Filogènia Molecular dels Bilaterals:
una aproximació multigènica

Jordi Paps

Tesi Doctoral

2008

UNIVERSITAT DE BARCELONA

Departament de Genètica

Programa de Genètica

Bienni 2001-2003

Filogènia Molecular dels Bilaterals: una aproximació multigènica

Memòria presentada per

Jordi Paps Montserrat

Per optar al grau de

Doctor

Barcelona Juliol del 2008

La directora   L'autor

Dra. Marta Riutort León   Jordi Paps Montserrat

Resultats

Capítol I

A phylogenetic analysis of Myosin Heavy Chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians

I. Ruiz-Trillo*, J. Paps*, M. Loukota†, C. Ribera†, U. Jondelius‡, J. Baguñà*, and M. Riutort*

*Departament de Genètica and †Departament de Biologia Animal, Universitat de Barcelona, Av. Diagonal, 645 08028 Barcelona, Spain

‡Evolutionary Biology Centre, University Uppsala, Norbyvägen 18D, 752 36 Uppsala, Sweden

Resum

Aquest article és el primer fruït del nostre objectiu de cercar nous marcadors moleculars útils a l'hora de resoldre la filogènia dels bilaterals. Els gens alternatius al 18S que s'havien fet servir fins a la data (p.ex. l'Histona H3, l'Elongation Factor 1-alpha o la subunitat petita de la RNA Polimerasa II) no eren prou resolutius per si sols; així que calia trobar nous gens, obtenir una mostra representativa dels bilaterals i avaluar si contenien prou informació per inferir la filogènia d'aquest grup. La primera molècula que vam obtenir és la que es presenta en aquest capítol: la cadena pesant de la miosina de tipus II.

Es va iniciar un esforç per obtenir mostres biològiques dels diferents fílums de bilaterals. A partir de les mostres es va obtenir cDNA que es va fer servir per amplificar i seqüenciar 750 parells de bases la miosina per 29 taxa pertanyents a 12 grups de bilaterals més l'outgroup. Aquest va ser el primer cop que s'obtenien i analitzaven seqüències de miosina per tants fílums per tal d'obtenir una filogènia dels bilaterals. Les anàlisis conduïdes inclouen el Relative Rate Test per determinar quins organismes eren fast-clock i l'ús de diferents mètodes d'inferència filogenètica (maximum likelihood, inferència bayesiana, neighbor-joining i màxima parsimonia). Aquestes anàlisis es varen realitzar tant per les dades de miosina i 18S individualment com per la matriu concatenada d'ambdues molècules. Les possibles topologies alternatives varen ser testades per tots els sets de dades.

Si bé la representació taxonòmica d'aquest estudi és només una part de la que vam aconseguir posteriorment, va ser suficient per demostrar que:

1. El gen de la miosina és un candidat ideal per realitzar estudis filogenètics dels bilaterals.
2. El gen de la miosina confirma grups prèviament indicats pel 18S, com per exemple l'existència dels tres superclades de bilaterals o l'inclusió dels *Platyhelminthes* dins dels Lophotrochozoa.
3. Els acels i els nemertodermàtides són el grup germà de la resta de bilaterals, tant a les filogènies d'un sol gen com al concatenat d'ambdues molècules i per totes les metodologies utilitzades.

Aportació personal al treball

1. Extracció de RNA, retrotranscripció, amplificació i seqüenciació del gen la miosina per 16 representants dels lofotrocozous.
2. Assistència a l'alineació de les seqüències del gen 18S, basada en la estructura secundària, i de la miosina basada en aminoàcids.
3. Contribució en les anàlisi d'inferència filogenètica dels tres jocs de dades: 18S, miosina i 18S + miosina.
4. Assistència a la redacció de l'article.

Aportació personal al treball

1. Extracció de RNA, retrotranscripció, amplificació i seqüenciació del gen la miosina per 16 representants dels lofotrocozous.

Capítol II

Ribosomal RNA genes and Bilateria phylogeny revisited: new insights into Lophotrochozoa internal phylogeny

Jordi Paps, Jaume Baguñà & Marta Riutort*

Departament de Genetica, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Spain.

In preparation

Resum

A més del nostre interès en desenvolupar nous marcadors moleculars en el camp de la filogènia dels bilaterals, també volíem explorar la informació disponible en els dos marcadors millor mostrejats fins a la data: els gens ribosomals. Si bé ja existien estudis combinant ambdues molècules, aquests presentaven una representació dels bilaterals molt incomplerta o bé es centraven exclusivament en un dels superclades de bilaterals. Per tant varem decidir realitzar un estudi dels gens ribosomals que, per una banda, maximitzés el mostratge de la diversitat dels bilaterals, i per l'altra, apliqués un conjunt d'estratègies que minimitzessin els problemes associats amb aquests marcadors, principalment el LBA.

A partir de les bases de dades, vam obtenir 564 seqüències pel 18S i 142 pel 28S. Per cadascuna de les molècules, vam triar els representants de cada fílum amb branques més curtes (basant-nos en les distancies obtingudes per maximum likelihood). Les seqüències del 18S i del 28S dels taxa escollits es van concatenar en un alineament de 3.700 parells de bases per 104 taxa pertanyents a 28 fílums de bilaterals més l'outgroup. Aquest joc de dades es va fer servir per inferir una filogènia complerta dels bilaterals mitjançant mètodes d'inferència i models evolutius que contraresten els artefactes com el LBA. A més, aquesta filogènia es va fer servir per detectar els grups problemàtics i analitzar-los per separat, per tal de compartimentar la seva problemàtica. A més, per cada grup problemàtic es van testar les possibles topologies alternatives per tots els jocs de dades.

El conjunt d'arbres mostra que:

1. El conjunt d'estratègies emprades realment minimitza l'efecte del LBA, ja que els grups fast-clock no s'agrupen entre ells ni es posicionen prop de l'outgroup.
2. La filogènia total recupera els tres grans clades de bilaterals, així com la posició basal dels acels i nemertodermàtides
3. Les filogènies basades en els subjocs són congruents amb la filogènia total, i a més per molts grups el suport estadístic és superior a les anàlisis compartimentades.
4. Les filogènies internes de deuterostomats i ecdisozous coincideixen amb estudis anteriors, tot i que els suports són baixos.
5. Es mostra una nova filogènia dels lofotrocozous, a on destaquen:
    a. gastrotrics i gnatostomúlids com a grups basals,
    b. l'estatus polifilètic dels lofoforats,
    c. l'aparició d'un grup de fílums acelomats com a grup germà dels clades espirals

d. l'estatus parafilètic dels Spiralia degut a la posició dels braquiòpodes i foronidis (Brachiozoa) com a grup germà dels mol·luscs.

6. Es confirma que els acels i els nemertodermàtides són bilaterals basals i es suggereix l'afiliació dels quetògnats amb els ecdisozous.

Ribosomal RNA genes and Bilateria phylogeny revisited: new insights into Lophotrochozoa internal phylogeny

Jordi Paps, Jaume Baguñà & Marta Riutort*

Departament de Genetica, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Spain.

The relationships of the animal phyla are a key biological problem still pending to be solved. Morphology cannot solve the relationships among most phyla, and although molecular data have unveiled a new evolutionary scenario, they show its own limitations. Ribosomal genes (18S and 28S rDNA) have effectively been used for many years. However, they are considered of limited use to resolve deep divergences such as the origin of the bilaterians due to certain drawbacks as the long-branch attraction (LBA) problem. Here we attempt to overcome these pitfalls by combining several strategies suggested in previous studies but not applied yet to any bilaterian phylogeny based on these genes: use of Maximum Likelihood and Bayesian Inference methods, application of models with rate-heterogeneity across sites, a wide taxon sampling, and compartmentalized analyses for each problematic clade. The results obtained show that the combination of the above-mentioned strategies minimizes the LBA effect, and a well resolved Lophotrochozoa phylogeny with *Gnathostomulida* and *Gastrotricha* as earliest branching representatives emerges. Also, the Acoelomorpha (*Acoela* and *Nemertodermatida*) are confirmed with maximum support as the first branching bilaterians. Therefore it can be said that the ribosomal RNA genes can still be a reliable source for the study of deep divergences in the metazoan tree provided that a careful treatment is performed.

*Corresponding author, e-mail: mriutort@ub.edu

Introduction

Resolving the relationships among animal phyla is a key problem in modern biology, since they are instrumental to understand the evolution of many biological features including, among others, body plans, embryonic development and gene networks. Unfortunately, morphology falls short of the aim of clarifying the precise relationships among most phyla. Twenty years ago, the interest in this field was intensified by the introduction of the small ribosomal subunit RNA gene (18S rDNA or SSU) into metazoan phylogenies (Field et al. 1988; Lake 1990). However, molecular phylogenies also have their own downsides and raised new problems. The SSU's lack of resolving power on some regions of the metazoan tree (Abouheif et al. 1998; Adoutte et al. 2000; Philippe et al. 1994), and the long-branch attraction (LBA, Felsenstein 1978) are the major concerns on phylogeny resolution and credibility (Anderson and Swofford 2004).

Different sources of data have been added to the SSU sequences to overcome these drawbacks. Morphological matrices, together with molecular data, have been analyzed with Maximum Parsimony methods with varying results (Giribet et al. 2000; Peterson and Eernisse 2001; Zrzavý et al. 1998). Sequences from the large ribosomal subunit RNA gene (28S rDNA or LSU) have been analyzed together with the SSU using probabilistic methods such as Maximum Likelihood and Bayesian Inference (Mallatt and Giribet 2006; Mallatt and Sullivan 1998; Mallatt and Winchell 2002; Mallatt et al. 2004; Medina et al. 2001; Passamaneck and Halanych 2006; Telford et al. 2003; Winchell et al. 2002). Other approaches searched for new nuclear markers (Anderson et al. 2004; Peterson et al. 2004; Ruiz-Trillo et al. 2002), mitogenomics (Boore et al. 2005), high-throughput strategies such as phylogenomics (Dunn et al. 2008; Philippe and Telford 2006), or molecular synapomorphies such as the provided by the microRNA expression distribution across taxa (Sempere et al. 2006). These approaches mostly backed up the SSU division of bilaterians, but have not, however, solved all the questions raised by previous SSU studies. Moreover, the increase in information obtained in these cases is counteracted by the lower sampling of animal phyla for these new markers and methods.

The most recent and complete analysis done (Dunn et al. 2008) shows a better resolution within groups, such as the Lophotrochozoa, so far poorly resolved. Such resolution, however, comes at the price of skipping key taxa (gastrotrichs, gnathostomulids, rotifers, acoelomorphs, bryozoans, chaetognaths...) before the final analyses were performed.
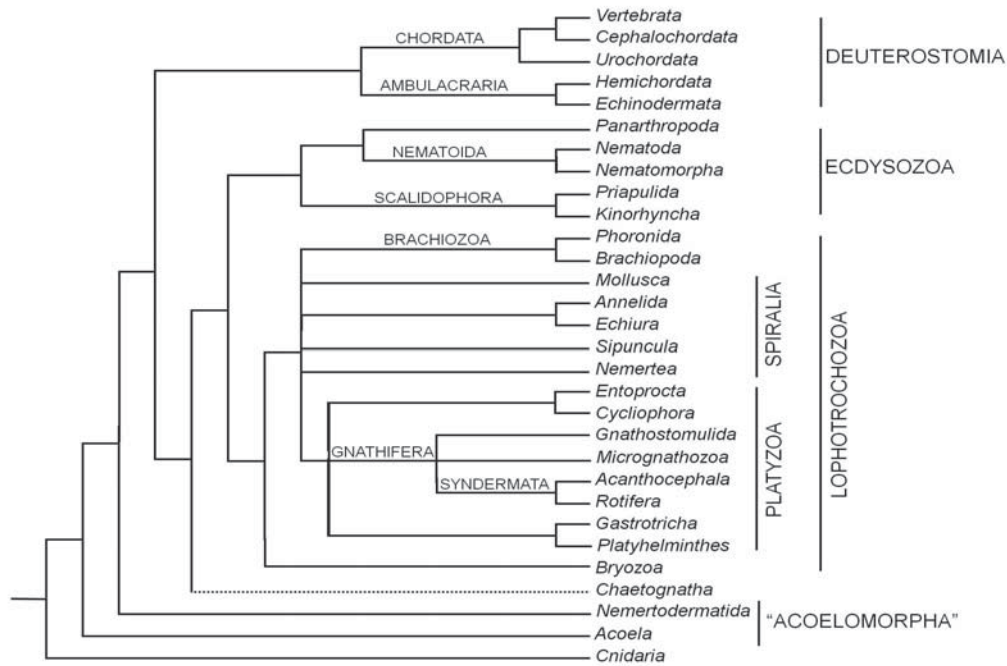
Fig. 1. Tree summarizing the bilaterian relationships based on molecular data. It shows the three main bilaterian clades, as well as some of the accepted phyla relationships and superclades. Modified after Halanych (Halanych 2004)

Despite these shortcomings, a consensus tree of the Bilateria has been portrayed by various authors (Adoutte et al. 2000; Balavoine and Adoutte 2003; Giribet 2002; Halanych 2004; Halanych and Passamaneck 2001; Telford 2006), as displayed in Figure 1 (after Halanych 2004). This tree shows the three main groups defined by molecular data, Lophotrochozoa, Ecdysozoa and Deuterostomia, as well as the most accepted relationships among the phyla they include. From this scheme it clearly stems out that Lophotrochozoa is the most problematic group, due to its high number of phyla and poor internal resolution. Furthermore, some groups with unsolved affinities might hold the key to understanding essential transitions in the bilaterian tree, namely the *Acoelomorpha*, the *Chaetognatha*,

the Gnathifera and the *Gastrotricha*. For a thorough discussion on the relationships proposed in this tree (Fig. 1) see Halanych (2004), as well as the discussion section of this paper.

The main aim of this work is to combine different approaches to unravel the status of the three big clades, mainly the Lophotrochozoa's internal relationships and the position of groups of uncertain affinities. For this endeavor we have attempted to maximize the metazoan phyla sampling and, at the same time, minimize the LBA effect, these being two of the factors that were suggested to cause uncertainties and lack of resolution in previous studies. We applied a careful analysis involving two steps. First, we applied several strategies that have been proposed to avoid

LBA in previous studies based on real and simulated data (Anderson and Swofford 2004; Bergsten 2005): using methods less sensitive to LBA, such as Maximum Likelihood (ML) or Bayesian Inference (BI); employing model modifications such as rate-heterogeneity across sites with a discrete gamma-distribution parameter; using the shortest branched representatives available for each phyla; and searching for the widest taxon sampling. Second, we have compartmentalized the analysis of the still problematic lineages (those with extremely long branches or with unstable or unreliable situations) by removing all of them from the analysis and adding them again one at a time to evaluate their respective position and the support they receive. Despite the numerous SSU and LSU sequences present in the databases, an extensive analysis of SSU and LSU for all bilaterians, applying these strategies, has yet to be performed.

In summary, the steps followed comprise the gathering of SSU and LSU sequences from a comprehensive representation of 29 phyla, the richest phyla sampling gathered for both genes yet, selecting the shortest-branched sequences available for each phylum. Next, the combined SSU and LSU dataset of 3.7 kbs, comprising 104 representatives for 29 phyla (Suppl. Data Table 1), was used to infer the bilaterian phylogeny. This phylogeny was also used to detect clades that either contain long-branched taxa or show unstable or anomalous groupings, and the sensitivity of results to the presence/absence of these groups was tested. In addition, these phylogenies were used to perform a topology comparison test.

