Terrisswalkerius windsori Jamieson, 1995	Windsor Tableland, Queensland	AY048494	AY101552	AF406587
Dichogastrini Jamieson, 1971				
Dichogaster Beddard, 1888 sp. (sexprostatic)	Plateau Boucher, Martinique	AY101555	AY101549	AF406571
Dichogaster saliens (Beddard, 1893)	Samford, Queensland	AY048493	AY048470	AF406573
Didymogaster sylvaticus Fletcher, 1886	Hornsby, New South Wales	AY048491	AY101554	AF406575
Digaster anomala Jamieson, 1970	Brisbane, Queensland	AY048480	AY048462	†
Digaster lingi Jamieson, 1995	Binna Burra, Queensland	AY101561	AY048459	AF406583
Megascolecini Jamieson, 1971				
Amynthas rodericensis (Grube, 1879)	Brisbane, Queensland	AY101562		
Begemius queenslandicus Easton, 1982	Mt Lewis, Carbine Tableland, Queensland	AY101563	AY048465	AF406578
Propheretima hugalli Jamieson, 1995	Boat Harbour, Lismore, Queensland	AY048505	AY101550	
Spenceriella cormieri Jamieson & Wampler, 1979	O'Reilly's, Border Ranges. Queensland	AY101564	AY048458	AF406589
Spenceriella Michaelsen, 1907 sp.	Boat Harbour, Lismore, New South Wales	AY048475	AY048454	AF406572
Number of sequences		59	34	28

TABLE 2. — Primers employed. Position numbers are from the Lumbricus terrestris mtDNA (accession number U24570).

Primer	3' position	Sequence 5' to 3'	Gene	Author
12SE1	10538	AAAACATGGATTAGATACCCRYCTAT	12S rRNA L	present work
12SH	10919	ACCTACTTTGTTACGACTTATCT	12S rRNA H	present work
16SAR	11639	CGCCTGTTTATCAAAAACAT	16S rRNA L	Palumbi (1996)
16SBR	12120	CCGGTCTGAACTCAGATCACGT	16S rRNA H	Palumbi (1996)
C1'		ACCCGCTGAATTTAAGCAT	28S rRNA "coding"	Tillier
D2		TCCGTGTTTCAAGACGG	28S rRNA	Tillier
C2		TGAACTCTCTCTTCAAAGTTCTTTTC	28S rRNA	Tillier
C2'		GAAAAGAACTTTGRARAGAGAGT	28S rRNA	Tillier

MATERIALS AND METHODS

The list of the species used in this analysis is presented in Table 1. The list follows the classification of Jamieson (1971a-c, 1988). Voucher specimens are lodged in the Queensland Museum Brisbane, Australia.

MOLECULAR DATA

One fresh specimen of *Fletcherodrilus sigillatus* mtDNA was subjected to differential centrifugation and CsCl density gradient ultra centrifugation for DNA purification. All other specimens were preserved in ethanol. For these, DNA was purified by SDS detergent based lysis in the presence of Proteinase K followed by extraction with phenol/chloroform and ethanol precipitation, or by digesting tissue at 55°C in 5 % chelex water and 5 ml of proteinase K (10 mg/ml). Some of the DNA extractions were performed using CTAB (Cetyltrimethylamonium bromide) (Winnepenninckx *et al.* 1993).

The primers used are indicated in Table 2. For 28S genes, universal primers located in the C1 domain and just 3' to the end of the D2 domain, designed by A. Tillier, amplified a *c.* 770 bp fragment covering most of the C1, and all of the D1, C2, and D2 domains. A 450 bp fragment of the mtDNA 12S was amplified and sequenced using modified versions of the standard 12S primers (Palumbi 1996) (developed using *Lumbricus terrestris*, GenBank accession number U24570). The primer 12SE1 was designed around the traditional 12S1 site, while the primer 12SH was placed 31 bp 3' to the 12S2 site (see Table 2). MtDNA 16S was amplified and sequenced using standard 16S rDNA primers 16SAR and 16SBR (Palumbi 1996). Nuclear 28S PCR (polymerase chain reaction) included 3.5 % DMSO. Most PCR products were gel purified (some directly by addition of Exonuclease and Shrimp alkaline phosphatase) and sequenced using ABI dye terminator automated sequencing for, in most cases, both strands. Some were manually sequenced. The taxa analysed in the study are drawn from a larger set (unpublished) which includes multiple representatives of eight of the OTUs (operational taxonomic units) presented here: all these within-species gene comparisons are congruent.

Nucleotide sequences were first aligned using CLUSTAL W (Thompson *et al.* 1994). These were then modified according to accepted secondary structure models for 12S (Hickson *et al.* 1996) and 16S (see De Rijk *et al.* 1999). Across the 28S rDNA sequence there were several regions of ambiguous alignment within the D1 and D2 variable domains. In the Crassiclitellata and Megascolecidae 28S data-set, three regions were removed: one in the D1 domain (10 bp between the B13 and B13 1 stems) and two in the D2 domain (10 and 12 bp), leaving 622 sites. The higher level oligochaete 28S data-set was more difficult to align and more sections of the D2 expansion region were removed resulting in 549 sites (alignment available from authors). For 12S rDNA, four loop regions of variable length between stem regions 40'-39', 42-42', 47-47' and 48-48' totalling 30-50 bp were removed, leaving 315 sites. Using the nomenclature of De Rijk *et al.* (1999) small sections of the E25 and G3 loops were removed from the 16S rDNA alignment, totalling 10-20 bp, leaving 435 sites.

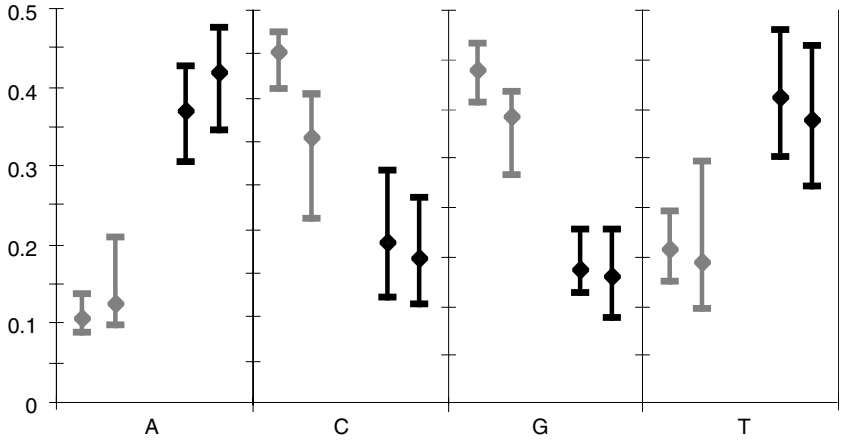

FIG. 1. — Average and range of base content of variable sites across all available taxa (59 for 28S, 25 for mtDNA genes). Left pair for each base: the 28S rDNA for Megascolecidae and outgroup taxa. Right pair for each base: megascolecid 12S rDNA and 16S rDNA.

PHYLOGENETIC ANALYSES

PAUP* 4.0b3a (Swofford 2000) was used for neighbour joining, parsimony and likelihood analyses, bootstrapping (BS), tree topology distances, base content, partition homogeneity tests, Bremer values, and branch length and pairwise divergence estimates. PHYLIP 3.5 (Felsenstein 1995) was used to generate certain bootstrap resamplings and trees from base-content distances. Random and various constraint trees were generated in MacClade 3.07 (Maddison & Maddison 1995). Parametric data-sets were generated with 100 (sometimes 200) replicates using SeqGen 1.1 (Rambaut & Grassley 1997) with model parameters drawn from maximum likelihood (ML) estimates. Modeltest 3.0 of Posada & Crandall (1998) provided information on likelihood model choice. Maximum parsimony (MP) analyses used heuristic search with tree bisection reconnection (TBR) and stepwise random sequence addition. ML model parameters and trees were optimised by a heuristic procedure of successive searches and re-optimisations using SPR (subtree pruning and regrafting) (Swofford in PAUP* release notes; Huelsenbeck 1998). Throughout we refer to relative log-likelihood (ΔlnL) as the difference in -lnL between topologies, typically a ML tree versus others.

Assessing congruence between mitochondrial and nuclear genes

Apparent conflict between the nuclear and mitochondrial genes was explored with the character based partition homogeneity test (ILD [incongruence length difference]; Farris *et al*. 1995; for excluding uninformative sites, see Lee 2001), likelihood ratio tests, the parametric test of Huelsenbeck & Bull (1996), and partition support using both MP and ML (Baker & DeSalle 1997; Lee & Hugall in press). Partition support values, both Bremer and relative lnL, were calculated in PAUP* using reverse constraint trees (best tree not containing a specified clade). Likelihood is partitioned using site -lnL values.

Combining mtDNA and nuclear gene sequence data

Complex models in likelihood procedures are designed to accommodate heterogeneity among sites. Although variation among the mtDNA genes concerned here is likely to be not much more than that within a gene, the differences between the mtDNA and nuclear genes are much more striking. As the sequence statistical analysis indicates that the mtDNA and nuclear genes are substantially different in their basic characteristics (Fig. 1; Table 3), we approach the issue of

dynamic heterogeneity by optimising models and parameters for each data partition but inferring a phylogeny from the total score. Within the mtDNA data, as the 12S and 16S sequences are sufficiently similar in character, neither dynamic nor historical heterogeneity is an issue (notwithstanding cryptic pseudogenes): they have been pooled, in favour of one model, in preference to splitting up the number of variable sites which might result in over specified models for the amount of data.

While combining data in MP is straightforward – all parts have the same model – for likelihood analyses we have chosen to combine the data in two ways: the first, called the COMBO model, optimises the model to the combined data; the second, called the SUM model, adds the likelihoods of each partition optimised individually (Edwards 1972; Wilgenbusch & De Queiroz 2000). To provide a tree search space for the SUM method we evaluated 14586 trees drawn from 1000 near parsimonious reverse constraint trees for each node in the combined data tree (Fig. 4).

Partition support analysis

In addition to summary statistics of the fit or otherwise of each data-set to the other, hidden support or conflict can be apportioned to each node for each partition (Baker & DeSalle 1997) using the reverse constraint tree set. Here we can extend this by comparing the mtDNA and 28S components in the combined analysis (COMBO) and the components in the SUM method. The partitioned likelihood scores in the combined data model are sums of the site -lnLs.

Assessing the effect of model design on tree support space

We determined a suitable nucleotide substitution model on the basis of the relative likelihood scores of the MP bootstrap consensus tree as parameters were added to the model, following the method of Posada & Crandall (1998) but interpreting the ΔlnL conservatively in accordance with the observations of Takahashi & Nei (2000) and Yang (1997).

Different models and parameters values can change the absolute likelihood substantially but the relative log likelihood amongst competing trees – the likelihood support surface – may be less sensitive to the details of the model over the range of plausible trees (Yang *et al*. 1995). We have investigated the effect of different models by comparing the relative log likelihood (ΔlnL) among trees, using the 25 megascolecid taxa for which we have both 28S and mtDNA data-sets. This is presented in Figure 6 as the relative lnL to the best SUM model tree (Fig. 4) for a subset of the near parsimonious reverse constraint trees used in the SUM method search set. In particular we compare the SUM versus the COMBO models. These support surface reverse constraint trees cover a wide range of likelihoods with small ΔlnL between each.

Hypothesis testing

Specific hypotheses were investigated with likelihood ratio tests comparing the best tree with the best tree constrained to that hypothesis. As there is a number of competing tests, and some uncertainty over their qualities, we used the distribution of relative lnL to the ML tree among BS pseudoreplicates as the basis for the likelihood ratio tests of Kishino & Hasegawa (1989) and Shimodaira & Hasegawa (1999) as described in Goldman *et al.* (2000), and for the estimated confidence from expected likelihood weight (*c*) of Strimmer & Rambaut (2002). These methods were applied to the best trees consistent with the hypotheses listed in the results and in Table 5. The estimated likelihood weight confidence was also applied to the COMBO model reverse constraint trees (Table 4). The distribution of ΔlnL was calculated from 200 PHYLIP generated bootstrap pseudoreplicates, using fixed model parameters.

Parametric tests

Parametric bootstrapping is a statistical tool for producing independent replicates of a study based on parameters estimated from a unique data-set (Huelsenbeck & Hillis 1996). Parametric methods explore the variance in the explicit model by crea-

TABLE 3. — GTR-Γ model parameter estimates for the 25 taxa of Megascolecid 28S, mtDNA, and combined data-sets. Alpha = 0.196/0.295/0.254. 28S to the left, mtDNA middle, and combined data right (T-G rate is normalised to 1). Four discrete rate categories for gamma, empirical base content.

	C	G	T	base content
A	1.40/3.95/2.06	1.53/11.09/4.92	1.61/10.41/8.73	0.15/0.36/0.27
C		0.37/0.96/0.99	4.33/28.35/13.20	0.32/0.18/0.25
G			1	0.38/0.19/0.27
T				0.15/0.27/0.21

ting data-sets based on that model. It therefore is a question of how much faith one puts in the model to be a reasonable measure of the natural variance or uncertainty in the real data (Goldman 1993; Yang *et al.* 1995; Huelsenbeck & Hillis 1996). Parametric methods are however ideal for comparing models (Goldman 1993; Huelsenbeck & Bull 1996; Huelsenbeck & Rannala 1998). We have used both facets of parametric methods. In assessing reliability of phylogenetic inference we use the approach developed by Swofford *et al.* (1996), illustrated in Bromham & Degnan (1999), and which Goldman *et al.* (2000) dubbed the SOWH (Swofford, Olsen, Waddell and Hillis) test: can a result consistent with one hypothesis be obtained by chance out of parametric data-sets built on the best observed data model that is not consistent with that hypothesis?

Given the nature of our taxon and sequence sampling, where, because of logistic and other constraints, only subsets of taxa have the full compliment of sequence data, we approach the analyses in a hierarchical manner, establishing the existence of higher groups so that subsequent analyses of within group relationships can be conducted using subsets of taxa that have larger suites of sequences. The phylogenetic analyses here comprise a 55 taxon, 28S alignment of 549 sites; combined 28S and mtDNA for 42 species in which a given species may be represented by one, two or three genes; and a more detailed study of 25 species for which all three genes have been sequenced (see Table 1). This latter includes 622 28S sites, 749 combined 12S and 16S mitochondrial sites, with hypothesis testing.

RESULTS

PATTERNS OF VARIATION IN 28S AND MTDNA DATA

Across all (55) taxa, 210 of 622 sites are variable in the megascolecid 28S rDNA, 352 of 549 in the higher (suprafamilial) oligochaete 28S, and 403 out of 749 in the mtDNA thus amounting to high levels of observed differences. The nuclear and mitochondrial genes show substantially different base contents while the mtDNA genes are more similar to each other (Fig. 1). The 28S rDNA is CG rich while the mitochondrial genes are AT rich. The mitochondrial genes show a larger range in base content, as might be expected with higher levels of divergence. The higher oligochaete 28S data-set (Fig. 2) shows significant base composition variation ($p < 0.001$ PAUP* test).

Using MP trees, large gains in likelihood are made allowing for site rate heterogeneity with the gamma parameter (Γ) and also by allowing for substitution rate heterogeneity using the 6-way general time reversible (GTR) model (Yang 1994). Some gain is made allowing for invariant sites in the 25 Megascolecid taxa set (2.9 ΔlnL: significant by Modeltest 3.0). We have interpreted the ΔlnL conservatively and kept to the GTR-Γ model as adequate, in accordance with the observations of Takahashi & Nei (2000) and Yang (1997).

Despite the opposite bias in base content both gene systems show a high estimated T-C rate and low G-C rate (Table 3). Both show high levels of site rate heterogeneity (alpha < 0.3), the 28S the more so. The estimated parameters are noticeably

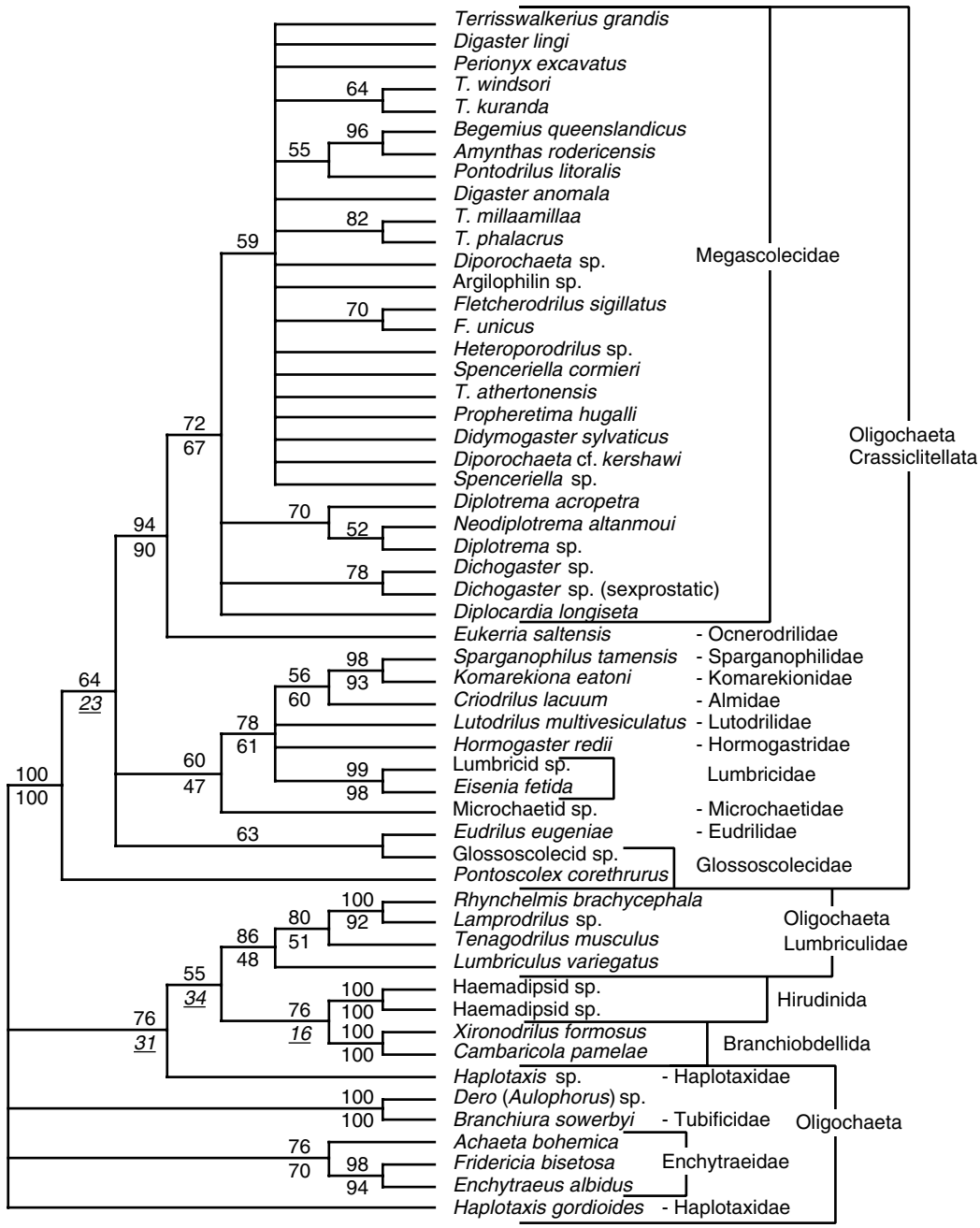

FIG. 2. — 28S rDNA maximum parsimony majority rule consensus of 1000 bootstrap resamples, five random sequence additions each with tree bisection reconnection swapping. Basal polytomy rooted with *Haplotaxis gordioides* (Hartmann, 1821). Maximum parsimony bootstraps above node, maximum likelihood bootstrap values below (200 resamplings from reduced 36 taxa set using only nine Megascolecidae). Those underlined were not recovered in the majority rule. Family and some higher level groups indicated (see Table 1 for details).

different from two-parameter models that distinguish transition/transversion rates only. The combined data model is a compromise between the two, resulting in much more even base contents and intermediate rate values (Table 3). Using the optimized GTR-Γ models estimated divergences (using the 25 taxa with all genes) range up to more than 0.30 for the 28S rDNA and to over 0.70 for the mtDNA (a substantial increase over observed difference), with a wide range of relative divergence between genes, 3-4 fold among taxa of similar 28S divergence. Thus, the data probably suffer saturation effects in the mtDNA relative to the 28S and substantial base content variation.

High level oligochaete phylogenetic inference using 28S

Initial analysis of the oligochaete 28S data-set contained 55 taxa including representatives of leeches, lumbriculids, branchiobdellids, enchytraeids and haplotaxids. Preliminary investigations indicated that addition of polychaete sequences provided little insight into the deeper root and was dependent on alignment. Considering the diversity of polychaetes in molecular data, which fail to recover monophyly (e.g., 18S see Winnepenninckx *et al.* 1998; elongation factor 1-alpha see Kojima 1998) or shows extreme polyphyly (Rota *et al.* 2001), the most appropriate outgroup of this data-set is uncertain but, by historical precedent, networks are presented as being been rooted with *Haplotaxis* species.

Figure 2 shows the result of a MP bootstrap with 1000 resamplings for the 549 bp of 28S, with the ML bootstrapping using GTR-Γ model below (reducing the number of taxa to 36 by representing the Megascolecidae with nine taxa across the Acanthodrilinae and Megascolecinae). Key conclusions from this are: the first two groupings (Crassiclitellata and the Megascolecidae + *Eukerria saltensis* [= Megascolecoidea, 94 %]) are statistically robust and, considering the level of our sampling, the Crassiclitellata and the Megascolecidae are each a monophyletic group.