Material and methods

Taxon Sampling, alignment and dataset assembling

1) "Preliminary" datasets: 564 SSU and 142 LSU sequences from 28 bilaterian metazoan phyla and the outgroup (7 cnidarian species) were downloaded from Genbank (for Accession numbers see Suppl. Data Table 1). The SSU alignment used in Wallberg et al. (2004) was downloaded and new sequences from genbank were added to complete the taxonomic sampling, namely the Nemertodermatida representatives, the added sequences were aligned using the Wallberg et al alignment as a profile (see that paper for a thorough discussion on the alignment methodology).

For LSU 142 sequences from former studies were aligned to secondary structure, with notation modified from Gillespie et al (2005). It has been shown that using secondary structure information facilitates the alignment in lenght-heterogeneous sequences (especially rDNAs, Gutell 1992; Gutell 1994; Gutell 1985; Hickson et al. 1996; Kjer 1995) while automated alignment programs tend to give bad results. Moreover, alignments constructed with reference to secondary structural model have also been shown to increase phylogenetic accuracy in the analysis of rDNA datasets (Cunningham et al. 2000; Dixon and Hillis 1993; Gonzalez and Labarere 2000; Hwang

and Kim 2000; Kjer 1995; Lydeard et al. 2000; Morin 2000; Morrison and Ellis 1997; Mugridge et al. 1999; Titus and Frost 1996; Uchida et al. 1998; Xia 2000; Xia 2003). Alignments were performed and checked on Bioedit (v.7.5, Hall 1999). The length-heterogeneous regions for which the nucleotides homology cannot be granted and the regions containing indels in the majority of sequences were removed prior to the analyses, using a very conservative criteria of keeping only unambiguosly conserved blocks. The final alignments contained 1,425 sites (out of 3.365 nucleotides) for SSU and 2,271 sites (out of 6,847 nucleotides)for LSU.

2) All taxa dataset (All-set): for each phyla the representatives with lowest patristic distance (irrespective of the absolute value of the distance, as calculated by ML with Treepuzzle) to the outgroup were selected in order to have the slowest evolving species for each one. Then they were merged into a combined SSU + LSU dataset with 104 representatives for 28 bilaterian phyla and the outgroup. Whenever possible, SSU and LSU sequences came from the same species (Suppl. Data Table 1). For those representatives lacking it, LSU was filled with Ns. In the *Chaetognatha*, the only LSU representative available was combined with the two chosen SSUs.

3) Subset datasets: after the first phylogenetic analyses with the All-set, five different subsamples were produced in order to examine more accurately the position of clades showing a high rate of substitutions (a taxon was considered fast evolving when its patristic distance to the outgroup was above 0.3 as calculated by ML with TreePuzzle) or presenting unstable or unreliable situations (such as the *Gastrotricha* polyphyly). With these compartimentalized analyses we also wanted to test the effect of the problematic groups on the general topology. These clades were removed from the All-set and five subsets were built adding only one of these groups at a time: Acoelomorpha (Acoel-set), Gnathifera (Gnat-set, *Gnathostomulida* lacks LSU), *Bryozoa* (Bryo-set), *Gastrotricha* (Gast-set, *Gastrotricha* lacks LSU) and *Chaetognatha* (Chaet-set).

4) Basic dataset (Basic-set): excludes the five groups containing taxa with patristic distances to the outgroup above 0.3 or presenting abnormalities. This dataset comprises 88 sequences from SSU and 87 from LSU (which lacks *Micrognathozoa* representatives) for 22 bilaterian phyla.

Phylogenetic Analyses

Modeltest (v. 3.6, Posada and Crandall 1998) was used to determine the evolutionary model best fitting each dataset. The specified model (GTR + Γ + I) was applied in all the algorithms where it was available. BI trees were inferred with a parallelized version of Mr Bayes software (v. 3.1, Ronquist and Huelsenbeck 2003), with and without partitioning of the dataset for the two ribosomal genes and with and without the

covarion model, running 3,000,000 generations in 2 independent analyses with a sample frequency of 1,000, allowing the two runs to converge onto the stationary distribution, shown by the average standard deviation of split frequencies approaching zero. To obtain the consensus tree and the BI supports, 1,000,000 generations were removed to avoid including trees sampled before likelihood values had reached a plateau. Treepuzzle (v.5.2, Schmidt et al. 2002) was employed to obtain the Gamma distribution parameters and site categories needed for fastDNAml, and was run with the options estimation accurate (slow), Tamura-Nei 93 model and a mixed rate heterogeneity model (8 Gamma categories + 1 invariable). A parallelized version of fastDNAml (v.1.2, Olsen et al. 1994) was employed with the options Global (G 3 2), Jumble and a mixed rate heterogeneity model calculated with Treepuzzle. Although they are fast algorithms, Treepuzzle and fastDNAml have not implemented the GTR model of evolution. Due to the computational cost of fastDNAml, bootstrap supports were not calculated for this method. RaxML (Stamatakis 2006), Treefinder (Jobb 2007) and Phyml (Guindon and Gascuel 2003)were used to infer ML trees and bootstrap values under the GTR+ Γ + I model, and Neighbor Joining trees were estimated using MEGA with 1,000 bootstrap replicates using the Kimura 2-Parameters model and Pairwise deletion option.

Competing topologies were evaluated for different datasets. Alternative topologies were based on previous morphological or molecular studies (indicated in the footnote) or were variations based on our analyses (see Table 1). The alternative trees were constructed using Treeview (v. 1.6.6., Page 1996), and PAUP (Swofford, 2000) was used to calculate the site likelihoods for all trees and prepare the input dataset for CONSEL. CONSEL (v.0.1i, Shimodaira and Hasegawa 2001) was run to perform the approximately unbiased test (AU, Shimodaira, 2000; RELL; 1,000 replicates) (Shimodaira 2002). The analyses were run on 4 different computers: 1) 2 PCs running Windows XP and SUSE Linux 10.0; 2) a Supercomputer located at CESCA (Centre de Supercomputació de Catalunya, http://www.cesca.es/); and 3) the Marenostrum supercomputer located at the Barcelona Supercomputing Center (http://www.bsc.es/).

Results

NJ, Treefinder and Phyml gave similar results, in many cases showing clear differences when compared to BI, RaxML and fastDNAml trees. The differences observed can most probably be explained as a consequence of LBA affecting the former group of algorithms. The trees obtained with NJ, Treefinder and PhyML are shown in the supplementary data (Supp Data Figure 1). Given the unreliability of the trees (and consequently of their bootstraps) inferred with these methods, we have relied on ML as inferred with fastDNAml and RaxML as well as BI results (from all datasets), and comparison of topologies to define which

clades are robustly recovered in our analyses, as described below.

All-set and subset analyses

BI, RaxML and fastDNAml phylogenies agree for all the datasets analyzed, and all the methods recover a steady topology within each dataset and across all of them, with only minor differences in some fastDNAml trees (not shown due to space limitations). No differences were seen in the topologies recovered by BI when covarion was used or when BI estimates were unlinked for the SSU and LSU partitions of the matrix. Figure 2 shows the BI and RaxML topology obtained from the All-set. Groups containing long-branched sequences (distance values to the outgroup above 0.3 in the ML patristic matrix) or showing anomalies (such as the Gastrotricha polyphyly) are boxed. These groups were selected for the following compartmentalized analyses, removing all but one each time.

The BI and RaxML trees for subsets Acoelomorpha (Acoel-set), Gnathifera (Gnat-set), Bryozoa (Bryo-set), Gastrotricha (Gast-set), and Chaetognatha (Chaet-set) can be seen in Figure 3. Finally, another dataset excluding all the problematic taxa (Basic-set, Fig. 3F) was used to test the effect on the support values when all the problematic groups were excluded. The overall topology of the tree agrees between the All-set (Fig. 2), the subsets and Basic-set (Fig. 3). However, the nodal support increases in the subsets and even more so in the Basic-set (except the Deuterostomia), as shown in Table 1. The fact that the supports

do not decrease when long branches are removed clearly indicates that high supports in the All-set are not a consequence of LBA misleading the method. The position of the long-branched taxa and problematic groups in the All-set tree is consistent with their position in the subset phylogenies (compare Fig.2 with Fig 3), though again the subsets show noticeably higher supports (Table 1).

On the whole, all the datasets recover the clades Lophotrochozoa, Ecdysozoa and Deuterostomia as well as the protostomates. All-set (Fig. 2) and Acoel-set (Fig 3A) show Acoela as the first branching bilaterians and the Nemertodermatida as sistergroup to the rest of the bilaterians. Regarding Ecdysozoa, Scalidophora (Priapulida+Kinorhyncha) and Panarthropoda (Arthropoda+Tardigrada) are well supported, but not Nematoida (Nematoda+Nematomorpha), the latter only showing low support in some subsets (Acoel-set, Gna-set, Bryo-set) and Basic-set (Fig. 3). Scalidophora and Panarthropoda are sistergroup in some subsets (Acoel-set, Gna-set and Basic-set, Fig. 3) while nematodes are Panarthropoda sistergroup in the others (Gast-set, Bryo-set and Chae-set, Fig 3) as well as in the All-set (Fig. 2).

Although the *Chaetognatha* representatives show very long branches, they fall within the ecdysozoans with maximum support (Fig. 3E), clustering with the Scalidophora in the All-set (Fig. 2) and the Chaet-set (Fig. 3E). As regards Deuterostomia, chordate monophyly is not recovered in the
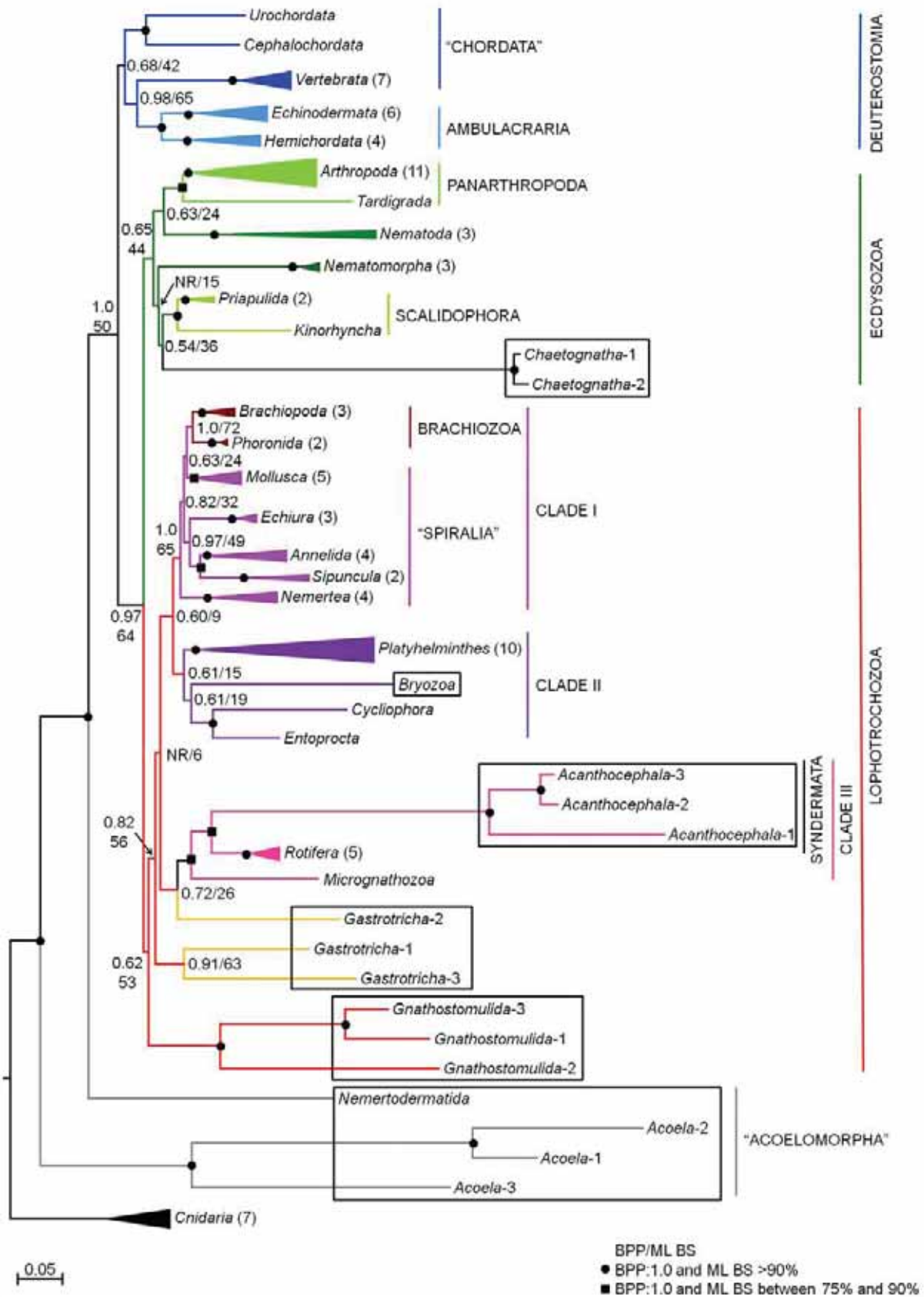
Fig. 2. Bayesian and RaxML topology (GTR + Γ + I) for All-set (104 metazoan representatives, Cnidaria as the outgroup. Posterior probabilities (PP) and ML bootstrap values (BV) are indictated with a bullet (PP=1.0 & BV>90%) or a square (PP=1.0 & 75%<BV≤90%) on the node; lower values are indicated. See table 1 to compare the effects of removing problematic taxa on boostrap values. NR stands for a node not resolved in the BI consensus tree. Problematic taxa are boxed (see text). Monophyletic phyla are collapsed (triangle size proportional to number of representatives included), the monophyly of each phylum has maximum support (except for gastrotrichs). The scale bar indicates the number of changes per site in ML inference. For species names corresponding to each terminal see Supp. Data table 1.
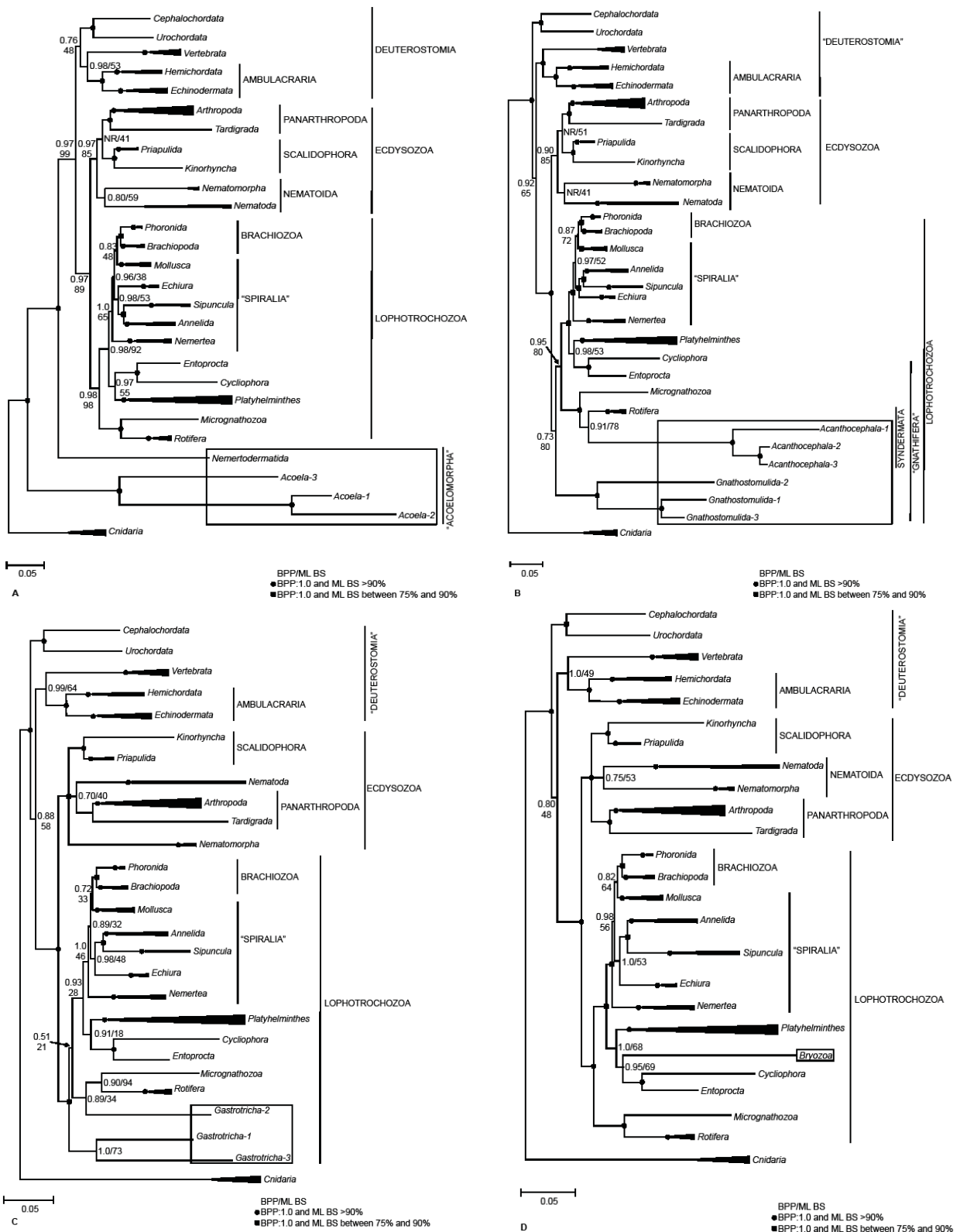
Figure 3. Bayesian and RaxML topologies (GTR + Γ + I) for the taxa subsets. Posterior probabilities (PP) and ML bootstrap values (BV) are indictated with a bullet (PP=1.0 & BV>90%) or a square (PP=1.0 & 75%<BV≤90%) on the node; lower values are indicated. For a comparison of the bootstrap values in the different trees see table 1. NR stands for a node not resolved in the BI consensus tree. Phyla of interest for each independent analysis are boxed. Monophyletic phyla are collapsed (triangle size proportional to number of representatives included), the monophyly of each phylum has maximum support (except for gastrotrichs). The scale bar indicates the number of changes per site. A: tree from Acoelomorpha dataset (Acoel-set); B: tree from Gnathifera dataset (Gna-set); C: tree from Gastrotricha dataset (Gast-set); D: tree from Bryozoa dataset (Bryo-set); E: tree from Chaetognatha dataset (Chae-set); and F: tree from the Basic-set. For species names corresponding to each terminal see Supp. Data table 1.
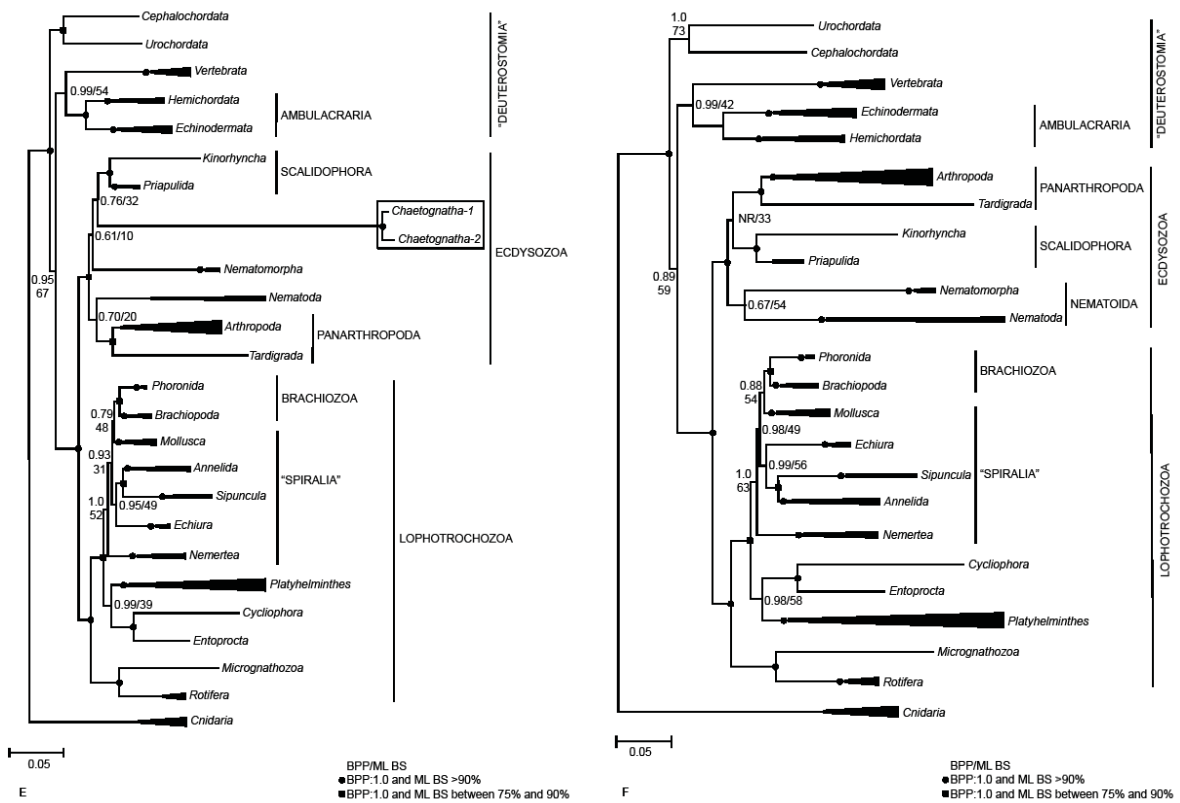
Figure 3. (Continued)

All-set due to urochordate and cephalochordate sequences appearing as the first splitting deuterostomates (Fig 2). The Acoel-set tree (Fig. 3A) also places *Urochordata + Cephalochordata* as sistergroup to the other deuterostomates, whereas the other subsets and the Basic-set place them as first-splitting bilaterians. When the urochordate sequence is removed from the Basic-set (not shown) the cephalochordate appears as sistergroup to the *Vertebrata*, with a 92% Bayesian posterior probability (BPP). Outside of cephalochordates and urochordates, the other deuterostomates cluster with maximum support, showing a well-resolved Ambulacraria as sistergroup to *Vertebrata*.

The *Gnathostomulida* are shown as sistergroup to the remaining Lophotrochozoa in the All-set (Fig. 2) and Gna-set trees (Fig. 3), although their basal position among lophotrochozoans shows low support; Lophotrochozoa monophyly support decreases when *Gnathostomulida* are present, while it is maximum in all the other sets. Regarding the other lophotrochozoans, the tree (Fig. 2) shows a polyphyletic *Gastrotricha* (Gastrotricha 1 and 3) as the next branch, followed by a clade containing *Micrognathozoa* + (*Rotifera + Acantocephala*) plus Gastrotricha-2 as sister group to the rest of lophotrochozoans. The *Gastrotricha* appears polyphyletic in both All-set and Gast-set (Fig. 3C), but their monophyly is shown as the best alternative by the AU test,

Table 1. Summary of the status of the main clades in different datasets. Bayesian posterior probabilities and RaxML bootstrap supports are shown.

Clade	All taxa Dataset (Fig. 2)	Sub-sets (Fig. 3)	Basic Dataset (Fig. 3F)
Acoela split from rest of bilaterians	1.0-99	1.0-99	-
PROTOSTOMIA	0.97-64		1.0-100
LOPHOTROCHOZOA	0.62-53		1.0-98
Gnathostomulida as Lophotrochozoa	0.62-53	0.73-53	-
Gastrotricha inside Lophotrochozoa	0.82-56	1.0-81	-
Micrognathozoa + Rotifera	-		1.0-99
Rotifera + **Acantocephala**	0.98-79	0.91-78	-
Clade III: (*Rotifera +* **Acantocephala**) + *Micrognathozoa*	0.83-77	1.0-96	-
Cycliophora + Entoprocta	1.0-92		1.0-98
(*Cycliophora + Entoprocta*) + *Platyhelminthes*	-		0.99-58
(*Cycliophora + Entoprocta*) + **Bryozoa**	0.61-19	0.95-69	-
Clade II: ((*Cycliophora + Entoprocta*) + **Bryozoa**) + *Platyhelminthes*	0.61-15	1.0-68	-
Clade I (Trochozoa including *Phoronida + Brachiopoda*)	1.0-65		1.0-63
((*Annellida + Sipuncula*) + *Echiura*) + ((*Phoronida+ Brachipoda*) + *Mollusca*)	0.82-32		0.99-49
Annelida + *Sipuncula*	1.0-77		1.0-76
(*Annelida + Sipuncula*) + *Echiura*	0.97-49		1.0-56
(*Phoronida+ Brachiopoda*)	1.0-72		1.0-83
(*Phoronida + Brachiopoda*) + *Mollusca*	0.63-24		0.84-54
ECDYSOZOA	0.65-44		1.0-97
Panarthropoda (*Arthropoda + Tardigrada*)	1.0-85		1.0-99
Nematoida (*Nematoda + Nematomorpha*)	NR-NR		0.65-54
Scalidophora (*Priapulida + Kinorhyncha*)	1.0-91		1.0-100
Chaetognatha as Ecdysozoa	0.65-44	1.0-90	-
DEUTEROSTOMIA	0.68-42		0.59*-25
Ambulacraria (*Hemichordata + Echinodermata*)	1.0-100		1.0-100
Chordata	NR-NR		0.92*-38

- Not applicable; NR Not recovered; * group only recovered when *Urochordata* are removed
Names in bold indicate the problematic phyla added in the independent datasets

while the same test rejects the topology obtained by BI, RaxML and fastDNAml (see discussion). The remaining lophotrochozoans are found in three main clades. The first (Clade I in Fig. 2) features *Nemertea* as a sistergroup of 2 subclades, one formed by *Echiura +*

(*Sipuncula + Annelida*) and the other including *Mollusca* + (*Brachiopoda + Phoronida*), and it is consistently recovered in all trees. Another clade (Clade II in Fig. 2) shows *Bryozoa* + (*Entoprocta + Cycliophora*) as sistergroup to *Platyhelminthes* (i.e. Catenulida + Rhabditophora; Baguñà and Riutort 2004), all of which make up a sistergroup to Clade I. Finally, despite their very long branches, the *Acantocephala* group with *Rotifera*, a cluster supported by both All-set and Gnat-set, with *Micrognathozoa* as their sistergroup (Clade III in Fig 2).

Comparison of Topologies

For each dataset, the best tree was statistically compared against alternative trees (Table 2). Competing topologies were selected either because they appeared in previous studies (see table footnotes) or because they were minor alterations of the trees obtained in this study. Concerning the subsets, all the alternative topologies tested were significantly worse than the original tree for all the sets with two exceptions: 1) the test based on the Gast-set (hypotheses 9 to 11) rejects the original polyphyletic *Gastrotricha* in favor of their monophyly, despite the fact that the former is found in BI and ML trees; 2) the hypothesis placing chaetognaths as sistergroup to ecdysozoans (hypotheses 13) cannot be rejected. The hypotheses rejected are *Acoelomorpha* as sistergroup to Ambulacraria (hypothesis 2) or to *Platyhelminthes* (hypothesis 3), a monophyletic *Gastrotricha* as sistergroup to Ecdysozoa (hypothesis 11), the

monophyly of Gnathifera (hypothesis 5), bryozoans as sistergroup of brachiopods and phoronids within Clade I (hypothesis 7), the monophyly of the Lophophorata (hypothesis 8), the polyphyly of Gastrotricha (hypothesis 9), and the *Chaetognatha* either as sistergroup to Lophotrochozoa (hypothesis 14) or to Protostomia (hypothesis 15).

The All-set allowed studying the same alternative hypotheses tested in the previous datasets as well as new ones. As regards the branching pattern among the ecdysozoan clades, many of the multiple alternatives tried (hypotheses 17, 18, 20, 21, 23, 24 and 25) were not rejected. Regarding the rest of bilaterians, other hypothesis not rejected by the All-set were those showing *Gastrotricha* monophyletic splitting after *Gnathostomulida* and sistergroup to the rest of Lophotrochozoa (hypothesis 27), and Deuterostomia monophyly (hypothesis 33). The topologies rejected are Scalidophora + Panarthropoda and paraphyletic Nematoida (hypothesis 19), *Chaetognatha* as sistergroup to an ecdysozoan clade (where scalidophorans and nematoidans are sistergroup to panarthropodans, hypothesis 22) or Protostomia (hypothesis 26), *Gastrotricha* monophyletic sistergroup to Ecdysozoa (hypothesis 28), *Bryozoa* sistergroup to *Phoronida* + *Brachiopoda* (hypothesis 29), Lophophorata monophyly (hypothesis 30), *Acoelomorpha* sistergroup to

Table 2. Topology tests results.

	Subsets	Ln Likelihood	AU
1.	***Acoelomorpha* dataset original tree (basal bilaterians, Fig 3A)**	-64053,3559	best
2.	*Acoelomorpha* sistergroup to Ambulacraria (Philippe et al. 2007)	-64104,8488	0,001*
3.	*Acoelomorpha* sistergroup to Platyhelminthes (Rieger 1991)	-64136,2360	0,001*
4.	**Gnathifera dataset original tree (paraphyletic, Fig. 3B)**	-62365,0881	best
5.	Gnathifera monophyletic (Haszprunar and 41–48. 1996)	-62405,8423	0,003*
6.	***Bryozoa* dataset original tree (sistergroup to *Entoprocta* + *Cycliophora*, Fig. 3C)**	-59151,7621	best
7.	*Bryozoa* moved to be sistergroup to (*Phoronida + Brachiopoda*), sistergroup to *Mollusca* [#]	-59175,2591	0,008*
8.	Lophophorata = *Bryozoa + Entoprocta* + (*Phoronida + Brachiopoda*), sistergroup to *Mollusca* (Hyman 1959)	-59203,2772	0,002*
9.	***Gastrotricha* dataset original tree (polyphyletic, Fig. 3D)**	-59152,5900	0,006*
10.	*Gastrotricha* monophyletic and sistergroup to Lophotrochocozoa[#]	-59115,3544	best
11.	*Gastrotricha* monophyletic and sistergroup to Ecdysozoa (Schmidt-Rhaesa 2003)	-59131,5161	0,003*
12.	***Chaetognatha* dataset original tree (sistergroup to Scalidophora, Fig. 3E)**	-59610,9699	best
13.	*Chaetognatha* sistergroup to Ecdysozoa (Zrzavý et al. 1998)	-59617,9625	0,058
14.	*Chaetognatha* sistergroup to Lophotrochozoa (Matus et al. 2006)	-59624,3330	0,045*
15.	*Chaetognatha* sistergroup to Protostomia (Marletaz et al. 2006)	-59628,0121	0,000*
	All taxa dataset	Ln Likelihood	AU
16.	**Best tree (Fig. 2)**	-72590,3401	best
17.	Nematoida monophyletic, sistergroup to Panarthropoda (Mallatt and Giribet 2006)	-72590,6644	0,658
18.	Nematoida + (Scalidophora + Panarthropoda) (Glenner et al. 2005b) (Glenner et al. 2004)	-72595,0884	0,342
19.	(Scalidophora + Panarthropoda) and paraphyletic Nematoida [#]	-72601,3860	0,032*
20.	((Scalidophora+*Chaetognatha*)+Nematoida) + Panarthropoda[#]	-72592,0407	0,474
21.	((Scalidophora + *Chaetognatha*) + Panarthropoda) and paraphyletic Nematoida [#]	-72598,6865	0,055
22.	*Chaetognatha* + ((Scalidophora+Nematoida) + Panarthropoda) [#]	-72599,7541	0,017*
23.	Nematoida + ((Scalidophora+*Chaetognatha*) + Panarthropoda) [#]	-72592,3215	0,532
24.	*Chaetognatha* sistergroup to Ecdysozoa (Zrzavý et al. 1998)	-72594,8507	0,231
25.	*Chaetognatha* sistergroup to Lophotrochozoa (Matus et al. 2006)	-72598,2265	0,224
26.	*Chaetognatha* sistergroup to Protostomia (Marletaz et al. 2006)	-72602,4018	0,008*
27.	*Gastrotricha* monophyletic, splitting after *Gnathostomulida* in the Lophotrochozoa[#]	-72603,7584	0,078
28.	*Gastrotricha* monophyletic, sistergroup to Ecdysozoa (Schmidt-Rhaesa 2003)	-72619,5605	0,003*
29.	*Bryozoa* sistergroup to (*Phoronida + Brachiopoda*), together sistergroup to *Mollusca*[#]	-72614,8445	0,012*
30.	Lophophorata = *Bryozoa + Entoprocta* + (*Phoronida + Brachiopoda*), sistergroup to *Mollusca* (Hyman 1959)	-72662,9733	0,000*
31.	*Acoelomorpha* sistergroup to Platyhelminthes (Rieger 1991)	-72700,9191	0,000*
32.	*Acoelomorpha* sistergroup to Ambulacraria (Philippe et al. 2007)	-72693.3459	0,000*
33.	Deuteromia monophyletic (Cavalier-Smith 1998)	-72610,9121	0,107
34.	Gnathifera monophyletic, sistergroup to the rest of Lophotrochozoa[#]	-74836,0737	0,000*
35.	Platyzoa (without *Acoela*) sistergroup to the rest of Lophotrochozoa (Giribet et al. 2000)	-74363,6413	0,000*
36.	Platyzoa (including *Acoela*) sistergroup to the rest of Lophotrochozoa (Giribet et al. 2000)	-75801,0545	0,000*