A by-product is support for the published finding from 18S and COI for the 1) Lumbriculidae; 2) Enchytraeidae; 3) Naididae-Tubificidae; and 4) a leech-branchiobdellid clade. This last has been controversial owing to the concern that molecular artifacts (under the rubric of long branch attraction) undermine confidence in the result. Here, there is significant base composition variation in the 28S ($p < 0.001$ PAUP* test and see Fig. 1), in particular the leeches and the branchiobdellids being distinct from the others. Bootstrap support is fairly high in MP (76 %) but not sustained in ML (16 %), and the leech-branchiobdellid grouping is conspicuously absent using LogDet, a method that is supposed to compensate for nonstationarity (Lockhart *et al.* 1994). Further, parametric data-sets made to trees that do not contain this clade, when analysed with MP, group them at high bootstrap frequency (> 70 %; following the method of Huelsenbeck 1998, results not shown); we therefore suspect that MP support is artifactually high. Conversely, the low ML support may be due to long branch repulsion effects (Pol & Siddall 2001). Notwithstanding limited outgroups, that the 28S is tightly linked to the 18S and that it suffers a molecular bias that could confound methodology, the results here may be seen as some corroboration of this proposal.

The present study considers chiefly relationships in the Crassiclitellata and the Megascolecidae but some comment on wider clitellate relationships is given in the Discussion below.

Crassiclitellate taxa included in the morphocladistic analysis (Jamieson 1988) were the aquamegadrile families (with superfamilies there recognized) Sparganophilidae (Sparganophiloidea Michaelsen, 1918), Biwadrilidae (Biwadriloidea Jamieson, 1971), Almidae (including *Criodrilus*), Lutodrilidae (both Almoidea Dubosq, 1902, *sensu* Jamieson 1988) and, tentatively included, Kynotidae (new familial status for Kynotinae Jamieson, 1971), and the terrimegadrile families Ocnerodrilidae (Ocnerodriloidea), Eudrilidae (Eudriloidea), Microchaetidae, Hormogastridae, Glossoscolecidae, and Lumbricidae (all Lumbricoidea) and the Megascolecidae (Megascolecoidea). In the present molecular analyses, all but the Biwadrilidae and Kynotidae are represented.

Separate and combined 28S rDNA and mtDNA phylogenetic analysis of the Megascolecidae

Here we attempt to give greater resolution among the Megascolecidae by retrieving more sites within the 28S data by eliminating most outgroup taxa and adding combined 12S and 16S mitochondrial rDNA sequences. Figure 3 shows a strict consensus MP tree, with bootstrap values added from a separate analysis, for 42 species of Megascolecidae and outgroup families for combined 28S, 12S and 16S gene sequences. These make a useful comparison with the 28S only analysis (Fig. 2) and subsequent analyses of the 25 taxa available with all data but only two outgroup families. Despite the heterogeneity of the number of genes sampled, the tree in Figure 3 is highly informative and largely consistent with other trees for reduced taxon sets with uniform gene samples (see below).

As previously noted, we restrict the main mixed gene analyses (Figs 4; 5; Table 4) to 25 taxa for which we have all three genes: 622 sites of 28S and 749 sites of combined 12S and 16S mitochondrial rDNA sequence. Given the not unreasonable possibilities of dynamic and historical heterogeneity, in the first instance we analyse the two data partitions separately, then compare this with combined analysis results. These analyses are performed with both ML and MP methods to compare the effect of allowing for complex process models. The likelihood model parameters for the separate and combined analyses are shown in Table 3.

Incongruence and compatibility

The ML trees for the two partitions conflict in more than a few nodes so that the semi-strict consensus of the trees (Fig. 5A) loses much resolution. Neither tree has many strongly supported nodes so that majority rule bootstrap trees (not shown) also have low resolution, but where the trees differ, all but one bootstrap value are < 50%. Maximum parsimony produces much the same results (not shown). The character based partition homogeneity test (ILD) is borderline insignificant at $p = 0.066$. This could also reflect the substantially different characteristics of the 28S and mtDNA data (Dolphin *et al*. 2000). However, likelihood ratio tests, both parametric and non parametric, also record apparent phylogenetic incongruence: the optimal ML tree for each data partition is incompatible with the other by the SH test (54.3/86.9 ΔlnL, 28S/mtDNA, $p < 0.001$) but not with the combined data tree (24.6/5.7 ΔlnL, 28S/mtDNA, $p > 0.1$); the partitions are also significantly incongruous ($p < 0.01$) according to the parametric test of Huelsenbeck & Bull (1996). Despite evidence of incongruence, the two partitions combine to produce a more robust result. In addition to the congruence in the consensus tree (Fig. 5A), this can be seen in the average Bremer values of the best tree for each partition as compared with that for the combined data. Thus, the combined data value is greater than the sum of the partitions, indicating hidden support (128 vs 117), and there is also an increase in bootstrap levels (not shown).

Combining support versus combining data

We compare two methods of combining information in the 28S and mtDNA data: combining the data and analysing with a single model – the COMBO method; and combining the likelihoods of each partition using separate models – the SUM method. For the latter we evaluated 14586 unique topologies drawn from 1000 near parsimonious reverse constraint trees for each node in Figure 4, to represent the likelihood "support surface". At least in this case the combined data ML and the MP methods are close enough to provide the essential candidate topologies, which contain all the combined data ML best reverse constraint trees. It therefore represents a reasonable search space for the absolute best SUM model trees. The likelihoods for each of these topologies were calculated for both SUM and COMBO models. Table 4 tabulates various measures of support for each of these nodes: the partitioned likelihood support; the analogous MP Bremer values; MP and ML bootstrap values; and, in addition, the Strimmer & Rambaut (2002) expected likelihood weight estimated confidence (*c*) for the set of ML and reverse constraint trees.

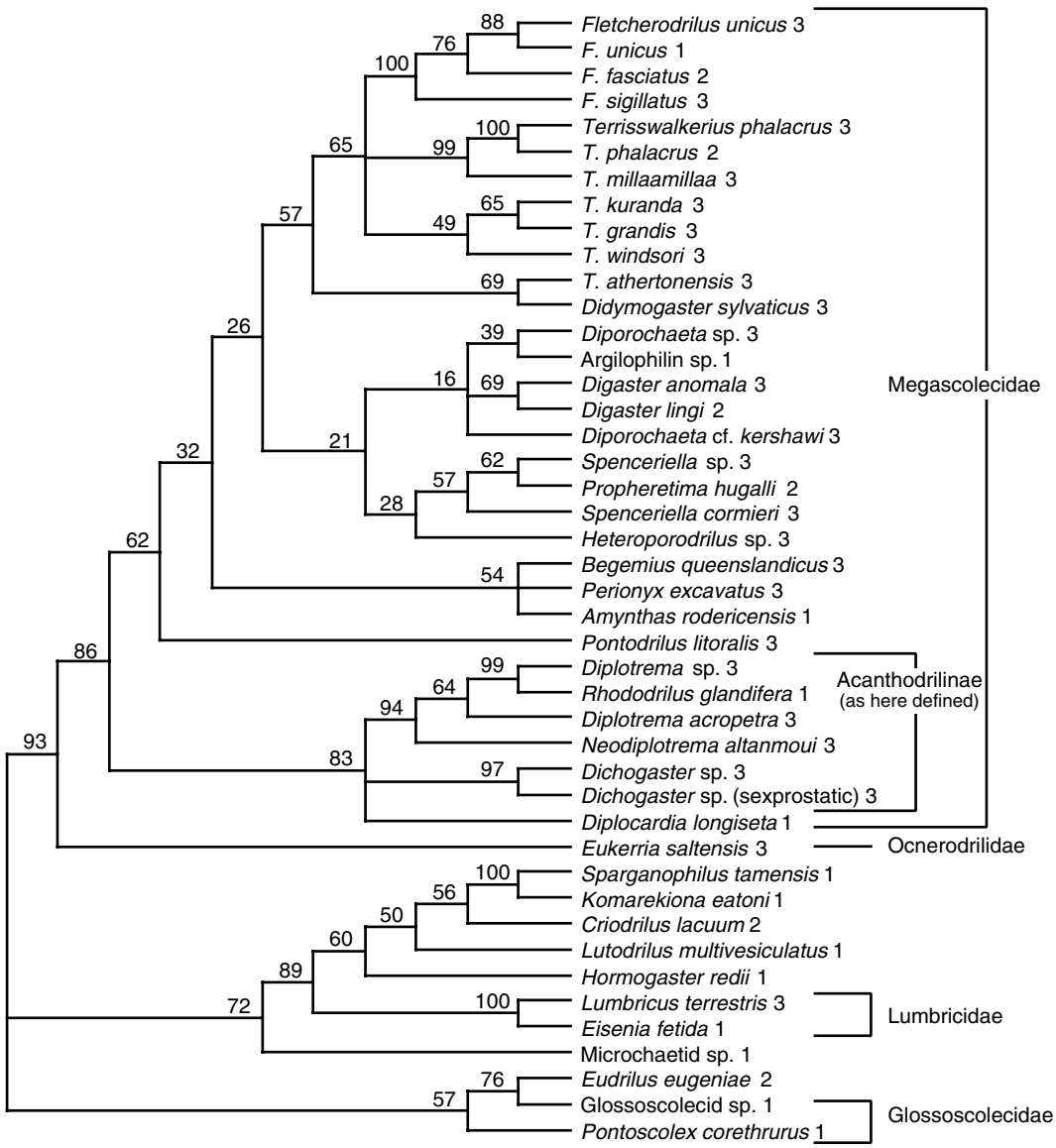

FIG. 3. — Strict consensus of six maximum parsimony trees using the combined data. Taxa with one, two or three genes indicated. Heuristic search, 1371 sites unweighted 50 random additions tree bisection reconnection with steepest decent, with bootstrap consensus values from 500 resamplings (heuristic search, unweighted, five random additions, tree bisection reconnection with steepest descent). **1**, 28S (12S only for *Rhododrilus glandifera* Jamieson, 1995); **2**, 28S + 12S; **3**, 28S + 12S + 16S.

The SUM model estimated best tree (Fig. 4) is the same as one of two equally most parsimonious trees, the other being 11.09 lnL units away according to the SUM model. The COMBO model best tree differs in two nodes: the positions of *Eukerria saltensis* and of *Pontodrilus litoralis*, at 0.28 ΔlnL according to the COMBO model.