AU, approximately unbiased test p-values; in bold, the original tree obtained by BI against with alternative hypotheses are tested; * hypothesis rejected when p < 0.05 for AU test; Numbers in parentheses refer to the studies in the bibliography on which this hypothesis is based; [#] hypotheses partially based on bibliography but modified to accomodate the topology obtained from our dataset.

Platyhelminthes or Ambulacraria (hypotheses 31 and 32), Gnathifera monophyly (hypothesis 34), and the Platyzoa (with and without acoels, hypotheses 35 and 36).

Discussion

The analyses here shown represent the widest taxonomic sampling of SSU and LSU sequences reported and analyzed to date using probabilistic methods. Overall, our results confirm that a combination of a wide taxon sampling, the use of short-branched representatives and careful analyses turns ribosomal sequences into molecules still able to furnish new answers.

Methodological Approach

The use of probabilistic methods and an evolutionary model with a gamma-distribution has proved to be a reliable way to avoid the LBA effect. Also, the selection of "short-branched" representatives has contributed, producing a consistent phylogeny across all the datasets and obtaining BI, RaxML and fastDNAml trees that agree in their topology, with only minor discrepancies in some fastDNAml analyses. Consistent differences were found between these three methods when compared to NJ, Treefinder and PhyML. The main difference is the strong tendency in NJ and PhyML to group long branches together, either within the ingroup or close to the outgroup (see examples in

Supplementary Data Figure 1), in some cases also affecting the Basic-set topology. While we expected of NJ, a distance-based method, to be proner than the ML or BI methods to LBA effects, to our surprise a similar result was found for many PhyML and Treefinder topologies. Both methods start the heuristic search with a NJ inferred tree that could result in the search being trapped in local minima close to the topology obtained with the NJ algorithm. Altogether, these results should warn on using algorithms starting from a NJ tree with complex datasets.

On the other hand, the other ML algorithms and BI place long branches deep inside the ingroup, as clearly shown in the All-set (Fig. 2) and the subset analyses (Fig. 3). If the LBA effect were active, the long branches would appear near the outgroup or close to one another. In our view, this suggests that LBA mostly does not affect our results obtained with BI and ML. As described, while the addition of long-branched phyla, either independently or simultaneously, does not have a drastic effect on the bilaterian topology, it does change the support of all nodes on the tree. In general, the subset trees show higher supports than the All-set, and the Basic-set shows even higher supports (Table 1). Moreover, in the topology comparison test (Table 2), while the alternative topologies are mostly rejected in comparisons of independent subset analyses, some of the same topologies are not rejected for the All-set. This could be

just an effect from the lesser taxon sampling and/or could stem from the increase of homoplasy due to the simultaneous presence of fast-evolving sequences. Therefore homoplasy, while not misleading the inference method, would reduce the number of sites supporting a node and would turn the differences among alternative topologies non-significant.

Bilaterian Phylogeny

BI and ML results from all datasets, together with the comparison of topologies, were used to define which clades are robustly recovered in our analyses and are summarized in the tree depicted in Figure 4. For the first time in such a comprehensive SSU+LSU analysis, the monophyly of Protostomia is recovered and remarkably Lophotrochozoa and Ecdysozoa also appear with high support (see Subsets in Fig. 3 and Table 1). Compared to Figure 1, the most noticeable difference is the increase in resolution obtained within the Lophotrochozoa. In contrast, our datasets did not support a monophyletic deuterostomate clade due to the anomalous *Urochordata+Cephalochordata* position. It is also worth noting the fact that we recover the monophyly of all phyla with high support (which was not always the case in previous studies).

The *Acoelomorpha*, formerly belonging to *Platyhelminthes*, appear as a paraphyletic group at the base of the remaining bilaterians, with the *Acoela* branching first. Their affinities to platyhelminths (Rieger 1991) or to deuterostomates (Philippe et al. 2007) were rejected by topology tests, altogether confirming earlier results. Regarding deuterostomate and chordate monophyly, only the Acoel-set recovers the monophyly of Deuterostomia (Fig. 3A), as does the All-set (Fig. 2), but not the chordate monophyly. The only urochordate representative, *Ciona intestinalis* has been considered problematic in many former studies (Glenner et al. 2005a; Mallatt and Giribet 2006; Mallatt and Winchell 2002; Philippe and Telford 2006; Telford et al. 2003; Winchell et al. 2002) and when it is removed from the Basic-set, the cephalochordate clusters with the vertebrates with high support. This is likely due to an LBA effect produced by the urochordate sequence, cancelled by the shortening of the long branch that separates the outgroup from the ingroup when acoelomorphs are added. Urochordates aside, deuterostomate internal relationships are resolved into two clades, *Vertebrata* and Ambulacraria (*Hemichordata* + *Echinodermata*), in agreement with most morphological (Gutmann 1981; Jollie 1973) and molecular analyses (Dunn et al. 2008; Giribet et al. 2000; Peterson and Eernisse 2001; Telford et al. 2003; Winchell et al. 2002; Zrzavý et al. 1998).

The internal relationships of Ecdysozoa resolve into three main clades: 1) Scalidophora (*Priapulida* + *Kinorhyncha*); 2) Nematoida (*Nematoda* + *Nematomorpha*); and 3) Panarthropoda (*Tardigrada* + *Arthropoda*).
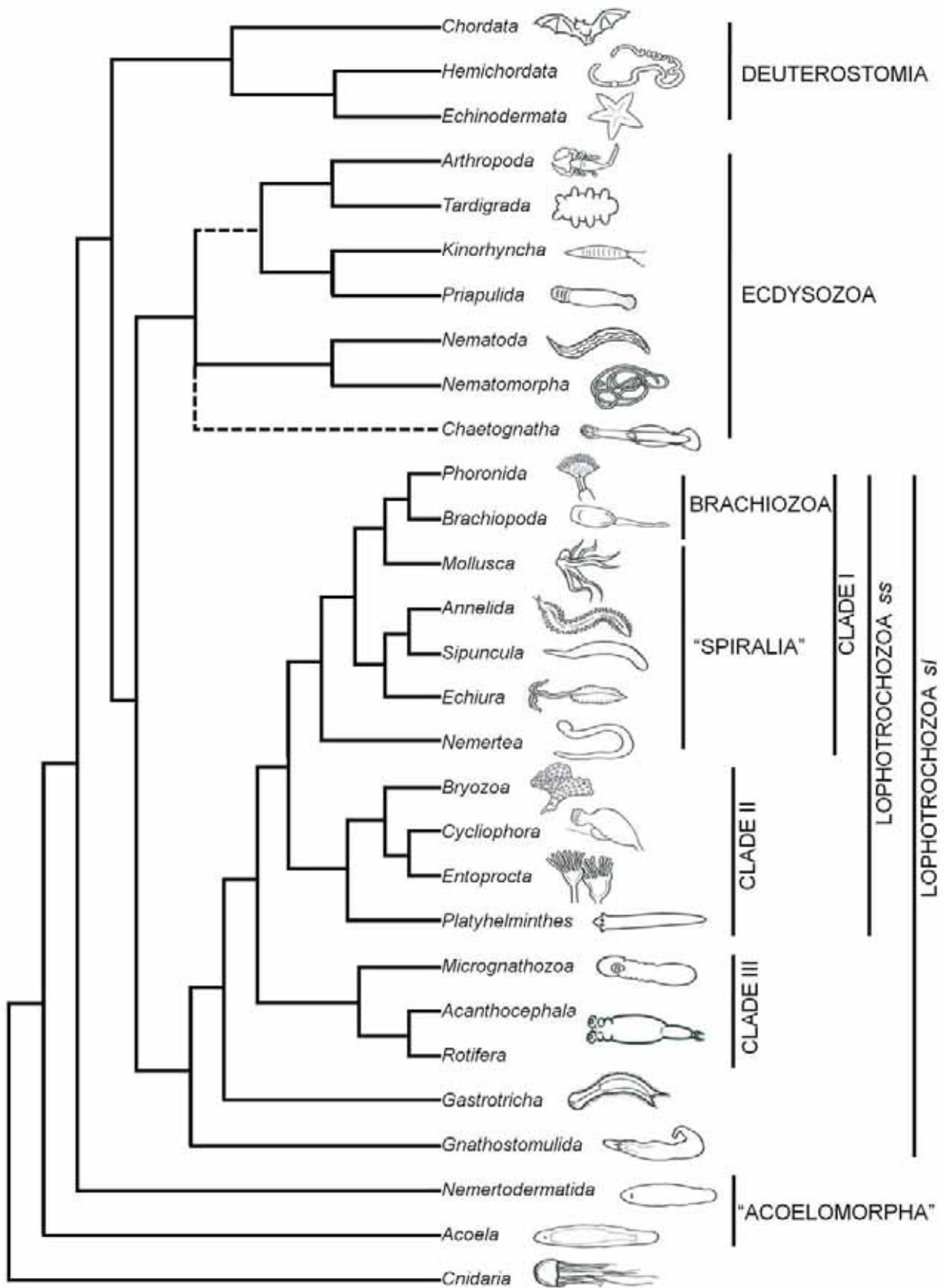
Figure 4. Summary tree from our results. See text for discussion.

These clades agree with morphology (Ehlers et al. 1996; Lemburg 1995; Schmidt-Rhaesa 1996; Schmidt-Rhaesa et al. 1998) and other molecular studies (Aguinaldo et al. 1997; Aleshin et al. 1998; Dunn et al. 2008; Giribet et al. 2000; Mallatt et al. 2004; Peterson and Eernisse 2001; Zrzavý et al. 1998). While Scalidophora and Panarthropoda show maximum supports, Nematoida is poorly supported only in some subsets; the relationship Scalidophora + Panarthropoda is found in some subsets, albeit with low support, and other scenarios could not be rejected by our data.

Chaetognatha shows one of the longest branches of the tree. Even so, they cluster with high support within ecdysozoans in the Chae-set, as sistergroup to Scalidophora, albeit the topologies tests do not reject their position as basal ecdysozoans (hypotheses 18 and 30) but reject their sistergroup relationship to protostomates or lophotrochozoans. This is in disagreement with other studies based on SSU and LSU (Mallatt and Winchell 2002), mtDNA (Helfenbein et al. 2004; Papillon et al. 2004) and multigenic approaches (Marletaz et al. 2006; Matus et al. 2006; Philippe et al. 2007), which supported a basal position among the protostomes.

Lophotrochozoa sensu stricto

The lophotrochozoans are of special interest because they include the greatest body plan diversity of the three main bilaterian superclades. In its original node-based definition(Halanych et al. 1995), the group included only annelids, molluscs and lophophorates. Subsequently, this definition has been broadened to include nearly all the phyla not belonging to either Deuterostomia or Ecdysozoa. Since most of them do not fit the original definition of having either trochophora larvae (Trochozoa) or a lophophore (Lophophorata), we consider this extended lophotrochozoan assemblage as the "Lophotrochozoa *sensu lato*" (Fig 4). If the original node-based definition is here applied (Halanych et al. 1995), the Lophotrochozoa *sensu stricto* is formed by two groups that we have named Clade I and Clade II.

Clade I receives high support in the analyses of all the datasets. It is an assemblage constituted by some of the phyla with spiral cleavage (nemertines, annelids, molluscs, echiurans and sipunculans) and two lophophorate phyla with radial cleavage (brachiopods and phoronids). Although affinities among these spiralian groups were already hinted in previous studies (Giribet et al. 2000; Passamaneck and Halanych 2006; Peterson and Eernisse 2001; Winchell et al. 2002; Zrzavý et al. 1998), the internal phylogeny shown here has never been recovered in any of them. Only in the recent EST study based on 150 genes (Dunn et al. 2008) appears the same group with very similar internal relationships, varying only the positions of nemertines and molluscs that are interchanged. The most basal group in Clade I is *Nemertea*. Apparent lack of coelom grouped them in the past with platihelminths (Hyman 1951; Nielsen 1995). Nowadays, nemertines

are known to bear a coelomic cavity (Turbeville et al. 1992), and the hox signatures of lophotrochozoans (Balavoine et al. 2002; de Rosa et al. 1999). Branching next to nemertines we found a highly supported (*Echiura* + (*Annelida* + *Sipunculida*)) group. A close relationship of *Echiura* to *Annelida* has been proposed both on morphological grounds (Hessling 2002; Nielsen 1995) and molecular data (Giribet et al. 2000; Mallatt and Winchell 2002; McHugh 1997; Peterson and Eernisse 2001). In turn, sipunculans have developmental affinities to both annelids (Clark 1969; Rice 1985) and molluscs (Scheltema 1993), though recent mtDNA and multigenic studies situate them closer to annelids (Boore and Staton 2002; Struck et al. 2007). In earlier SSU and combined analyses, sipunculans never show a highly supported relationship with any other phyla (Giribet et al. 2000; Glenner et al. 2005a; Mallatt and Winchell 2002; Passamaneck and Halanych 2006; Peterson and Eernisse 2001; Zrzavý et al. 1998), and a close relationship with annelids as presented here has never been proposed in previous SSU and LSU studies (Mallatt and Winchell 2002; Passamaneck and Halanych 2006). The sistergroup to the *Annelida* assemblage is a clade made up by *Mollusca* and *Phoronida* + *Brachiopoda*. The brachiopod-phoronid affinity has already been shown on the basis of SSU data (Cohen 2000; Cohen et al. 1998; Peterson and Eernisse 2001). Moreover, both SSU (Aguinaldo et al. 1997; Giribet et al. 2000; Halanych et al. 1995; Peterson and Eernisse 2001) and Hox genes (de Rosa et al. 1999) had already related them

to spiralians. Finally, mitochondrial gene data and a former SSU + LSU analysis (Helfenbein and Boore 2004; Stechmann 1999) pointed out a close relationship of brachiopods and phoronids to molluscs (Mallatt and Winchell 2002), a placement corroborated in our trees.

Clade II is made up by three acoelomate phyla (*Platyhelminthes*, *Entoprocta* and *Cycliophora*), together with the coelomate *Bryozoa.* Former SSU studies have already shown that bryozoans are not closely related to lophophorates (Cohen 2000; Littlewood et al. 1998) but their position was left open (Giribet et al. 2000; Passamaneck and Halanych 2006; Peterson and Eernisse 2001; Zrzavý et al. 1998). In our analyses bryozoans cluster with *Entoprocta* + *Cycliophora* with maximum support, and their relationships to *Brachiopoda* + *Phoronida* or the Lophophorata are rejected; this is in agreement with Nielsen's (Nielsen 1995) results on the basis of ontogenetic and metamorphosis features. *Cycliophora* has been related to entoprocts in morphological analyses (Funch and Kristensen 1995; Sørensen 2000; Zrzavý et al. 1998) and in the most recent SSU + LSU study (Passamaneck and Halanych 2006), while SSU data related them to rotiferans (Giribet et al. 2000; Peterson and Eernisse 2001; Winnepenninckx et al. 1998) and gnathostomulids (Giribet et al. 2004). Our trees strongly support a *Cycliophora-Entoprocta* relationship, both related with bryozoans, whereas no affinities to syndermatans or gnathostomulids are shown. This is the first molecular evidence backing up

the original morphological hypothesis of *Bryozoa* and *Cycliophora* related to entoprocts.

Recent molecular phylogenies have shown *Platyhelminthes* (*Catenulida* + *Rhabditophora*) as basal lophotrochozoans (Peterson and Eernisse 2001; Ruiz-Trillo et al. 1999) or within the Platyzoa (Giribet et al. 2000; Passamaneck and Halanych 2006). In our tree, platyhelminths appear in an unprecedented new position as sistergroup to the *Bryozoa* + (*Cycliophora*+*Entoprocta*) clade; support for this position is high in the Basic-set and Bryo-set (table 2). Albeit platyhelminths, cyclioforans and entoprocts share a negative trait, the acoelomate condition, no evident morphological synapomorphies could be thought of to group these phyla.

Lophotrochozoa sensu lato

Gnathostomulida are the first branching lophotrochozoans, a sistergroup to a highly supported clade made by the rest of the Lophotrochozoa s.*l.* (Fig. 3B). Although appearing in BI and ML trees, this sistergroup relationship has never had significant support, likely due to the absence of LSU sequences for this phylum. This position contrasts with previous molecular studies (Littlewood et al. 1998; Zrzavý et al. 1998) which related the gnathostomulids to the Ecdysozoa or suggested it formed a third protostome group with chaetognaths (Glenner et al. 2005a). Recent morphological studies place the *Gnathostomulida* close to rotifers and acantocefals forming the Gnathifera (Ahlrichs

1997; Rieger and Tyler 1995). Our trees do not recover a monophyletic Gnathifera and comparison of topologies rejects this hypothesis (Gnat-set and All-set, Table 2).

Regarding the *Gastrotricha*, they appear to be polyphyletic, but this is contradicted by the comparison of topologies where their monophyly is either shown as the best tree (Gast-set) or cannot be rejected (All-set). Gastrotrich SSU sequences have presented conflicting results in previous studies, showing them as polyphyletic within lophotrochozoans (Giribet et al. 2004; Manylov 2004) or monophyletic in the most recent study (with low support, Todaro et al. 2006). Their problematic nature, together with lack of gastrotrich LSU in our dataset, may explain our failure to recover its monophyly.

Despite their polyphyly, their clustering to the rest of lophotrochozoans (except gnathostomulids) has maximum support (Fig. 3C and Table 1), and any relationship to Ecdysozoa is rejected by the comparison of topologies. Moreover, a close relationship between gnathostomulids and gastrotrichs has been suggested on morphological grounds (Rieger 1976; Sterrer et al. 1985), while molecular studies have related gastrotrichs with gnathostomulids within Ecdysozoa (Neotrichozoa, Zrzavý et al. 1998), in the Platyzoa clade as sistergroup to platyhelminths (Giribet et al. 2000) or closely related to *Rotifera* and *Cycliophora* within lophotrochozoans (Todaro et al. 2006). Unfortunately neither gnathostomulids nor

gastrotrichs were included in recent SSU + LSU analyses due to the lack of LSU (Mallatt and Winchell 2002; Passamaneck and Halanych 2006). Our trees support their affinities with gnathostomulids (splitting after them) and syndermatans (splitting before them), placing them robustly within the Lophotrochozoa.

Clade III includes *Rotifera, Acantocephala and Micrognathozoa*. The relationships among *Rotifera* and *Acantocephala* have been suggested by morphology (see examples in Schmidt-Rhaesa 2003) and SSU (Syndermata, Garey et al. 1996; Garey and Schmidt-Rhaesa 1998; Zrzavý et al. 1998). In our analyses this grouping is solidly recovered despite the very long branches of acantocephalans. As regards *Micrognathozoa*, the presence of several homologous jaw elements related them to syndermatans (Kristensen and Funch 2000; Sørensen 2003), although recent molecular data are more ambiguous, placing them either near syndermatans and *Cycliophora* or near entoprocts, depending on the genes analyzed (Giribet et al. 2004). Our analyses clearly recover the clade (*Micrognathozoa* + (*Rotifera* + *Acantocephala*)) with maximum support in the Gnat-set (Fig. 3B) while placing cycliophorans together with entoprocts within Lophotrochozoa *ss*. Overall, our phylogeny of the Protostomia conflicts with proposals like Gnathifera (*Gnathostomulida* + *Micrognathozoa* (*Rotifera* + *Acantocephala*); Ahlrichs 1997; Nielsen 2001; Sørensen 2000), Cycloneuralia (sensu lato, Gastrotricha +

Nematoida + Scalidophora; Nielsen 2001; Peterson and Eernisse 2001; Schmidt-Rhaesa et al. 1998; Sørensen 2000; Zrzavý 2003), Neotrichozoa (Gastrotricha + Gnathostomulida; Zrzavý et al. 1998) and Platyzoa (Cavalier-Smith 1998; Garey 2001; Passamaneck and Halanych 2006).

In our trees we do not recover a monophyletic Platyzoa and the comparison of topologies from the All-set completely rejects their monophyly; this is in agreement with the dismissal of this clade in recent analyses (Dunn et al. 2008), where is claimed that Platyzoa might be an artefact of attracting unstable long-branch species to *Platyhelminthes* vicinity. If acoels are excluded from the platyzoan definition, the platyzoan representatives are spread mainly as a paraphyletic assemblage at the base of the Lophotrochozoa, a branching pattern of paraphyletic Neotrichozoa and paraphyletic Gnathifera made up by *Gnathostomulida* (*Gastrotricha* + (*Micrognathozoa* (*Rotifera* + *Acantocephala*))+ Lophotrochozoa ss). Hence, the characters that have been proposed as synapomorphies for these two groups may be reconsidered as plesiomorphic states for the Lophotrochozoa *sl*.

Evolutionary implications

The results here set forth (Fig. 4) have some interesting evolutionary derivations. First, the paraphyletic branching of the acoelomorphs at the base of the bilaterians suggests that the last common ancestor of all bilaterians, however different to present-day

acoels and nemertodermatids, was a small, benthonic, acoelomate worm, with an anterior concentration of nerve cells (primitive brain), a blind gut, a mesoderm which forms musculature and mesenchymal cells, and direct development. Second, the early branching of gnathostomulids within the lophotrochozoans agrees with their acoelomate nature and its presumed lack of a permanent anus, which maybe plesiomorphies shared with acoelomorphs and diploblasts. Next to gnathostomulids branch the gastrotrichs, which are also acoelomate worm-like animals, but with a through-gut with anus and not bearing jaws. According to the scenario drawn from Figure 4, gnathostomulids and gastrotrichs might represent a bridge from an acoel-like ancestor to more complex lophotrochozoans.

From this hypothetical lophotrochozoan acoelomate ancestor two further groups arose. The first clade (Clade III in Fig. 4) would have had a pseudocoelomate last common ancestor that kept the gnathiferan jaws of gnathostomulids (and, hence, lost in gastrotrichs) and from which derived rotifers, acantocephalans, and micrognathozoans. The second clade (Clade I + Clade II in Fig. 4) originated from an acoelomate ancestor, likely without gnathiferan jaws, which split into Clade II (acoelomate platyhelminths, entoprocts, cycliophorans and the coelomate bryozoan) and Clade I. This latter is formed by the coelomate Brachiozoa and the classical spiralian/thochophoran phyla such as nemertines, molluscs and, after segmentation evolved, the annelids and relatives. The

uncertain position of chaetognaths inside the Ecdysozoa turns difficult to produce an educated guess into their morphological evolution. Finally, the phylogenetic scheme produced makes clear that some morphological features, such as the presence and type of coelomic cavities or the type of cleavage, classically considered good phylogenetic characters for the metazoa, have independently appeared more than once.

Conclusions

To summarize, the study here reported demonstrates that the combination of complete taxon sampling, the application of adequate methodologies to avoid LBA, and careful compartmentalized analyses of problematic taxa allows inferring a highly supported tree of the bilaterian animals with better resolution than previous similar studies. Although some of the phylogenetic hypotheses here suggested had already been in part pointed out in previous studies, this is the first report which supports all of them using a single dataset.

Moreover, we suggest a new internal phylogeny for the Lophotrochozoa, and the overlapping of some clades of our with recent ESTs studies (i.e. compare our Clade I with the Clade C shown in Dunn et al. 2008) is, in our opinion, an extra prove in favour of the goodness of the methology used in our study. Furthermore, the vast taxonomic sampling available for the ribosomal genes allow us to test the position of some key clades poorly or not sampled for the new genetic markers yet. Altogether, these observations point to the fact

that the ribosomal RNA genes are still a reliable source for the study of deep divergences in the metazoan tree.

Supplementary Material

Supplementary Data Figure 1 (All-set NJ and PhyML trees) and Table 1 (species names and accession numbers used in this study) are available at http://

Authors' contributions

MR and JP designed the study and performed the analyses. MR, JP and JB prepared the manuscript. All three authors read and approved the final manuscript

Acknowledgements

References

Abouheif E, Zardoya R, Meyer A (1998) Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. J Mol Evol 47:394-405

Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R (2000) The new animal phylogeny: reliability and implications. Proc Natl Acad Sci U S A 97:4453-6

Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489-93

Ahlrichs WH (1997) Epidermal ultrastructure of Seison nebaliae and Seison annulatus, and a comparison of epidermal structures within the Gnathifera. Zoomorphology 117:41-48

Aleshin VV, Milyutina IA, Kedrova OS, Vladychenskaya NS, Petrov NB (1998) Phylogeny of Nematoda and Cephalorhyncha derived from 18S rDNA. J Mol Evol 47:597

Anderson FE, Cordoba AJ, Thollesson M (2004) Bilaterian phylogeny based on analyses of a region of the sodium-potassium ATPase alpha-subunit gene. J Mol Evol 58:252-68

Anderson FE, Swofford DL (2004) Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. Mol Phylogenet Evol 33:440-51

Baguñà J, Riutort M (2004) Molecular phylogeny of the Platyhelminthes. Can J Zool:168-193

Balavoine G, Adoutte A (2003) The Segmented Urbilateria: A Testable Scenario. Integr. Comp. Biol. 43:137-147

Balavoine G, de Rosa R, Adoutte A (2002) Hox clusters and bilaterian phylogeny. Molecular Phylogenetics and Evolution 24:366-373

Bergsten J (2005) A review of long-branch attraction. Cladistics 21:163-193

Boore JL, Macey JR, Medina M (2005) Sequencing and comparing whole mitochondrial genomes of animals. Methods Enzymol 395:311-48

Boore JL, Staton JL (2002) The Mitochondrial Genome of the Sipunculid Phascolopsis gouldii Supports Its Association with Annelida Rather than Mollusca. Mol Biol Evol 19:127-137

Cavalier-Smith T (1998) A revised six-kingdom system of life. Biol. Rev. 73:203

Clark RB (1969) Systematics and phylogeny: Annelida, Echiura, Sipuncula. Academic Press, New York & London

Cohen BL (2000) Monophyly of brachiopods and phoronids: reconciliation of molecular evidence with Linnaean classification (the subphylum Phoroniformea nov.).

Proceedings of the Royal Society B: Biological Sciences 267:225-231

Cohen BL, A.Gawthrop, Smith TC (1998) Molecular phylogeny of brachiopods and phoronids based on nuclear-encoded small subunit ribosomal RNA gene sequences. Philosophical Transactions of the Royal Society B: Biological Sciences 353:2039-2061

Cunningham CO, Aliesky H, Collins CM (2000) SEQUENCE AND SECONDARY STRUCTURE VARIATION IN THE GYRODACTYLUS (PLATYHELMINTHES: MONOGENEA) RIBOSOMAL RNA GENE ARRAY. Journal of Parasitology 86:567-576

de Rosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A (1999) HOX genes in brachiopods and priapulids and protostome evolution. Nature 399:772

Dixon MT, Hillis DM (1993) Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis. Mol Biol Evol 10:256-267

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SHD, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745-749

Ehlers U, Ahlrichs W, Lemburg C, Schmidt-Rhaesa A (1996) Phylogenetic systematization of the Nemathelminthes (Aschelminthes). Verh Dtsch Zool Ges 89:8

Felsenstein J (1978) Cases in which Parsimony or Compatibility Methods Will be Positively Misleading Syst Zool 27:401-410

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselin MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239:748-753

Funch P, Kristensen RM (1995) Cycliophora is a new phylum with affinities to Entoprocta and Ectoprocta. Nature 378:711-714

Garey J, Near T, Nonnemacher M, Nadler S (1996) Molecular evidence for Acanthocephala as a subtaxon of Rotifera. Journal of Molecular Evolution 43:287-292

Garey JR (2001) Ecdysozoa: the relationship between Cycloneuralia and Panarthropoda. Zool Anz 240:321

Garey JR, Schmidt-Rhaesa A (1998) The Essential Role of "Minor" Phyla in Molecular Studies of Animal Evolution. Amer. Zool. 38:907-917

Gillespie JJ, Yoder MJ, Wharton RA (2005) Predicted Secondary Structure for 28S and 18S rRNA from Ichneumonoidea (Insecta: Hymenoptera: Apocrita): Impact on Sequence Alignment and Phylogeny Estimation. Journal of Molecular Evolution 61:114-137

Giribet G (2002) Current advances in the phylogenetic reconstruction of metazoan evolution. A new paradigm for the Cambrian explosion? Mol Phylogenet Evol 24:345-57

Giribet G, Distel DL, Polz M, Sterrer W, Wheeler WC (2000) Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. Syst Biol 49:539-62

Giribet G, Sorensen MV, Funch P, Kristensen RM, Sterrer W (2004) Investigations into the phylogenetic position of Micrognathozoa using four molecular loci. Cladistics 20:1-13

Glenner H, Hansen AJ, Sorensen MV, Ronquist F, Huelsenbeck JP, Willerslev E (2004) Bayesian Inference of the Metazoan Phylogeny: A Combined Molecular and Morphological Approach. Current Biology 14:1644-1649