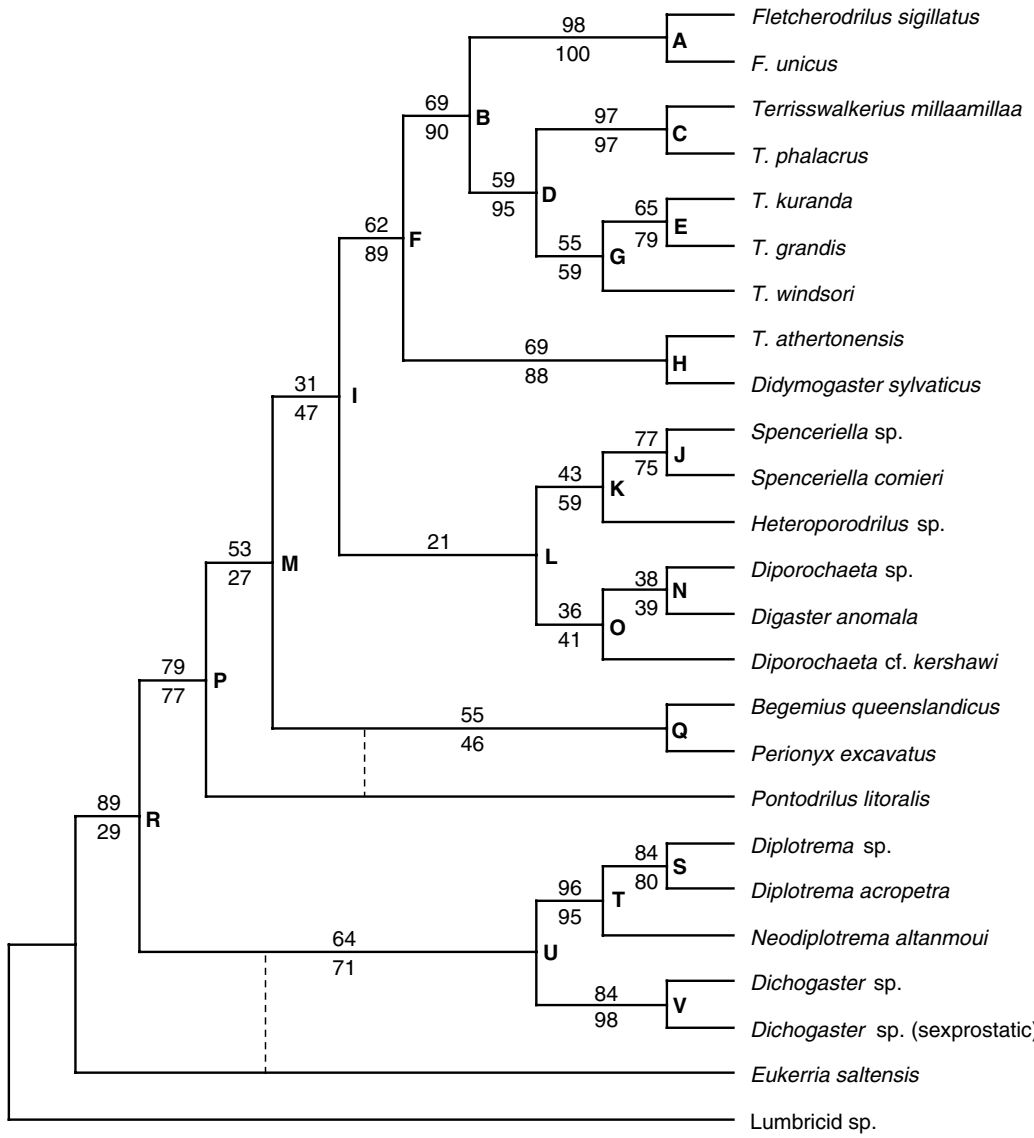

FIG. 4. — Maximum likelihood tree from the sum of the 28S rDNA and mtDNA partitions optimized for GTR-Γ model separately (SUM model best tree). Selected from 14586 near parsimonious reverse constraint trees. The maximum parsimony bootstrap tree is the same. Dashed lines indicate the two topology differences found in the combined data model maximum likelihood tree (COMBO model). Bold letters are clade labels for Table 4. Maximum parsimony and maximum likelihood boostrap values above and below branches respectively. Partition support shown in Table 4.

Effect of model design

The effect of model design (SUM versus COMBO models) as well as adding invariant sites and parameter optimization is graphically represented in the comparison of the relative lnL among reverse constraint trees (Fig. 6). These trees are a subset of the SUM method search set. The deviation in relative likelihood among trees using fixed parameters versus parameters optimised for each tree is negligible (SUM versus

TABLE 4. — Separate and combined partition support for nodes in the SUM model tree. All values refer to the support by each dataset for clades defined in Fig. 4. Partition support refers to best reverse constraint tree for that method relative to SUM model best tree COMBO model lnL split into 28S sites and mtDNA sites. Shading indicates conflict. *c*, expected likelihood weight confidence set from COMBO model reverse constraint trees and maximum likelihood best tree. Arranged in order of *c*, bold values indicate 95% confidence set.

c	Bootstrap		Bremer			Partition ΔlnL separate models			combined data analysis			
COMBO	COMBO	MP	all	mt	28S	SUM	mt	28S	all	mt	28S	clade
0.00	100	98	14	6	8	35.15	28.32	6.83	34.67	28.17	6.50	A
0.00	97	97	15	4.5	10.5	15.95	6.85	9.10	13.73	6.98	6.76	C
0.00	89	62	7	5	2	13.90	18.48	–4.59	16.62	10.59	6.03	F
0.00	98	84	10	3.7	6.3	20.82	12.12	8.70	22.54	13.37	9.16	V
0.00	88	69	10	8.5	1.5	3.09	3.09	0.00	3.78	5.31	–1.53	H
0.00	95	96	12	10	2	12.10	7.26	4.83	13.65	6.21	7.44	T
0.01	95	59	0	–4	4	11.09	–0.75	11.83	12.80	–1.27	14.07	D
0.02	46	55	4	3	1	0.99	–0.47	1.45	4.81	1.88	2.93	Q
0.02	80	84	4	3	1	0.99	0.95	0.05	2.56	3.02	–0.45	S
0.02	75	77	4	–1	5	5.88	3.26	2.61	4.09	2.42	1.67	J
0.03	90	69	4	3	1	7.13	0.67	6.46	11.08	1.56	9.52	B
0.03	59	43	4	0.4	3.6	2.75	3.00	–0.26	4.75	4.36	0.38	K
0.04	71	64	2	–2	4	4.09	2.02	2.07	3.10	0.19	2.91	U
0.04	27	53	3	1	2	0.62	0.62	0.00	–0.28	0.85	–1.13	M
0.04	29	89	10	6.7	3.3	2.75	2.75	0.00	–0.28	0.85	–1.13	R
0.06	59	55	2	–5	7	2.54	0.79	1.75	0.77	1.33	–0.56	G
0.06	23	14	2	0	2	2.23	3.14	–0.91	1.50	3.38	–1.87	L
0.07	79	65	5	7	–2	0.88	8.85	–7.97	4.56	10.83	–6.27	E
0.10	77	79	7	4	3	7.53	4.68	2.85	6.63	3.79	2.84	P
0.10	47	31	3	–1	4	5.23	5.23	0.00	5.58	5.40	0.18	I
0.10	41	36	3	0	3	3.59	2.75	0.84	5.58	5.40	0.18	O
0.17	39	38	3	–1.5	4.5	3.19	–3.40	6.60	1.53	–4.41	5.94	N
	68.4	64.4	5.8	2.3	3.5	7.4	5.0	2.4	7.9	5.0	2.9	average per node

optimized), saving a considerable computational burden. While there can be substantial difference in relative likelihood between the SUM and COMBO models (average difference 2.8 lnL), the lnL support for each node in the ML tree seems less affected (see Table 4). The COMBO model is a substantially less good fit to the data (> 350 ΔlnL, d. f. = 108, $p < 0.001$; Wilgenbusch & De Queiroz 2000), although it cannot be rejected by the Goldman (1993) parametric test as an adequate representation of the real data ($p > 0.26$, 100 replicates). Compared with this, introducing an invariant sites term in the model (as directed by standard likelihood ratio tests) produces little difference (average difference 1.1 lnL).

Partition support (Baker & DeSalle 1997) can be a guide to distribution of conflict or support. Table 4 provides such a summary of the similarities and differences between partitions and among methods for node support, providing a basis for interpreting these measures of support and conflict.

The combined data tree resembles the mtDNA tree more than it does the 28S tree for both ML

and MP (comparing the symmetric difference metric of Penny & Hendy [1985]). The mtDNA contributes more to the relative lnL but the 28S more to the Bremer score. Combined analysis, either SUM or COMBO methods, still shows conflict but also indicates hidden support: for example Fig. 4 nodes B, J, A, G – in conflict in separate analysis, are both supported in combined methods. The mtDNA has more conflict in Bremer; less in ML. Overall (as might be expected of a simpler model) the COMBO model has slightly higher levels of support and conflict than the SUM resulting in very similar proportions of conflict; MP Bremer shows a greater proportion of conflict. A reduction in conflict in ML versus MP suggests a more complex model is appropriate; reduction in conflict in the SUM versus COMBO methods may reflect reduced dynamic heterogeneity. In each method the conflict is spread across similar number of nodes 7-8/22 with the majority concentrated into a few nodes. Each method has a slightly different mix of nodes and values among methods so that overall the component parts of 28S and mtDNA vary between the two methods more than the total ΔlnL and rank order of the nodes. Of the combinations of conflict: 1) strong conflict in all methods (e.g., Fig. 4, nodes E, N), all in the same direction, may indicate historical heterogeneity; 2) difference in conflict between COMBO and SUM (e.g., Fig. 4, nodes F, G, H, M, R, S), most of which is slight (except F) may indicate dynamic heterogeneity.

Figure 5 shows the effect of different types of partition support on consensus: Fig. 5A is consensus of independent 28S and mtDNA ML trees; Fig. 5B consensus of SUM model reverse constraint trees that do not show conflict; Fig. 5C consensus of COMBO model reverse constraint trees that do not show conflict. When combined the partitions show less conflict than apart, indicating hidden support (Baker & DeSalle 1997; Lee & Hugall in press), reflected in the number of nodes resolved (11 vs 14). The SUM and COMBO models differ in the position of *Eukerria saltensis* – the SUM model is perhaps to be preferred as: a) ΔlnL and bootstrap values are weak in the COMBO model; and b) it is a more conventional arrangement – outside the Megascolecidae *s.l.*

Tests of molecular support for specific morphological classifications

The trees so far presented provide evidence against a number of morphological classifications. Overall data appear to combine so that the whole is greater than the sum of the parts but the bootstrap support for many nodes is weak. Here we test several morphological hypotheses of monophyly based on the likelihood ratio of the ML tree versus the best tree constrained to each of the alternative hypotheses. The possible genuine conflicts suggested in the partition analyses are

TABLE 5. — Comparison of likelihood ratio tests. †, best trees, fixed parameters; *, SOWH test, 100 replicates, fixed parameters; *c*¶, expected likelihood weight Strimmer & Rambaut (2002). Shading indicates rejected at $\alpha < 0.05$; dash = test not done; § 28S, Fig. 3 reduced taxa set data, Shimodaira-Hasegawa test only; 1, 28S; 2, mtDNA; 3, COMBO.