Glenner H, Hansen AJ, Sorensen MV, Ronquist F, Huelsenbeck JP, Willerslev E (2005a) Bayesian Inference of the Metazoan Phylogeny: A Combined Molecular and Morphological Approach. Current Biology 15:392-393

Glenner H, Hansen AJ, Sorensen MV, Ronquist F, Huelsenbeck JP, Willerslev E (2005b) Erratum: Bayesian Inference of the Metazoan Phylogeny: A Combined Molecular and Morphological Approach. Current Biology 15:392-393

Gonzalez P, Labarere J (2000) Phylogenetic relationships of Pleurotus species according to the sequence and secondary structure of the mitochondrial small-subunit rRNA V4, V6 and V9 domains. Microbiology 146:209-221

Guindon S, Gascuel O (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Systematic Biology 52:696 - 704

Gutell R (1992) Evolutionary characteristics of 16S and 23S rRNA structures. In: Hartman H, Matsuno, K. (ed) The origin and evolution of prokaryotic and eukaryotic cells. World Scientific, Hackensack, NJ, p 243–309

Gutell R (1994) Collection of Small Subunit (16S- and 16S-like) ribosomal RNA structures. Nucleic Acids Research 22:3502-3507

Gutell R, Weiser, B., Woese, CR., Noller, HF. (1985) Comparative anatomy of 16S-like ribosomal RNA. Prog Nucleic Acid Res Mol Biol 155–216

Gutmann WF (1981) Relationships between invertebrate phyla based on functional-mechanical analysis of the hydrostatic skeleton Amer. Zool.:63-81

Halanych KM (2004) The new view of animal phylogeny. Annual Review of Ecology, Evolution, and Systematics 35:229-256

Halanych KM, Bacheller JD, Aguinaldo AM, Liva SM, Hillis DM, Lake JA (1995) Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science 267:1641-3

Halanych KM, Passamaneck Y (2001) A Brief Review of Metazoan Phylogeny and Future Prospects in Hox-Research. Amer. Zool. 41:629-639

Hall T (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser.:95−98

Haszprunar GbPaP, 41–48. PtJZSER (1996) Platyhelminthes and Platihelminthomorpha—Paraphyletic taxa. J. Zoo. Syst. Evol. Res. 34:41-48

Helfenbein KG, Boore JL (2004) The Mitochondrial Genome of Phoronis architecta--Comparisons Demonstrate that Phoronids Are Lophotrochozoan Protostomes. Mol Biol Evol 21:153-157

Helfenbein KG, Fourcade HM, Vanjani RG, Boore JL (2004) The mitochondrial genome of Paraspadella gotoi is highly reduced and reveals that chaetognaths are a sister group to protostomes. PNAS 101:10639-10643

Hessling R, Westheide, W. (2002) Are Echiura derived from a segmented ancestor? Immunohistochemical analysis of the nervous system in developmental stages of Bonellia viridis. Journal of Morphology 252:100-113

Hickson RE, Simon C, Cooper A, Spicer GS, Sullivan J, Penny D (1996) Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol Biol Evol 13:150-169

Hwang S-K, Kim J-G (2000) Secondary Structural and Phylogenetic Implications of Nuclear Large Subunit Ribosomal RNA in the Ectomycorrhizal Fungus Tricholoma matsutake. Current Microbiology 40:250-256

Hyman LH (1951) The invertebrates: Platyhelminthes and Rhynchoela. The acelomate bilateria. Vol 2., New York

Hyman LH (1959) The invertebrates: Smaller coelomate. Groups Chaetognatha, Hemicordata, Pogonophora, Phoronida, Ectoprocta, Brachiopoda, Sipunculida. The coelomate bilateria. Vol 5., New York

Jobb G (2007) TREEFINDER version of June 2007. Munich, Germany. Distributed by the author at www.treefinder.de .

Jollie MJ (1973) The origin of the chordates Acta Zool:81-100

Kjer KM (1995) Use of rRNA Secondary Structure in Phylogenetic Studies to Identify Homologous Positions: An Example of Alignment and Data Presentation from the Frogs. Molecular Phylogenetics and Evolution 4:314-330

Kristensen RM, Funch P (2000) Micrognathozoa: A new class with complicated jaws like those of Rotifera and Gnathostomulida. Journal of Morphology 246:1-49

Lake JA (1990) Origin of the Metazoa. PNAS 87:763-766

Lemburg C (1995) Ultrastructure of the introvert and associated structures of the larvae of Halicryptus spinulosus (Priapulida). Zoomorphology 115:11

Littlewood DTJ, Telford MJ, Clough KA, Rohde K (1998) Gnathostomulida--An Enigmatic Metazoan Phylum from both Morphological and Molecular Perspectives. Molecular Phylogenetics and Evolution 9:72-79

Lydeard C, Holznagel WE, Schnare MN, Gutell RR (2000) Phylogenetic Analysis of Molluscan Mitochondrial LSU rDNA Sequences and Secondary Structures. Molecular Phylogenetics and Evolution 15:83-102

Mallatt J, Giribet G (2006) Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol Phylogenet Evol 40:772-94

Mallatt J, Sullivan J (1998) 28S and 18S rDNA sequences support the monophyly of lampreys and hagfishes. Mol Biol Evol 15:1706-18

Mallatt J, Winchell CJ (2002) Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. Mol Biol Evol 19:289-301

Mallatt JM, Garey JR, Shultz JW (2004) Ecdysozoan phylogeny and Bayesian inference: first use of nearly complete 28S and 18S rRNA gene sequences to classify

the arthropods and their kin. Mol Phylogenet Evol 31:178-91

Manylov OGa, Vladychenskaya, N.S.b , Milyutina, I.A.b , Kedrova, O.S.b , Korokhov, N.P.c , Dvoryanchikov, G.A.c , Aleshin, V.V.b , Petrov, N.B.b (2004) Analysis of 18S rRNA gene sequences suggests significant molecular differences between Macrodasyida and Chaetonotida (Gastrotricha). Mol Phylogenet Evol 30:850-854

Marletaz F, Martin E, Perez Y, Papillon D, Caubit X, Lowe CJ, Freeman B, Fasano L, Dossat C, Wincker P, Weissenbach J, Le Parco Y (2006) Chaetognath phylogenomics: a protostome with deuterostome-like development. Current Biology 16:R577-R578

Matus DQ, Copley RR, Dunn CW, Hejnol A, Eccleston H, Halanych KM, Martindale MQ, Telford MJ (2006) Broad taxon and gene sampling indicate that chaetognaths are protostomes. Current Biology 16:R575-R576

McHugh D (1997) Molecular evidence that echiurans and pogonophorans are derived annelids. PNAS 94:8006-8009

Medina M, Collins AG, Silberman JD, Sogin ML (2001) Evaluating hypotheses of basal animal phylogeny using complete sequences of large and small subunit rRNA. Proc Natl Acad Sci U S A 98:9707-12

Morin L (2000) Long branch attraction effects and the status of ''basal eukaryotes'': Phylogeny and structural analysis of the ribosomal RNA gene cluster of the free-living diplomonad *Trepomonas agilis*. J Eukaryot Microbiol 47:167–177

Morrison DA, Ellis JT (1997) Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa. Mol Biol Evol 14:428-441

Mugridge NB, Morrison DA, Johnson AM, Luton K, Dubey JP, Votýpka J, Tenter AM (1999) Phylogenetic relationships of the genus Frenkelia: a review of its history and new knowledge gained from comparison of large subunit ribosomal ribonucleic acid gene sequences. International Journal for Parasitology 29:957-972

Nielsen C (1995) Animal Evolution: Interrelationships of the Living Phyla. Oxford University Press, Oxford

Nielsen C (2001) Animal Evolution: Interrelationships of the Living Phyla.

Oxford University Press; 2Rev Ed edition (31 Mar 2001), Oxford

Olsen G, Natsuda H, Hagstrom R, Overbeek R (1994) FastD-NAML: A Tool for construction of phylogenetic trees of DNA sequences using maximum-likelihood. Comput Applic Biosci:41-48

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences:357-358

Papillon D, Perez Y, Caubit X, Le Parco Y (2004) Identification of Chaetognaths as Protostomes Is Supported by the Analysis of Their Mitochondrial Genome. Mol Biol Evol 21:2122-2129

Passamaneck Y, Halanych KM (2006) Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. Mol Phylogenet Evol 40:20-8

Peterson KJ, Eernisse DJ (2001) Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol Dev 3:170-205

Peterson KJ, Lyons JB, Nowak KS, Takacs CM, Wargo MJ, McPeek MA (2004) Estimating metazoan divergence times with a molecular clock. Proc Natl Acad Sci U S A 101:6536-41

Philippe H, Brinkmann H, Martinez P, Riutort M, Baguñà J (2007) Acoel Flatworms Are Not Platyhelminthes: Evidence from Phylogenomics. PLoS ONE 8:e717.

Philippe H, Chenuil A, Adoutte A (1994) Can the Cambrian Explosion Be Inferred through Molecular Phylogeny. Development:15-25

Philippe H, Telford MJ (2006) Large-scale sequencing and the new animal phylogeny. Trends Ecol Evol 21:614-20

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817-818

Rice ME (1985) Sipuncula: developmental evidence for phylogenetic inference. Oxford University Press, Oxford

Rieger RM (1976) Monociliated epidermal cells in Gastrotricha: significance for concepts of early metazoan evolution. Zool. Syst. Evol. Forsch:198-226

Rieger RM, Tyler S (1995) Sister-Group Relationship of Gnathostomulida and Rotifera-Acanthocephala Invertebrate Biology 114:186-188

Rieger RM, Tyler, S., Smith III, J. P. S., & Rieger, G. E. (1991) Platyhelminthes: Turbellaria. Wiley-Liss, New York

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under

mixed models. Bioinformatics 19:1572-1574

Ruiz-Trillo I, Paps J, Loukota M, Ribera C, Jondelius U, Baguna J, Riutort M (2002) A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. Proc Natl Acad Sci U S A 99:11246-51

Ruiz-Trillo I, Riutort M, Littlewood DT, Herniou EA, Baguna J (1999) Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. Science 283:1919-23

Scheltema AH (1993) Aplacophora as progenetic aculiferans and the coelomate origin of molluscs as the sister taxon of Sipuncula. Biological Bulletin 57-78

Schmidt-Rhaesa A (1996) The nervous system of Nectonema munidae and Gordius aquaticus, with implications for the ground pattern of Nematomorpha. Zoomorphology 116:133

Schmidt-Rhaesa A (2003) Old trees, new trees - is there any progress? Zoology 106:291-301

Schmidt-Rhaesa A, Bartolomaeus T, Lemburg C, Ehlers U, Garey JR (1998) The position of the Arthropoda in the phylogenetic system. Journal of Morphology 238:263-285

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502-504

Sempere LF, Cole CN, McPeek MA, Peterson KJ (2006) The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zoolog B Mol Dev Evol 306:575-88

Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. Systematic Biology:492-508

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246-1247

Sørensen MV (2003) Further structures in the jaw apparatus of Limnognathia maerski (Micrognathozoa), with notes on the phylogeny of the gnathifera. Journal of Morphology 255:131-145

Sørensen MV, P. Funch, E. Willerslev, A. J. Hansen, and J. Olesen. (2000) On the phylogeny of the Metazoa in light of Cycliophora and Micrognathozoa. Zool Anz 239:297

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses

with thousands of taxa and mixed models. Bioinformatics 22:2688-2690

Stechmann A, Schlegel, M. (1999) Analysis of the complete mitochondrial DNA sequence of the brachiopod Terebratulina retusa places Brachiopoda within the protostomes. Proceedings of the Royal Society B: Biological Sciences 266:2043-2043

Sterrer W, Mainitz M, Rieger R (1985) Gnathostomulida: enigmatic as ever. Clarendon Press

Struck T, Schult N, Kusen T, Hickman E, Bleidorn C, McHugh D, Halanych K (2007) Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evolutionary Biology 7:57

Telford MJ (2006) Animal phylogeny. Current Biolology 16:R981-5

Telford MJ, Lockyer AE, Cartwright-Finch C, Littlewood DT (2003) Combined large and small subunit ribosomal RNA phylogenies support a basal position of the acoelomorph flatworms. Proc Biol Sci 270:1077-83

Titus TA, Frost DR (1996) Molecular Homology Assessment and Phylogeny in the Lizard Family Opluridae (Squamata: Iguania). Molecular Phylogenetics and Evolution 6:49-62

Todaro MA, Telford MJ, Lockyer AE, Littlewood DTJ (2006) Interrelationships of the Gastrotricha and their place among the Metazoa inferred from 18S rRNA genes. Zoologica Scripta 35:251-259

Turbeville JM, Field KG, Raff RA (1992) Phylogenetic position of phylum Nemertini, inferred from 18S rRNA sequences: molecular data as a test of morphological character homology. Mol Biol Evol 9:235-249

Uchida H, Kitae K, Tomizawa K-I, Yokota A (1998) Comparison of the Nucleotide Sequence and Secondary Structure of the 5.8S Ribosomal RNA Gene of <i>Chlamydomonas tetragama</i> with those of Green Algae. DNA Sequence 8:403 - 408

Wallberg A, Thollesson M, Farris JS, Jondelius U (2004) The phylogenetic position of the comb jellies (Ctenophora) and the importance of taxonomic sampling. Cladistics 20:558-578

Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J (2002) Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. Mol Biol Evol 19:762-76

Winnepenninckx BMH, Backeljau T, Kristensen RM (1998) Relations of the new phylum Cycliophora. Nature 393:636-638

Xia X (2000) Phylogenetic relationship among horseshoe crab species: The effect of substitution models on phylogenetic analyses. Systematic Biology 49:87-100

Xia X, Xie, Z., Kjer, KM. (2003) 18S ribosomal RNA and tetrapod phylogeny. Systematic Biology 52:283-295

Zrzavý J (2003) Gastrotricha and metazoan phylogeny. Zoologica Scripta 32:61-81

Zrzavý J, Mihulka S, Kepka P, Bezdek A, Tietz D (1998) Phylogeny of the Metazoa Based on Morphological and 18S Ribosomal DNA Evidence. Cladistics 14:249-285

Supplementary data

The results obtained with NJ (supp. Data Fig. 1A), Treefinder and PhyML (supp. Data fig. 1B) show some groupings that never appear in the trees inferred by BI and RaxML (Fig. 2 main text). The differences found could most likely be explained by LBA affecting the first group of methods. As an example:

1. NJ (Supp Data Fig. 1A) places Micrognathozoa within the Arthropoda, while Chaetognaths appear splitting after acoelomorphs and sistergroup to the rest of bilaterians.