	ΔlnL			BS ML†			KH test†			SH test†			Parametric test*			*c*¶		
Hypothesis of monophyly	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Acanthodrilidae	62.6	67.5	126.3	0.00	0.00	0.00	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	–	–	–	<0.01	<0.01	<0.01
Dichogastrini	33.9	53.4	88.5	0.01	0.01	0.00	0.01	0.00	<0.01	0.10	<0.01	<0.01	–	–	–	0.01	<0.01	<0.01
Octochaetidae	7.0	14.3	21.0	0.25	0.03	0.00	0.57	0.10	0.02	0.58	0.40	0.37	0.01	<0.01	–	0.26	0.03	<0.01
Racemose Megascolecidae	16.7	19.0	31.5	0.03	0.05	0.03	0.08	0.09	0.05	0.25	0.26	0.16	<0.01	<0.01	–	0.02	0.06	0.03
Aquamegadrili §	28.2	–	–	–	–	–	–	–	–	0.032	–	–	–	–	–	–	–	–
Lumbricoidea §	35.7	–	–	–	–	–	–	–	–	0.011	–	–	–	–	–	–	–	–

not relevant to the hypotheses tested here. Generic names given in the tests represent the species listed for those genera in Table 1; author names for these genera are not therefore required. Table 5 summarizes the information on likelihood ratios tests both non-parametric (bootstrap distribution and derived KH [Kishino-Hasegawa], SH [Shimodaira-Hasegawa] and expected likelihood weight test statistics) and parametric tests (SOWH). All tests use fixed parameter values optimized to the hypothesis in question. Results in Figure 6 comparing fixed versus optimized parameters indicate that the use of the RELL (resampling estimated log-likelihood) approximation is reasonable. The following tests within the Megascolecidae use the 25 taxa in Figure 4. We may invoke congruence between the mtDNA and the 28S as additional confidence. The first two megascolecid classifications (1 and 2 below) are inconsistent with the consensus tree Figure 5A; the others are unresolved; all are inconsistent with the SUM and COMBO consensuses. Bootstrap values against these hypotheses are generally low for separate data-sets but high for the combined data; the SH test is significant for only the strongest tests while the parametric tests are significant for weakest, and therefore certainly significant for all. The expected likelihood weight 95 % confidence set rejects most.

Hypothetical taxonomic groupings tested by maximum likelihood are as follows:

1. Acanthodrilidae Vejdowski, 1884 as redefined by Gates (1959, 1972): (*Diplotrema*, *Diporochaeta* sp., *Fletcherodrilus*, *Diporochaeta kershawi*, *Pontodrilus*) versus the rest.

2. Monophyly of the Dichogastrini Jamieson, 1971: (*Dichogaster*, *Didymogaster*, *Digaster*) versus the rest.

1 and 2. Both the 28S and the mtDNA reject monophyly of the Dichogastrini Jamieson, 1971 and the Acanthodrilidae *sensu* Gates 1959.

3. Octochaetidae *sensu* Gates 1959 (Gates 1959, 1972): (*Dichogaster*, *Neodiplotrema*) versus the rest.

4. Racemose prostate Megascolecidae: (*Spenceriella*, *Didymogaster*, *Heteroporodrilus*, *Digaster*, *Begemius*, *Perionyx*) versus the rest.

The cases of the Octochaetidae and of the racemose Megascolecidae highlight the differences in methods of assessing support. Neither 28S mtDNA nor combined data trees are consistent with either hypothesis. BS support against is high (95 %, 88 % respectively) for combined data but low for partitions (e.g., 40 % in 28S); the KH test shows a trend towards significance, not apparent in the SH test; all are highly significant in the parametric test, but only 4/6 according to the expected likelihood weight. In conclusion, the Octochaetidae and Megascolecidae of Gates are not supported.

Two higher classifications within the Crassiclitellata can also be tested using the 28S data (Fig. 2):

5. Aquamegadrili Jamieson, 1988 – Terrimegadrili Jamieson, 1988: (*Sparganophilus*, *Criodrilus*, *Lutodrilus*) versus the rest.

6. Lumbricoidea: (*Lumbricus*, microchaetid, glossoscolecid, *Hormogaster*) versus the rest.

Both are rejected by the conservative SH test. The 28S data-set is inconsistent with monophyly of the Aquamegadrili (aquatic megadriles) owing to the inclusion of *Komariekiona* as sister-taxon of *Sparganophilus*. The Lumbricoidea is rejected principally because the glossoscolecid never groups with the remainder in the best trees but with *Eudrilus*.

DISCUSSION

Consensus, combining data and likelihood model design

Concerning model choice and combining of data, we applied a three-way procedure: 1) consensus of separate analyses; 2) combine data in the COMBO model; and 3) combine likelihoods in the SUM method. The separate analyses allowed each partition to be analysed under its own optimum model but the SUM method constrains the data in such a way that it chooses among combined score best trees. Multi-models represents a design "parameter" that is of more substantial influence than other parameters available (Fig. 6) but despite this we find that, in this particular case

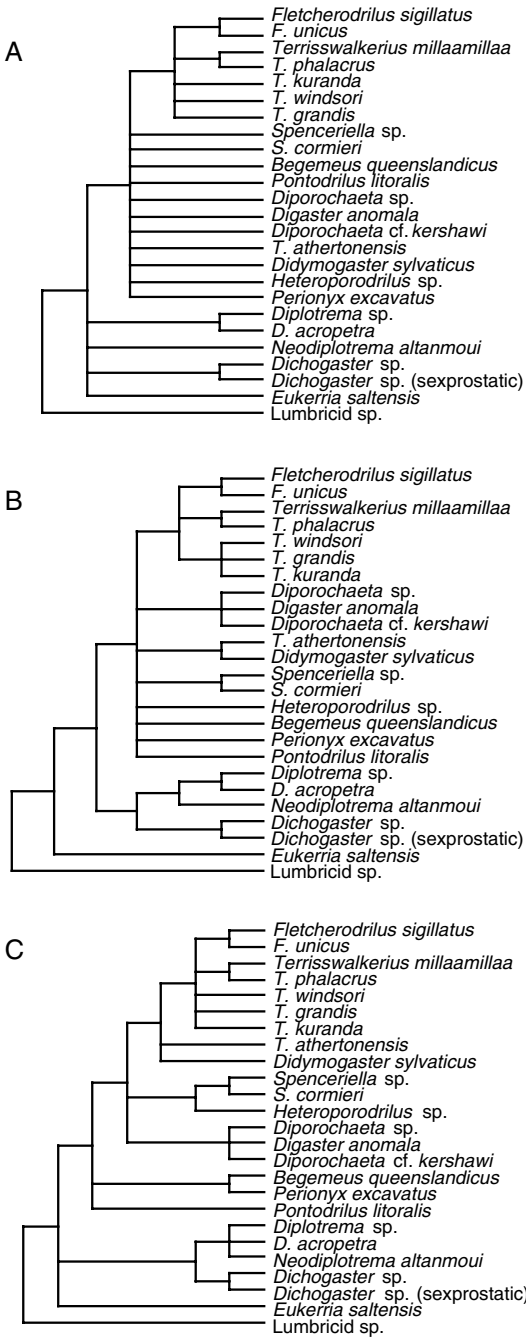

FIG. 5. — Semistrict consensus trees for the Megascolecid 28S and mtDNA data-sets; **A**, consensus of 28S and mtDNA maximum likelihood trees; **B**, consensus of SUM model reverse constraint tree showing conflict between 28S and mtDNA partitions (see Table 4); **C**, COMBO model partition conflict consensus.

at least, the results presented are insensitive to the details of the choice of model. This is especially seen in the similarity of ΔlnL between the SUM and COMBO methods, and that the ML and MP results are very similar (Table 4). The different characteristics of the nuclear and mtDNA combine to give a model more similar to the MP approach than either data-set, and the variance in the GTR-Γ model is sufficient to accommodate all sites. Congruence among the methods, and more particularly, the two data-sets (nuclear and mitochondrial) can then be invoked as support for the robustness of our findings.

By combining data at the level of tree space, the SUM method allows each partition maximum independence, without imposing any assumptions of one upon the other. An intermediate procedure would be to employ separate models for each partition but to use a combined data estimate of branch length. The general principle of combining information at the level of (relative) likelihoods of topologies provides a way of combining disparate data, providing a model. Some examples are: combining the PHYLIP likelihood method for analysing allele frequency data (allozymes, etc.); DNA-DNA hybridization data; ML method for morphology (Lewis 2001). Most of these types of data cannot directly be combined into a matrix of characters with states (even applying mixed models) but they do provide relative likelihood for topologies – the likelihood support surface. Likelihood is recognized as suited to combining disparate types of data – combining the likelihood support surface by adding up the likelihoods contributed by each type of data for each hypothesis. Further, it provides a distinct way of dealing with data-sets that are not completely matched – topologies involving taxa that are not represented will have a flat support surface, as opposed to the alternate procedure in the combined data model of integrating over all possible states (Swofford 2000). The multi-model partition analysis is inconvenient to apply because of the lack of automation. For further use and elaboration (such as bootstrapping) it needs to be computerized; apparently under development for future versions of PAUP*.

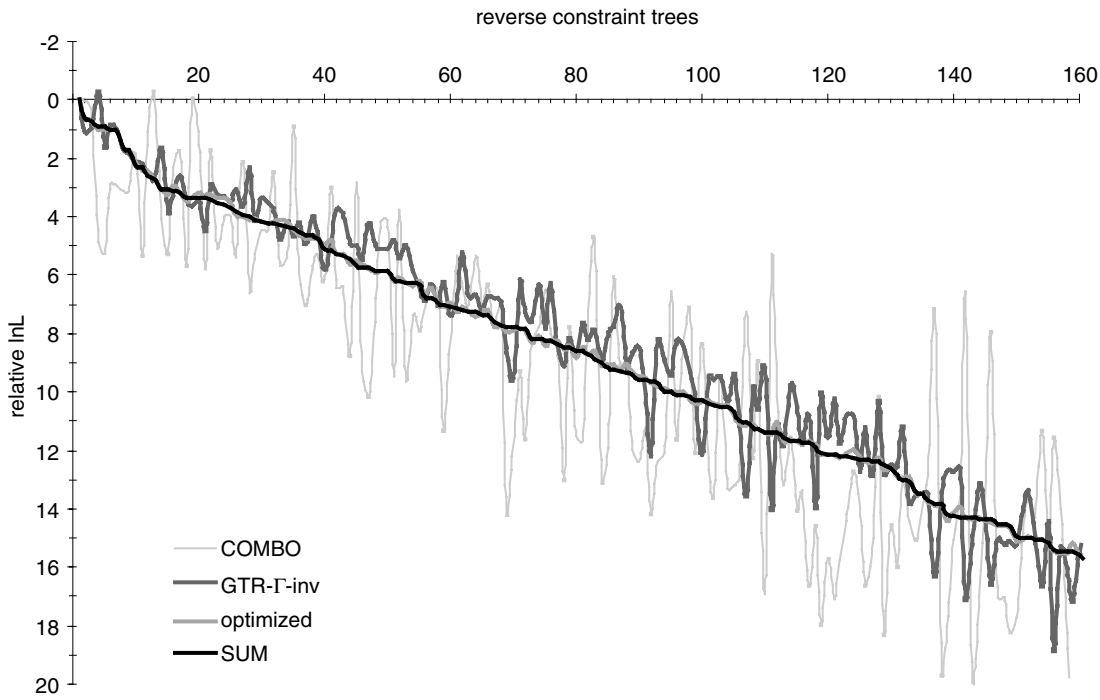

FIG. 6. — Effect of model design on likelihood support surface across reverse constraint trees. Relative lnL of a selection of near parsimonious reverse constraint trees compared to the best: SUM model tree score for the SUM model; the SUM GTR-Γ models with fixed parameters; same but with extra invariant sites parameter; SUM GTR-Γ models with optimized parameters for each topology; COMBO model fixed parameters. Line joining individual ΔlnL values is to aid visualization.