2. PhyML (Supp Data Fig 1B) recovers a group made up by nemertodermatids, chaetognaths and micrognathozoans inside ecdysozoans, as sistergroup to Panarthropoda

3. Treefinder (not shown) anomalies appear mainly in the subsets trees:

    a. *Acoelomorpha*: Tardigrada at the bilaterian base, splitting after *Nemertodermatida*; *Micrognathozoa* within Arthropoda.

    b. *Gnathifera*: *Urochordata* at bilaterian base; *Micrognathozoa* within Arthropoda.

    c. *Bryozoa: at the bilaterian base shows Bryozoa + ((Cycliophora+Entoprocta) + rest of bilaterians); Micrognathozoa within Arthropoda.*

    d. *Chaetognatha: Micrognathozoa within Arthropoda.*
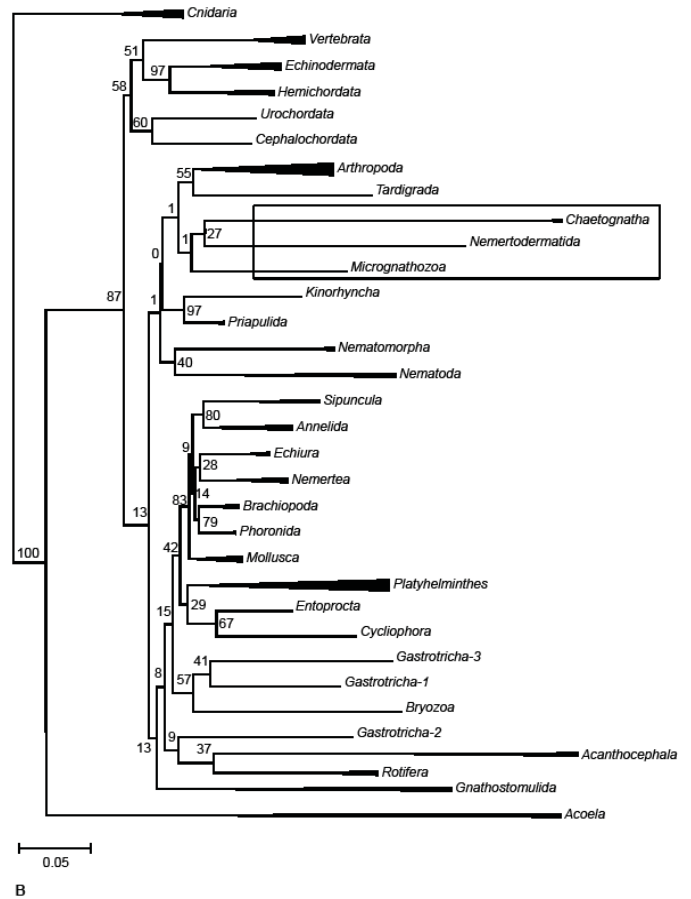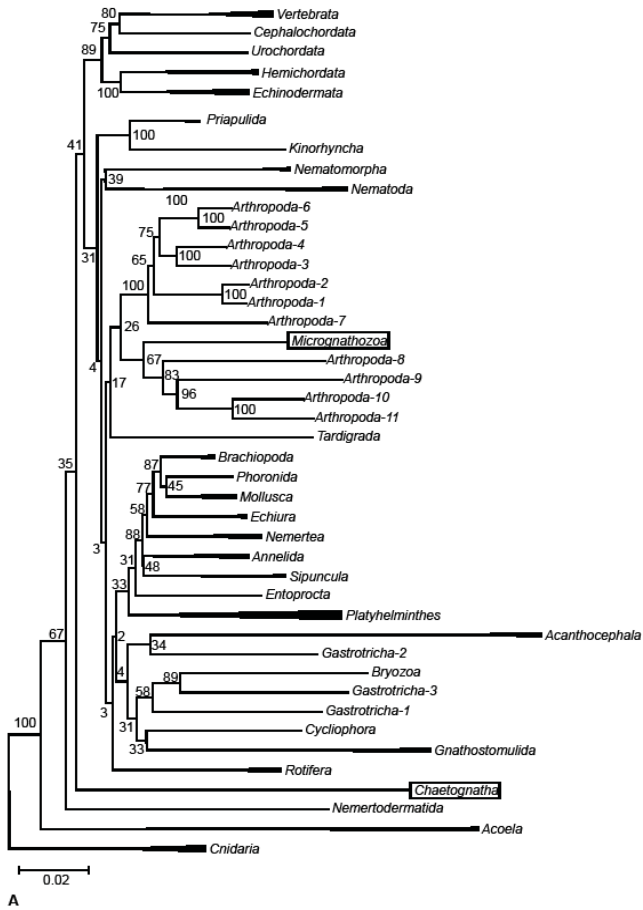
Supp. Data Figure 1. Topology obtained by NJ (A) and PhyML (B) with the All-set.

Bilaterian Phylogeny: A Wide Sampling for 13 Genes Provides a New Lophotrochozoa Phylogeny and Solves Some Elusive Relationships

Jordi Paps, Jaume Baguñà & Marta Riutort*

Departament de Genetica, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Spain.

Resum

Aquest capítol es el resultat final d'aquest projecte de tesi, essent el més recent dels nostres estudis i el més equilibrat en quant a mostratge taxonòmic i genètic. En aquest treball es conjuga la filosofia d'obtenir seqüències per nous marcadors del Capítol I i l'aproximació de maximitzar el mostratge taxonòmic del Capítol II. El resultat és un set de dades que consta de 90 taxa pertanyents a 27 grups de bilaterals i una extensió de 8.880 parells de bases provinents de 13 gens diferents.

Aquesta matriu es va fer servir per inferir una filogènia extensa dels bilaterals, mitjançant mètodes d'inferència de tipus probabilístic i tenint en compte les mesures presses al Capítol II per minimitzar els artefactes com el LBA. A diferència dels estudis anteriors, en aquest joc de dades es va permetre la presència de missing data (del 40%), el qual també ens va permetre afegir grups interessants com onicòfors, *Xenoturbella* o els pogonòfors.

El conjunt d'arbres mostra que:

1. L'existència dels tres superclades
2. Les filogènies internes de la majoria dels fílums estan ben resoltes.
3. Les filogènies internes de deuterostomats ens mostra la monofilia de cordats i els fílums dels Ambulacraria (hemicordats + equinoderms), situant *Xenoturbella* com a grup germà d'aquests últims.
4. Les relacions dins dels ecdisozous mostren als nemàtodes com a grup germà dels panartròpodes; els nematomorfs s'agrupen amb els priapúlids i cinorrincs (Scalidophora).
5. Es mostra una nova filogènia dels lofotrocozous, a on destaquen:
   a. gastrotrics i gnatostomúlids com a grups basals,
   b. l'estatus polifilètic dels lofoforats
   c. l'estatus parafilètic dels Spiralia degut a la posició dels braquiòpodes i foronidis (Brachiozoa) com a grup germà dels mol·luscs.
6. Es confirma que els acels i els nemertodermàtides són bilaterals basals i es suggereix l'afiliació dels quetògnats amb els ecdisozous.

Bilaterian Phylogeny: A Wide Sampling for 13 Genes Provides a New Lophotrochozoa Phylogeny and Solves Some Elusive Relationships

Jordi Paps, Jaume Baguna, & Marta Riutort [*]

Departament de Genetica, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona, Spain

During the past decade, great progress has been made in clarifying the relationships among bilaterian animals. Studies based on a limited number of markers established new hypotheses such as the existence of three superclades (Deuterostomia, Ecdysozoa and Lophotrochozoa) but left major questions unresolved. Many have sought to address these questions through phylogenomic approaches. However, many such studies have yielded trees that are inconsistent with a number of previous molecular findings. Poor taxon sampling and matrices that include a high proportion of missing data are two of the factors suspected to be responsible for these surprising results. Thus, existing datasets are biased (too few genes or too few taxa) and have failed to provide definitive answers. We performed phylogenetic analyses using a more balanced molecular matrix drawn from 13 genes (8,880 base pairs) in 90 taxa belonging to 27 bilaterian phyla. Probabilistic analyses of nucleotide sequences using adequate models of evolution robustly support the three superclades, the monophyly of *Chordata*, a spiralian clade including Brachiozoa, the basal position of a paraphyletic *Acoelomorpha*, and the ecdysozoan nature of *Chaetognatha*. This new phylogeny agrees with most classical molecular results, but also provides new insights into the relationships between lophotrochozoans and challenges the results obtained using high-throughput strategies, highlighting the problems associated with the current trend to increase gene number rather than taxa.

*Corresponding author, e-mail: mriutort@ub.edu

Introduction

Small ribosomal subunit RNA gene (18S rDNA or SSU) sequences were the first and most widely used source of information to establish the new, widely accepted bilaterian phylogeny, which features three large superclades, namely, the Lophotrochozoa, the Ecdysozoa and the Deuterostomia (Halanych 2004). The relationships within these superclades and the phylogenetic position of some enigmatic phyla still remain elusive to SSU analyses, in part due to long-branch attraction (LBA) artifacts (Felsenstein 1978) and to their recognized limited resolution (Abouheif, Zardoya, and Meyer 1998). In an attempt to overcome this problem, other markers such as the large ribosomal subunit RNA gene (28S or LSU; Mallatt and Giribet 2006; Passamaneck and Halanych 2006) or protein-coding genes (Ruiz-Trillo et al. 2002; Anderson, Cordoba, and Thollesson 2004; Peterson et al. 2004), were introduced. Although these approaches were instrumental in resolving some internal nodes of the tree, they are associated with similar problems to those encountered with SSU genes, namely stochastic errors and artifacts due to LBA (Philippe and Telford 2006). Alternative sources of information such as sequence signatures in the Hox genes (Balavoine, de Rosa, and Adoutte 2002), mitogenomics (Boore, Macey, and Medina 2005) or the microRNAs (Sempere et al. 2006) were subsequently proposed. Unfortunately, the binary nature of these qualitative characters (present/absent, shared/not shared) has only allowed definition of one clade versus another, and has not helped to resolve internal relationships.

Phylogenomics, based mainly on expressed sequence tag (EST) projects, is perhaps the leading approach through which to address this problem. Recent phylogenomic studies using up to 183 genes, together with the development of new models of protein evolution, have lent support to the three large superclades (Philippe, Lartillot, and Brinkmann 2005; Bourlat et al. 2006; Delsuc et al. 2006) but have been unable to produce a clear and robust internal phylogeny of these clades. Phylogenomics claims to overcome stochastic errors by incorporating a high number of characters; however, it is also susceptible to poor taxon sampling, systematic errors and paralogy problems (Philippe and Telford 2006). Indeed, reduced sampling might explain conflicting results either supporting or rejecting the Ecdysozoa over the Coelomata (Dopazo, Santoyo, and Dopazo 2004), as well as uncertainties related to the true position of the *Acoela* (Philippe et al. 2007) or the tunicates (Delsuc et al. 2006). Even the incorporation of one taxon per phylum does not guarantee a systematic error-free phylogeny (Philippe and Telford 2006), although the incorporation of new EST projects into future analyses will hopefully overcome those errors.

As of today, all of the molecular matrices used are highly asymmetric: they include either many phyla and few markers, or few phyla and many markers, each case bearing its own flaws. This produces dark areas in some regions of the bilaterian tree; e.g. the internal relationships within the Lophotrochozoa, the monophyletic status of *Chordata*, and the position of groups like the *Acoelomorpha* and the *Chaetognatha.* To provide a more robust basis on which to analyze the phylogenetic relationships of these problematic regions, we developed and analyzed a more balanced dataset. We evaluated twenty-six genes for their potential phylogenetic information, and 11 were selected. Sequences already present in GenBank were downloaded and 89 new sequences produced. The final matrix contains 90 representatives from 27 phyla and is 8,880 nucleotides long for the 11 protein-coding genes in addition to the two ribosomal RNA genes with a value of 40% missing data.

Methods

Sampling Thanks to the kind collaboration of many experts, 125 samples were collected for 96 species belonging to 31 phyla (see Suppl Data Table 1). Some groups where not sampled due to their rich representation in GenBank (such as *Craniata, Nematoda* and *Arthropoda*) or because we had no access to them (*Acantocephala*, *Micrognathozoa*,

Loricifera, Mesozoa, Pogonophora, Myzostomida).

Molecular techniques RNA was extracted from live animals or preserved in RNAlater (Ambion) with TRIzol reagent (Amersham Pharmacia Biotech) and cDNA was obtained by standard reverse transcription with M-MLV reverse transcriptase (Promega). When the yield of RNA or cDNA was low, the SMART protocol (Invitrogen) was used to increase the number of cDNA copies by PCR with adapters, according to the manufacturer's instructions (cDNAs obtained by SMART are noted in Supplementary Table 1). Gene fragments were amplified by PCR: 25 μl, with 1 unit of Dynazyme polymerase (Fynnzimes), 40 cycles of 45 seconds at 94°C, 45 seconds at the annealing temperature for each primer pair (Supplementary Table 2), and 55 seconds at 72°C. PCR products were purified (Microcon PCR columns, Millipore) and directly cycle-sequenced from both strands (Big Dye Terminator V.2.0, Applied Biosystems), ethanol precipitated, and run on an ABI Prism 3700 (Applied Biosystems) automated sequencer. Contigs were assembled with SeqEd VER. 1.0.3 (Applied Biosystems).

Gene selection The phylogenetic potential for 26 genes was evaluated, these genes had been preselected either because they have proven to be useful in previous phylogenetic studies, showed good phyla

sampling in the genetic databases or had interesting qualities regarding their rates of evolution. The genes assessed were as follows: 14-3-3, sodium-potassium ATPase alpha-subunit (ATPase alpha), cathepsin, cell division cycle 42 (Cdc 42), cAMP response element-binding (CREB), elongation factor alpha 1 (EF1), elongation factor alpha 2 (EF2), eukaryotic translation initiation factor 4E (eIF4E), forkhead, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), histone H3 (H3), intermediate filaments (IFs), laminin binding protein, myosin heavy chain type II (Myosin), ribosomal protein L13 (RP L13), ribosomal Protein L22 (RP L22), tropomyosin, tubulin, aldolase, ATP synthase beta-chain, methionine adenosyltransferase (MAT), DNA helicase, kinesin, phosphofructokinase (PFK), catalase and actin.

Gene sequences were downloaded from GenBank and each gene was aligned independently based on the amino acid sequence using ClustalX 1.81, and the resulting alignments were checked with Bioedit. Regions of ambiguous alignment were removed using Gblocks (Castresana 2000) with the default options except Allowed gap positions (set to "With half"), and the case of 28S, where the "Minimum number of sequences for flank positions" was set up to the same value as "Minimum number of sequences for a conserved position". For each gene, taxa lacking representatives were amplified and sequenced, and a Blast search was performed with the new sequences to confirm their identity. They were added to their respective alignments and their orthology was assessed with single-gene phylogenies. Genes which produced poorly resolved phylogenies (comb trees), contained poor taxon sampling or produced trees that were highly inconsistent with previous phylogenetic studies (e.g. placing molluscs inside chordates) were discarded. The final selected genes were ATPase alpha, GAPDH, H3, IFs, myosin, tropomyosin, aldolase, ATP synthase beta, MAT, PFK and catalase. Independent alignments for SSU and LSU sequences from a previous study (J.P., J.B. and M.R. unpublished data) were also used. In order to have the same number of OTUs for all the genes, the missing representatives for each gene were classified as missing data (filled with Ns).

Dataset The independent alignments were concatenated into a dataset containing 90 OTUs representing 27 phyla and 8,880 positions for 13 genes and has 40% missing data value. A summary of the sequences included in is provided in Supplementary Table 3, and a more detailed description for each OTU (species, classification, number of genes available and accession numbers) is shown in Supplementary Table 4 (CD). In all cases where it was possible, sequences for the same species or genus have been merged,

when it was not possible it is indicated in the table and by the name of the final OTU.

Phylogenetic analyses The saturation of the data was evaluated by comparing the observed nucleotide differences between pairs of sequences calculated with DAMBE (Xia and Xie 2001) with the inferred distance between the same two sequences determined by maximum likelihood in TreePuzzle (Schmidt et al. 2002). This program was also employed to carry out the likelihood mapping analyses, with the options estimation accurate (slow), Tamura-Nei 93 model, 4 Gamma categories and 1 invariable and 10,000 quartets. Modeltest (Posada and Crandall 1998) was used to determine the evolutionary model that showed the best fit for each gene, following the Akaike information criterion. The specified model ($GTR + \Gamma + I$) was used in all the algorithms. Bayesian inference trees were inferred with a parallelized version of MrBayes software (Ronquist and Huelsenbeck 2003), using a partitioned dataset and running 1,000,000 generations in 2 independent analyses with a sample frequency of 100. To obtain the consensus tree and bayesian inference supports, 500,000 generations were removed to discard trees sampled before likelihood values had reached a plateau. Maximum likelihood trees were inferred with RaxML (Stamatakis 2006), run using the model $GTR + \Gamma + I$ (4 gamma categories + 1 invariable), using a partitioned dataset and 1,000 bootstrap replicates.