The partition analysis identifies the distribution of conflict, amongst data-sets and amongst methods: much is spread across nodes at low levels and can be dismissed as due to data limitations. If some of the conflict is due to dynamic heterogeneity this may be reduced in the SUM method, while localized consistent conflict across methods needs to be identified. Even though nodes may be reasonably supported in a combined data analysis such nodes must remain questionable (e.g., Fig. 4, node E).

Congruence among data can be powerful evidence. However, direct consensus of independent results can obscure underlying similarities. On the other hand, profound conflict can be obscured in the combined data approach. Both aspects are revealed in the consensuses representing three levels of combining information shown in Figure 5A, B, C respectively: 1) completely separate; 2) separate models but arbitrating over combined information results; and 3) combined model and data, subsequently partitioned.

Given the current state of uncertainty as to the robustness and sensitivity of phylogenetic likelihood ratios tests, we have taken the opportunity to present the ΔlnL, the of BS distribution values (proportion of resamples a tree is better than chosen ML tree, see also Stuart-Fox *et al.* [2002]), the derived KH, SH and *c* statistics, and the parametric (SOWH) test (Table 5). The uncertainty hinges on plausibility of null hypotheses for non-parametric tests and on how well the model fits the data for the parametric test. We explored this in Table 5, and Figure 6 together with likelihood ratio tests. As most of the model lies in the data patterns, the fixed parameter approximation is worth the trivial sacrifice in precision; as the 28S and mtDNA are very different, the multi-model (SUM) method of considerable gain. While the GTR-Γ model may be considered optimized, the

difference between parametric versus non-parametric tests (Table 5), suggests it is still too far from reality to justify parametric methods. Here the most noticeable difference between the parametric and the real data is distribution of apomorphies – more clustered in real data suggesting lack of independence among sites as assumed by the model (not shown). All the other approaches use non-parametric pseudoreplicates, which always have fewer patterns – the principle determinant of the likelihood model. The KH test has been criticized as unfair to the null hypothesis (Goldman *et al.* 2000) but on the other hand the SH test can be weakened to the point of futility by including highly unlikely trees (Goldman *et al*. 2000; Strimmer & Rambaut 2002). The expected likelihood weighting essentially reflects the primary BS ΔlnL distribution (the simplest form of all these tests) and captures most of the consensus of information (Tables 4; 5).

SIGNIFICANCE TO TAXONOMY AND MORPHOLOGICAL CLASSIFICATIONS

Crassiclitellata Jamieson, 1988

That the Crassiclitellata of Jamieson (1988), named for families in which the clitellum is multilayered, is a monophylum is confirmed in the molecular analyses. The MP majority rule consensus tree for 549 bp of 28S only (Fig. 2) gives 100% bootstrap support for the Crassiclitellata versus outgroup taxa. Maximum likelihood analysis differs little from a MP bootstrap tree. *Eudrilus* again associates with the glossoscolecid Gen. sp. but not with *Microchaetus*.

Aquamegadrili Jamieson, 1988 and Lumbricoidea sensu *Jamieson 1971*

The Aquamegadrili, named by Jamieson (1988) for aquatic crassiclitellates (see Introduction for geographic distribution), consist of the Sparganophilidae, Biwadrilidae, Almidae (mostly warm tropics, including *Criodrilus*, Mediterranean region, etc.) and Lutodrilidae (Southern Neartic). In the likelihood ratio tests the 28S data-set is inconsistent with monophyly of the Aquamegadrili but this is entirely due to the inclusion of *Komarekiona* as sister-taxon of *Sparganophilus.* The 28S data are inconsistent with partition of crassiclitellates into Aquamegadrili and Terrimegadrili (see families listed in Introduction); instead the representatives (*Sparganophilus*, *Criodrilus* and *Lutodrilus*) of the original aquamegadrile taxa, with *Komarekiona*, lie within a paraphyletic Terrimegadrili (Figs 2; 3). The Lumbricoidea is incompatible with the 28S data but principally because the glossoscolecid sp. never groups with the remainder in the best trees but groups equivocally with *Eudrilus*. However the unity of the majority of the remaining Lumbricoidea will necessarily require more sampling.

Eudrilidae Claus, 1880

The family Eudrilidae has formerly been associated with the Megascolecidae and Ocnerodrilidae in a superfamily Megascolecoidea (Jamieson 1978) or in a separate superfamily Eudriloidea as the sister-group of the Lumbricoidea + Megascolecoidea *s.s.*, in the morphocladistic analysis (Jamieson 1988). In the present analysis, *Eudrilus* always has the glossoscolecid Gen. sp. (a lumbricoid *sensu* Jamieson 1978, 1988) as its sister-taxon, with moderate BS support of 63-84 % (MP, ML, Figs 2; 3). The other glossoscolecid, *Pontoscolex*, may or may not link with these. There is no molecular support here for regarding eudrilids as the unique sister-group of the Ocnerodrilidae + Megascolecidae assemblage. Omodeo (2000) derived eudrilids independently (from alluroidids), thus also noting their distinctness but that origin goes contrary to the present confirmation of the monophyletic nature of the Crassiclitellata.

Megascolecoidea (Megascolecidae Rosa, 1891 + Ocnerodrilidae Beddard, 1891)

In the morphocladistic analysis (Jamieson 1988), the superfamily Megascolecoidea contained only the Megascolecidae, with the two subfamilies Acanthodrilinae and Megascolecinae, and excluded the Ocnerodrilidae and Eudrilidae. However, all three of these families had been tentatively included in the superfamily Megascolecoidea in Jamieson (1978, 1980). Recognition by Lee (1959), Jamieson (1971a-c, 1978, 1980),

and Sims (1966, 1967) of an Acanthodrilinae + Megascolecinae assemblage together with the Ocnerodril(-inae) (-idae) is endorsed in the present molecular analyses. Monophyly of the ocnerodrile exemplar *Eukerria saltensis* with the Megascolecidae (Acanthodrilinae + Megascolecinae) is supported in all cladistic analyses that have additional outgroups. This relationship of the Ocnerodrilidae with the Megascolecidae has > 90% BS support in MP trees (see Figs 2; 4, also Jamieson 2000). In view of the sister-group relationship of the *Eukerria saltensis* with the Megascolecidae *s.l.*, it appears that the family Ocnerodrilidae may be included in the Megascolecoidea, rather than having a suprafamilial rank of its own, but this needs wider sampling to test monophyly, and the relationships of the two ocnerodrile tribes Ocnerodrilini and Malabarini.

Megascolecidae sensu *Gates 1972*

With regard to previous classifications of the Megascolecidae the following conclusions can be drawn from the present molecular analyses. The widely used system for internal classification of the Megascolecidae of Gates (1959, 1972) cannot be sustained, as already argued by Lee (1959, 1970) and Jamieson (1971a-c). Diagnosis of Megascolecidae in the restricted sense of Gates (1972) by the possession of racemose prostates (with holo- or meronephridia) is not supported in any MP trees, with multiple evolution of racemose and tubular prostates implied by the molecular phylogenies (Figs 3; 4). However, bootstrap values are low in the relevant section of the trees and, as no species with tubular prostates has a high bootstrap linkage with a species with racemose prostates [notwithstanding the *Terrisswalkerius athertonensis-Didymogaster* clade]. Likelihood ratio tests are equivocal on this but we suggest this is more a reflection on the SH test as overly conservative.

Octochaetidae sensu *Gates 1972 and Acanthodrilidae* sensu *Gates 1972*

The Octochaetidae were defined by Gates (1972) and, as the Octochaetinae, by Gates (1959) and Sims (1967), as all species with tubular prostates and more than one pair of nephridia per segment (meronephridia). Gates (1959, 1972) diagnosed the Acanthodril(-inae), (-idae), as having tubular prostates and a pair of nephridia (holonephridia) per segment. The relationship of the meronephric *Neodiplotrema* with the holonephric *Diplotrema* in all trees (either 28S, mtDNA or combined) argues against recognition of the Octochaet(-idae), (inae), and against the Acanthodril(-inae), (-idae) of Gates and of Sims. The *Neodiplotrema* + *Diplotrema* clade fully endorses the conclusion by Lee (1959, 1970) that phylogenetic pairs of holonephric with meronephric species of Acanthodrilinae are recognizable. We will now further consider subdivision of the Megascolecidae into the subfamilies Acanthodrilinae and Megascolecinae and division of the Megascolecinae into the tribes Perionychini, Dichogastrini, and Megascolecini in the classification of Jamieson (1971a-c).

Acanthodrilinae sensu *Jamieson 1971a*

The definition of the subfamily Acanthodrilinae *sensu* Jamieson (1971a) differs fundamentally from that of Gates (1959, 1972). Prostates are not only tubular but may also (rarely) be racemose and nephridia are not only holonephridia but may also be meronephridia. Unlike many acanthodrilids of Gates, the prostates usually do not discharge on segment 18 (doing so in *Rhododrilus*), typically opening on segments 17 and 19, as in the type-genus. Megascolecinae with homeotic displacement of male pores may correspond with this definition but show their affinities with megascolecines in other respects. Posterior nephridia in Acanthodrilinae *sensu* Jamieson (1971a) lacked the median funnel diagnostic of the Dichogastrini within the Megascolecinae *sensu* Jamieson. The alimentary and vascular systems differed from those of the Ocnerodrilinae in some of which, as in *Eukerria* Michaelsen, 1935, the male and prostate pores have the acanthodrilin arrangement. It has been shown in the molecular analysis that dichogastrins with acanthodrilin male pores must be transferred to the Acanthodrilinae and that the dichogastrin nephridial condition has arisen more than once (see below).

Megascolecinae

Perionychini Jamieson, 1971. The tribe Perionychini (defined by megascolecin male pores and holonephridia, irrespective of prostate type) is represented here by five genera. There is no convincing support in any analysis for a unified or monophyletic Perionychini although there is evidence that some of these holonephric megascolecines are closely related. There is, however, much instability in perionychin relationships, reflected in low BS values and lack of consensus (Figs 3-5). In all analyses the "perionychin" (acanthodrile *sensu* Gates) *Pontodrilus* diverges at or near the base of the Megascolecinae *s.l.*, and while exclusion from the Acanthodrilinae is upheld, its uncertain placement reflects its enigmatic affinities on morphological and ecological grounds. *Pontodrilus* is highly unusual among crassiclitellates in being euryhaline. It may be suspected of having had a long independent evolution. The tribe Perionychini, although a convenient taxonomic grouping, is thus a para- or polyphyletic assemblage in the molecular analyses. This has previously been suspected (Jamieson 1988) as the Perionychini is recognized on the basis of the symplesiomorphic possession of holonephridia (a condition seen throughout the Oligochaeta, whereas meronephry is virtually limited to Megascolecidae). It is therefore a grade rather than a clade.

Dichogastrini Jamieson, 1971 and Acanthodrilinae Vejdovsky, 1884. The Dichogastrini were defined by presence of a single stomate meronephridium median to astomate micromeronephridia on each side in caudal segments, in the absence of posterior enteronephry (Jamieson 1971a). It has, however, repeatedly been questioned (e.g., Jamieson 1978, 1981; Dyne 1984) that dichogastrins with acanthodrilin male pores (here represented by *Dichogaster*) are monophyletic with those with megascolecin pores (e.g., *Digaster*).

The present analyses relegate "acanthodrilin dichogastrins" (*Dichogaster* and *Neodiplotrema*) to the Acanthodrilinae and "megascolecin dichogastrins" (albeit represented only by *Digaster* and *Didymogaster*) to the Megascolecinae *s.l.* in combined and separate analyses for both mtDNA and 28S with overall good support (e.g., BS support of 77% in the combined ML tree, Fig. 4).

The Nearctic *Diplocardia longiseta* appears to lie within this "acanthodrilin" clade (83% BS, Fig. 3) but the only available data (the 28S) are not sufficient for further resolution and more taxa are required. *Diplocardia* Garman, 1888 has been shown by James (1990) to be closely similar in morphology to *Diplotrema*. The available 12S data support the argument (Jamieson 1995) that *Rhododrilus glandifera*, in the Wet Tropics of Queensland, is locally derived from a precursor with the acanthodrilin arrangement of male pores (probably *Diplotrema* with which it has a 99% BS value in Fig. 3) though this requires confirmation from analysis of larger numbers of sequences. *R. glandifera* thus appears to deserve a subgeneric rank in *Diplotrema* or generic rank separately from *Rhododrilus*, the type-locality of which is in New Zealand.

All trees from combined data endorse recognition of the Acanthodrilinae for worms with acanthodrilin male pores, including acanthodrilin Dichogastrini (*Dichogaster*) (though there is lack of resolution for 28S, Fig. 2) but not those ocnerodriles (*Eukerria*) with acanthodrilin or other male terminalia. The ocnerodriles (albeit represented only by *Eukerria*) are phylogenetically distinct in the present study and are well-defined morphologically.

Megascolecini *sensu* Jamieson 1971. The third tribe of the Megascolecidae, the Megascolecini, was defined by having male and prostate pores coincident on segment 18 (rarely segment 17), and meronephry in which a median stomate nephridium, if present, differed from those of dichogastrins in opening into the intestine (enteronephry). Prostates were racemose, tubular or tubuloracemose (Jamieson 1971a-c). In contrast, Gates (1959, 1972) attributed only worms with racemose prostates, irrespective of nephridial types, to his restricted Megascolecidae.

Resolution of the Megascolecini was not an aim of the present work and as few representatives have been included (*Amynthas* Kinberg, 1867, *Begemius*

Easton, 1982, *Propheretima* Jamieson, 1995, and *Spenceriella* Michaelsen, 1907) results must be regarded with caution. However, none of the analyses supports retention of the Megascolecini as defined by Jamieson (1971a-c). It is to be expected that the criteria of meronephry with enteronephry may have evolved more than once. A core of megascolecin genera, including among others *Begemius*, and *Amynthas*, is suspected to be monophyletic however (Jamieson 1981), though it no longer appears that *Spenceriella* is as close to the pheretimoids as previously argued.

Relevance to Clitellata (oligochaete, leech and branchiobdellid) relationships

Morphological and molecular support for paraphyly or polyphyly of the Oligochaeta and inclusion within this group of leeches and branchiobdellids has been outlined in the Introduction. Our 28S data are consistent with a leech-branchiobdellid-lumbriculid clade, within the Oligochaeta, less so for a leech-branchiobdellid grouping, considering the variation among methods. Although long branch/rate acceleration/base content has posed problems for the nuclear ribosomal genes analyses, consistency with data of different characteristics (mtDNA and COI) underlines these relationships. The lumbriculid relationship was proposed on morphological grounds by Michaelsen (1928-1932), Brinkhurst & Nemec (1986), Brinkhurst & Gelder (1989), Brinkhurst (1999a) and, less certainly, by Brinkhurst (1999b) and from molecular data by Siddall & Burreson (1998), Martin (2001) and Siddall *et al.* (2001). If accepted, the present phylogeny would confirm the Oligochaeta as a paraphyletic group as previously mooted (Jamieson *et al.* 1987; Jamieson 1988; Martin 2001), being merely the non-leech, non-branchiobdellidan clitellates (Purschke *et al.* 1993). Inclusion of leeches and branchiobdellids within the Oligochaeta would thus render the name Oligochaeta synonymous with Clitellata (or Euclitellata of Jamieson 1983), as proposed by Siddall *et al.* (2001) and in a study of molecular phylogeny of the Tubificidae (subsuming the Naididae) by Erséus *et al.* (2002).

Acknowledgements

We are grateful to the following colleagues who freely donated material used in this study: Ralph Brinkhurst, Mark Harvey, Alan Hodgson, Pierre Martin, Craig Moritz, Pietro Omodeo and Emilia Rota. Much material was collected by the authors. In addition to sequencing by the authors in Brisbane and Paris, some was performed by David Scheltinga and Katrina McGuigan. The courtesy of Mark Siddall in allowing us to see the typescript of the submitted paper on clitellate molecular phylogenetics is greatly appreciated. We thank Mike Lee and Stuart Baird for discussions on phylogenetic analyses, and C. Cunningham and P. Bouchet for improvements to the manuscript. The encouragement and provision of facilities by Professor Craig Moritz is also most gratefully acknowledged, as is provision of funding by the Australian Research Grants Council to BGMJ.

REFERENCES

Adachi J. & Hasegawa M. 1992. — MOLPHY: Programs for molecular phylogenetics I-PROTML: Maximum likelihood inference of protein phylogeny. *Computer Science Monograph* 27. Institute of Statistics and Mathematics, Tokyo.

Baker R. H. & DeSalle R. 1997. — Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *Systematic Biology* 46: 654-673.

Brinkhurst R. O. 1999a. — Retrospect and prospect: Reflections on forty years of study of aquatic oligochaetes. *Hydrobiologia* 406: 9-19.

Brinkhurst R. O. 1999b. — Lumbriculids, branchiobdellidans and leeches: An overview of recent progress in phylogenetic research on clitellates. *Hydrobiologia* 406: 281-290.

Brinkhurst R. O. & Gelder S. R. 1989. — Did the lumbriculids provide the ancestors of the branchiobdellidans, acanthobdellidans and leeches? *Hydrobiologia* 180: 7-15.

Brinkhurst R. O. & Nemec F. L. 1986. — A comparison of phenetic and phylogenetic methods applied to the systematics of the Oligochaeta. *Hydrobiologia* 155: 65-74.

Bromham L. D & Degnan B. M. 1999. — Hemichordates and deuterostome evolution: robust molecular phylogenetic support for a hemichordate + echinoderm clade. *Evolution and Development* 1: 166-171.

CRANDALL K. A., HARRIS D. J. & FELZNER J. W. JR. 2000. — The monophyletic origin of freshwater crayfish estimated from nuclear and mitochondrial DNA sequences. *Proceedings of the Royal Society of London* B Biological Sciences 267: 1679-1686.

DE RIJK P., ROBBRECHT E., DE HOOG S., CAERS A., VAN DE PEER Y. & DE WACHTER R. 1999. — Database on the structure of large subunit ribosomal RNA. *Nucleic Acids Research* 27: 174-178.

DOLPHIN K., BELSHAW R., ORME C. D. L. & QUICKE D. L J. 2000. — Noise and incongruence: interpreting results of the incongruence length difference test. *Molecular Phylogenetics and Evolution* 17: 401-406.

DYNE G. R. 1984. — *Systematics and Zoogeography of Australian Megascolecid Earthworms.* Unpublished Ph. D. thesis, University of Queensland, Brisbane, Australia, 500 p.

EDWARDS A. W. F. 1972. — *Likelihood; an account of the statistical concept of likelihood and its application to scientific inference*. Cambridge University Press, Cambridge, xv+235 p.

ERSÉUS C., KALLERSJO M., EKMAN M. & HOVMOLER R. 2002. — 18S rDNA phylogeny of the Tubificidae (Clitellata) and its constituent taxa: dismissal of the Naididae. *Molecular Phylogenetics and Evolution* 22: 414-422.

FARRIS J. S., KALLERSJO M., KLUGE A. G. & BULT C. 1995. — Testing significance of incongruence. *Cladistics* 10: 315-319.

FELSENSTEIN J. 1995. — *PHYLIP (Phylogeny Inference Package) Version 3.5c*. Department of Genetics, University of Washington, Seattle.

FLOOK P. K., KLEE S. & ROWELL C. H. F. 1999. — Combined molecular phylogenetic analysis of the Orthoptera (Arthropoda, Insecta) and implications for their higher systematics. *Systematic Biology* 48: 233-253.

GATES G. E. 1959. — On a taxonomic puzzle and the classification of the earthworms. *Bulletin of the Museum of Comparative Zoology Harvard* 121: 229-261.

GATES G. E. 1972. — Burmese earthworms. *Transactions of the American Philosophical Society* n. ser. 62: 1-326.

GOLDMAN N. 1993. — Statistical test of models of DNA substitution. *Journal of Molecular Evolution* 36: 182-198.

GOLDMAN N., ANDERSON J. P. & RODRIGO A. G. 2000. — Likelihood-based tests of topologies in phylogenetics. *Systematic Biology* 49: 652-670.

HICKSON R. E., SIMON C., COOPER A., SPICER G. S., SULLIVAN J. & PENNY D. 1996. — Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. *Molecular Biology and Evolution* 13: 150-169.

HUELSENBECK J. P. 1998. — Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Systematic Biology* 47: 519-537.

HUELSENBECK J. P. & BULL J. J. 1996. — A likelihood ratio test to detect conflicting phylogenetic signal. *Systematic Biology* 45: 92-98.

HUELSENBECK J. P. & HILLIS D. M. 1996. — Parametric bootstrapping in molecular phylogenetics: applications and performance, *in* FERRARIS J. D. & PALUMBI S. R. (eds), *Molecular Zoology: Advances, Strategies, and Protocols*. Wiley-Liss, New York: 19-45.

HUELSENBECK J. P. & RANNALA B. 1998. — Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* 276: 227-232.

JAMES S. W. 1990. — *Diplotrema murchei* and *D. papillata*, new earthworms (Oligochaeta: Megascolecidae) from Mexico. *Acta Zoológica Mexicana*, Nueva Serie 0 (38): 18-27.

JAMIESON B. G. M. 1971a. — A review of the megascolecoid earthworm genera (Oligochaeta) of Australia. Part I - Reclassification and checklist of the megascolecoid genera of the world. *Proceedings of the Royal Society of Queensland* 82: 75-86.

JAMIESON B. G. M. 1971b. — A review of the megascolecoid earthworm genera (Oligochaeta) of Australia. Part II - The subfamilies Ocnerodrilinae and Acanthodrilinae. *Proceedings of the Royal Society of Queensland* 82: 95-108.

JAMIESON B. G. M. 1971c. — A review of the megascolecoid earthworm genera (Oligochaeta) of Australia. Part III - The subfamily Megascolecinae. *Memoirs of the Queensland Museum* 16: 69-102.

JAMIESON B. G. M. 1978. — Phylogenetic and phenetic systematics of the opisthoporous Oligochaeta (Annelida: Clitellata). *Evolutionary Theory* 3: 195-233.

JAMIESON B. G. M. 1980. — Preliminary discussion of an Hennigian analysis of the phylogeny and systematics of opisthoporous oligochaetes. *Revue d'Écologie et de Biologie du Sol* 17: 261-275.

JAMIESON B. G. M. 1981. — Historical biogeography of Australian Oligochaeta, *in* KEAST A. (ed.), *Ecological Biogeography of Australia*. W. Junk, The Hague: 887-921

JAMIESON B. G. M. 1983. — Spermiogenesis in the oligochaetoid polychaete *Questa* (Annelida, Questidae). *Zoologica Scripta* 12: 179-186.

JAMIESON B. G. M. 1988. — On the phylogeny and higher classification of the Oligochaeta. *Cladistics* 4: 367-410.

JAMIESON B. G. M. 1992. — Oligochaeta, *in* HARRISON F. W. (ed.), *Microscopic Anatomy of Invertebrates*. Volume 7. *Annelida*. HARRISON F. W. & GARDINER S. L. (eds). Wiley-Liss, New York: 217-322.

JAMIESON B. G. M. 1995. — New species and a new genus of earthworms in the collections of the Queensland Museum (Megascolecidae: Oligochaeta). *Memoirs of the Queensland Museum* 38: 575-596.

JAMIESON B. G. M. 2000. — *Native Earthworms of Australia (Megascolecidae, Megascolecinae)*. Science Publishers, Enfield, New Hampshire, USA (CD ROM).

JAMIESON B. G. M., ERSÉUS C. & FERRAGUTI M. 1987. — Parsimony analysis of the phylogeny of some Oligochaeta (Annelida) using spermatozoal ultrastructure. *Cladistics* 3: 141-155.

KISHINO H. & HASEGAWA M. 1989. — Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *Journal of Molecular Evolution* 29: 170-179.

KOJIMA S. 1998. — Paraphyletic status of Polychaeta suggested by phylogenetic analysis based on the amino acid sequences of elongation factor-1 alpha. *Molecular Phylogenetics and Evolution* 9: 255-261.

KRAJEWSKI C., FAIN M. G., BUCKLEY L. & KING D. G. 1999. — Dynamically heterogeneous partitions and phylogenetic inference: an evaluation of analytical strategies with Cytochrome *b* and ND6 gene sequences in Cranes. *Molecular Phylogenetics and Evolution* 13: 302-313.

LEE K. E. 1959. — The earthworm fauna of New Zealand. *New Zealand Department of Scientific and Industrial Research Bulletin* 130: 1-486.

LEE K. E. 1970. — Application of numerical taxonomy to the classification of Megascolecidae (Annelida: Oligochaeta). *Pedobiologia* 10: 257-264.

LEE M. S. Y. 2001. — Uninformative characters and apparent conflict between molecules and morphology. *Molecular Biology and Evolution* 18: 676-680.

LEE M. S. Y. & HUGALL A. F. in press. — Partitioned likelihood support and the evaluation of data set conflict. *Systematic Biology*.

LEWIS P. O. 2001. — A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913-925.

LOCKHART P. J., STEEL M. A., HENDY M. D. & PENNY D. 1994. — Recovering evolutionary trees under a more realistic model of sequence evolution. *Molecular Biology and Evolution* 11: 605-612.

MADDISON W. P. & MADDISON D. R. 1995. — *MacClade: Analysis of Phylogenetic Character Evolution. Version 3.07*. Sinauer Associates, Sunderland, MS.

MARTIN P. 2001. — On the origin of the Hirudinea and the demise of the Oligochaeta. *Proceedings of the Royal Society of London* B Biological Sciences 268 (1471): 1089-1098.

MARTIN P., KAYGORODOVA I., SHERBAKOV D. Y. & VERHEYEN E. 2000. — Rapidly evolving lineages impede the resolution of phylogenetic relationships among Clitellata (Annelida). *Molecular Phylogenetics and Evolution* 15: 355-368.

MCHUGH D. 2000. — Molecular phylogeny of the Annelida. *Canadian Journal of Zoology* 78: 1873-1884.

MICHAELSEN W. 1928-1932. — Oligochaeta, *in* KUKENTHAL W. & KRUMBACH T. (eds), *Handbuch der Zoologie*, 2 (8). Walter de Gruyter, Berlin: 1-118.

MICHAELSEN W. 1933. — Die Oligochätenfauna Surinames mit Erörterung der verwanschaftlichen und geographischen Beziehungen der Octochätinen. *Tijdschrift der Nederlandsche dierkundige vereeniging, Leiden* 3: 112-131.

NYLANDER J. A. A., ERSÉUS C. & KALLERSJO M. 1999. — A test of monophyly of the gutless Phallodrilinae (Oligochaeta, Tubificidae) and the use of a 573-bp region of the mitochondrial cytochrome oxidase I gene in analysis of annelid phylogeny. *Zoologica Scripta* 28: 305-313.

OMODEO P. 1958. — La réserve naturelle intégrale du Mont Nimba. *Mémoires de l'Institut français d'Afrique noire* 53: 9-109.

OMODEO P. 2000. — Evolution and biogeography of megadriles (Annelida, Clitellata). *Italian Journal of Zoology* 67: 179-201.

PALUMBI S. A. 1996. — Nucleic Acids II: The Polymerase Chain Reaction, *in* HILLIS D. M., MORITZ C. & MABLE B. K. (eds), *Molecular Systematics*, 2nd edition. Sinauer, Sunderland, Mass.: 205-247.

PENNY D. & HENDY M. D. 1985. — The use of tree comparison metrics. *Systematic Zoology* 34: 75-82.

PICKFORD G. E. 1937. — *A Monograph of the Acanthodriline Earthworms of South Africa*. Heffer, Cambridge, 612 p.

POL D. & SIDDALL M. E. 2001. — Biases in maximum likelihood and parsimony: a simulation approach to a 10-taxon case. *Cladistics* 17: 266-281.

POSADA D. & CRANDALL K. A. 1998. — MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

PURSCHKE G., WESTHEIDE W., ROHDE D. & BRINKHURST R. O. 1993. — Morphological reinvestigation and phylogenetic relationship of *Acanthobdella peledina* (Annelida, Clitellata). *Zoomorphology* 113: 91-101.

RAMBAUT A. & GRASSLY N. C. 1997. — Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in Biosciences* 13: 235-238.

ROTA E., MARTIN P. & ERSÉUS C. 2001. — Soil-dwelling polychaetes: enigmatic as ever? Some hints on their phylogenetic relationships as suggested by a maximum parsimony analysis of 18S rDNA gene sequences. *Contributions to Zoology* 70: 127-138.

SHIMODAIRA H. & HASEGAWA M. 1999. — Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* 16: 1114-1116.

SIDDALL M. E., APAKUPAKUL K., BURRESON E. M., COATES K. A., ERSÉUS C., GELDER S. R., KALLERSJO M. & TRAPIDO-ROSENTHAL H. 2001. — Validating Livanow: Molecular data agree that leeches, branchiobdellidans, and Acanthobdella peledina form a

monophyletic group of oligochaetes. *Molecular Phylogenetics and Evolution* 21: 346-351.

SIDDALL M. E. & BURRESON E. M. 1998. — Phylogeny of leeches (Hirudinea) based on mitochondrial cytochrome c Oxidase subunit 1. *Molecular Phylogenetics and Evolution* 9: 156-162.

SIMS R.W. 1966. — The classification of the megascolecid earthworms: an investigation of oligochaete systematics by computer techniques. *Proceedings of the Linnean Society of London* 177: 125-141.

SIMS R.W. 1967. — Earthworms (Acanthodrilidae and Eudrilidae) from Gambia. *Bulletin of the British Museum of Natural History* 16: 1-43.

STEPHENSON J. 1930. — *The Oligochaeta*. Clarendon Press, Oxford, 978 p.

STRIMMER K. & RAMBAUT A. 2002. — Inferring confidence sets of possibly misspecified gene trees. *Proceedings of the Royal Society of London* B Biological Sciences 269: 137-142.

STUART-FOX D., HUGALL A. F. & MORITZ C. 2002. — A molecular phylogeny of rainbow skinks (Scincidae: *Carlia*): taxonomic and biogeographic implications. *Australian Journal of Zoology* 50: 39-51.

SWOFFORD D. L. 2000. — *PAUP*. Phylogenetic Analysis Using Parsimony and Other Methods*. Illinois Natural History Survey, Champaign, IL.

SWOFFORD D. L., OLSEN G. J., WADDELL P. J. & HILLIS D. M. 1996. — Phylogenetic inference, *in* HILLIS D. M., MORITZ C. & MABLE B. K. (eds), *Molecular Systematics*, 2nd edition. Sinauer Associates, Sunderland Mass.: 407-514.

TAKAHASHI K. & NEI M. 2000. — Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution* 17: 1251-1258.

THOMPSON J. D., HIGGINS D. G. & GIBSON T. J. 1994. — CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.

WILGENBUSCH J. & DE QUEIROZ K. 2000. — Phylogenetic relationships among the Phrynosomatid Sand Lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. *Systematic Biology* 49: 592-612.

WINNEPENNINCKX B. M. H., BACKELJAU T. & DE WACHTER R. 1993. — Extraction of high molecular weight DNA from molluscs. *Trends in Genetics* 9: 407.

WINNEPENNINCKX B. M. H., VAN DE PEER Y. & BACKELJAU T. 1998. — Metazoan relationships on the basis of 18S rRNA sequences: A few years later... *American Zoologist* 38: 888-906.

YANG Z. 1994. — Estimating the pattern of nucleotide substitution. *Journal of Molecular Evolution* 39: 105-111.

YANG Z. 1996. — Maximum-Likelihood models for combined analyses of multiple sequence data. *Journal of Molecular Evolution* 42: 587-596.

YANG Z. 1997. — How often do wrong models produce better phylogenies. *Molecular Biology and Evolution* 14: 105-108.

YANG Z., GOLDMAN N. & FRIDAY A. 1995. — Maximum likelihood tree from DNA sequences: a peculiar statistical estimation problem. *Systematic Biology* 44: 384-399.

Submitted on 21 June 2002;
accepted on 30 September 2002.