Competing topologies were evaluated. The alternative trees were constructed using Treeview (Page 1996) using the original maximum likelihood inference tree as a template. The alternative topologies tested were based on previous studies or those found in our analyses (Table 1). CONSEL (Shimodaira and Hasegawa 2001) was run to perform the approximately unbiased test.

All the analyses were run on 4 different computers: 1) 2 PCs running Windows XP and SUSE Linux 10.0, 2) a supercomputer located at CESCA (Centre de Supercomputació de Catalunya, http://www.cesca.es), and 3) the Marenostrum supercomputer located at the Barcelona Supercomputing Center (http://www.BPPc.es).

Results & Discussion

Methodological Problems and dataset information

The experimental work endured two bottlenecks causing the missing data in our matrix. The first were the unsuccessful RNA extractions, mainly from some marine
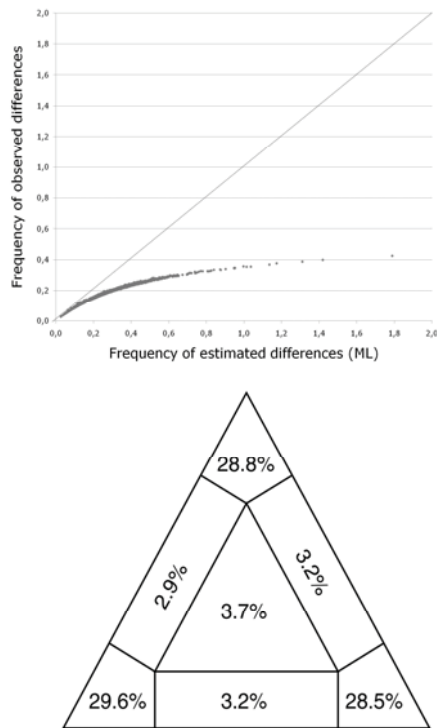

Fig. 1. Substitution saturation curves (up) and Likelihood Mapping analyses (down). The saturation curves show the frequency of observed differences between pairs of sequences (y axis) opposed to the inferred distance between the same two sequences determined by ML (x axis). The higher distances, belonging to the most divergent pairs of sequences, still show proportionality in the observed vs estimated ratio, indicating that we still have information for these comparisons. The Likelihood mapping analyses are represented as a triangle whose corners values indicate percentage of well-resolved phylogenies for all possible quartets, whereas central and lateral values are percentages of unresolved phylogenies.The dataset show a high proportion of well-resolved quartets, indicating that is phylogenetically informative.

tiny animals (*Porifera, Ctenophora, Placozoa, Myxozoa, Gnathostomulida, Cycliophora* or *Gastrotricha*), where a seemingly sufficient amount of tissue was available but the extraction did not yield enough quantity or quality for the following procedures. This explains the lack of some key phyla in our matrix, though these groups were collected several times. The second bottleneck was the PCR amplification, where some primer pairs worked successfully for some samples but not for others, while these very same samples amplified effectively for other pairs. In the end, in this study 89 new sequences were produced for the eleven selected genes. A complete list of the species and sequences used in this study can be seen in Supp Data Table 4.

The Likelihood mapping analysis and the saturation curves for the dataset are shown in Figure 1. The curve obtained does not reach a plateau, although for the higher distances the slope is less accentuated, indicating a certain degree of saturation. Regarding the Likelihood Mapping analysis, the dataset shows a high proportion of well-resolved quartets (87%), indicating that the dataset is phylogenetically informative.

A novel bilaterian phylogeny

Bayesian Inference (BI, Fig. 2) and Maximum Likelihood (ML, Fig. 3) result on a bilaterian phylogeny that sheds light on some current uncertainties. Regarding the statistical supports, in these analyses BI experiences a recognized "overconfidence" on its Bayesian posterior probabilities (BPP) while the ML phylogeny shows remarkably low Bootstrap Supports (BS) for the intermediate branches. However, both inference methods agree in the general topology that reproduces in both cases the three main bilaterian superclades with some minor disagreements in a few internal nodes.

Acoels and nemertodermatids

Both inference methods robustly show a paraphyletic *Acoelomorpha* as sistergroup to the other bilaterians. Their monophyly and their relationship to platyhelminths is rejected by the comparison of topologies with the approximately unbiased (AU) test (Table 1). This result confirms earlier studies showing them as a paraphyletic assemblage at the base of the bilaterians (Ruiz-Trillo et al. 1999; Ruiz-Trillo et al. 2002; Ruiz-Trillo et al. 2004; Sempere et al. 2007; Wallberg et al. 2007). But, it is in contrast with two recent phylogenomic reports, the first placing a single acoel species as a sister group to the deuterostomates (Philippe et al. 2007), an alternative also rejected here by comparison of topologies (Table 1, hypotheses 2 and 3),

and the second showing acoels within a clade that is sistergroup to the spiralian lophotrochozoans (although with no support, Dunn et al. 2008). While the analysis by Philippe et al (2007) contained more genes than those used here (68 genes vs 13), the number of phyla (13 vs 27) and OTUs (51 vs 90) was clearly lower than in our analysis. The higher number of *Acoelomorpha* included here (five acoels and two nemertodermatids) shortened the basal branch of that group and probably provided greater stability to this clade, thereby making our results more reliable. Moreover, we recover a partial internal phylogeny of the acoels that is in concordance to a recent systematic proposal based on molecular and sperm structure data (Hooge and Tyler 2006).

Deuterostomia and *Xenoturbella*

The Deuterostomia appear as a monophyletic clade, but with low support in both inference methods (Figure 2 and 3). Deuterostomates split into a robust *Chordata* and a weak clade including *Xenoturbella* and a strong Ambulacraria (*Hemichordata + Echinodermata)*. The low support for the deuterostomates has also been seen in many recent molecular studies based on ribosomal data (Mallatt and Winchell 2002), ESTs (Delsuc et al. 2006; Philippe et al. 2007) or mtDNA data (Bourlat et al. 2006). This low support for Deuterostomia in our results can be explained by the unstable nature of *Xenoturbella*, only 57% of the BI trees
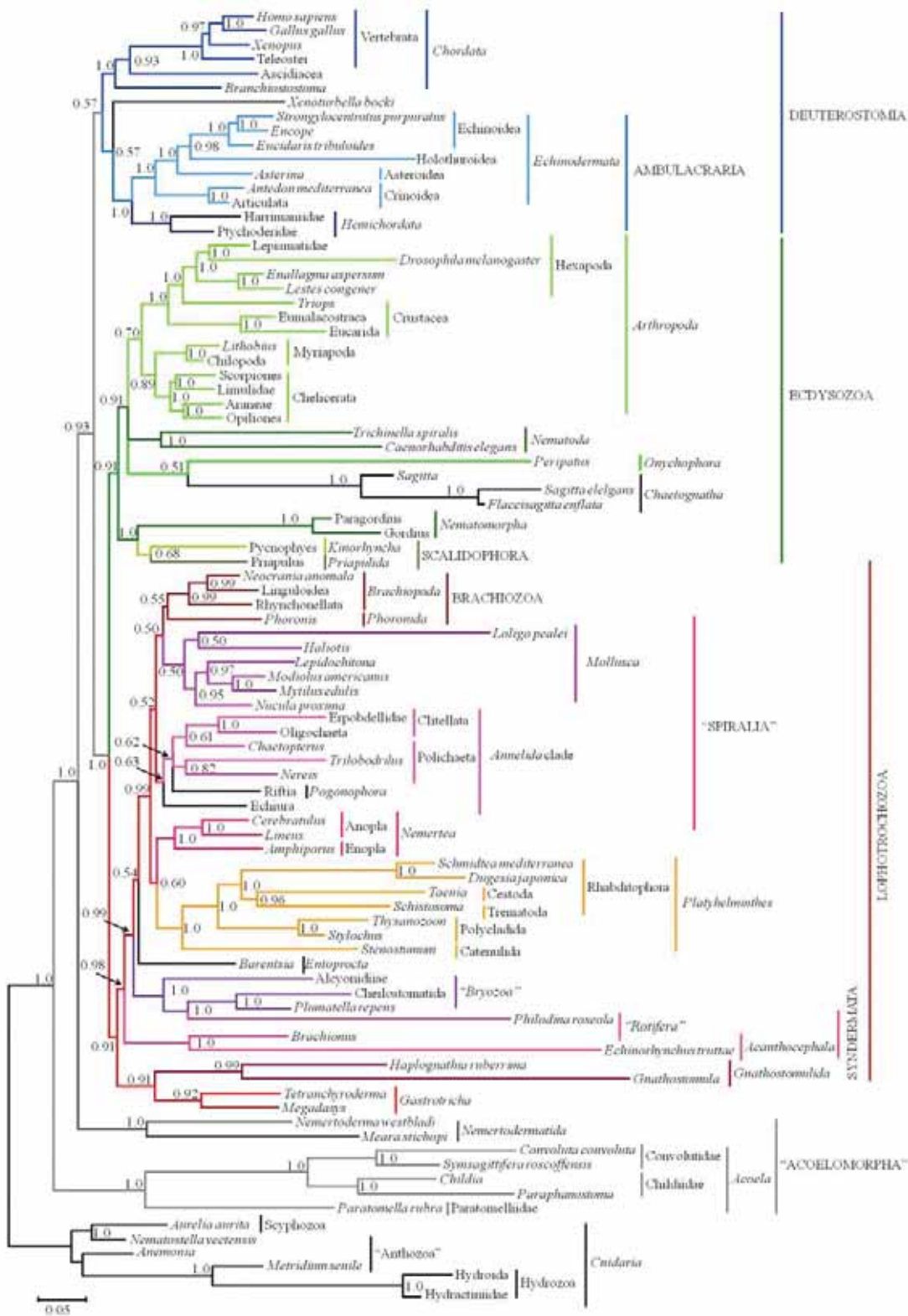
Figure 2. Phylogenetic tree for the Bayesian inference method. Bayesian Posterior Probabilities (BPP) are indicated in the nodes. The scale bar indicates the number of changes per site. Paraphyletic clades are indicated within quotation marks. For species names corresponding to each terminal see Supplementary Table 4.

Figure 3. Phylogenetic tree for the maximum likelihood method. Bootstrap Support values (BS) are indicated in the nodes. The scale bar indicates the number of changes per site. Paraphyletic clades are indicated within quotation marks. For species names corresponding to each terminal see Supplementary Table 4.
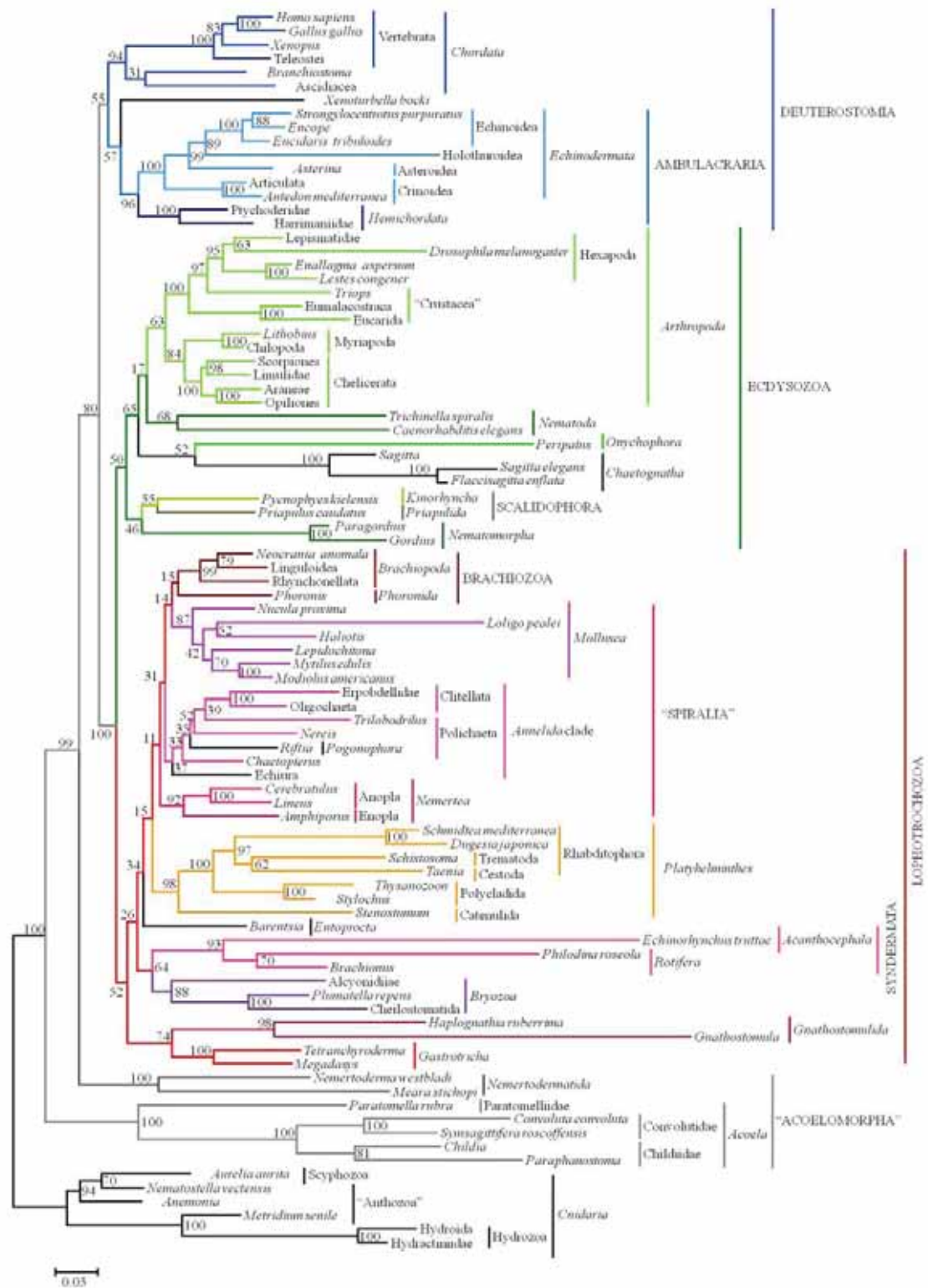
Table 1. Comparison of topologies using the approximately unbiased test

Topology	AU
1. **Original ML tree (Fig. 1)**	0.927
2. *Acoelomorpha* sistergroup to Ambulacraria (Philippe et al. 2007)	<0,001*
3. *Acoelomorpha* sistergroup to *Platyhelminthes* (Rieger 1991)	<0,001*
4. *Xenoturbella* sistergroup to Nephrozoa[#]	0.294
5. *Nematoda + Nematomorpha*, sistergroup to *Priapulida+Kinorhyncha* (Dunn et al. 2008)	0.091
6. *Nematoda + Nematomorpha*, sistergroup to *Arthropoda* (Mallatt and Giribet 2006)	0.087
7. *Chaetognatha* sistergroup to Ecdysozoa (Zrzavý et al. 1998)	0.003*
8. *Chaetognatha* sistergroup to Lophotrochozoa (Matus et al. 2006)	0.014*
9. *Chaetognatha* sistergroup to Protostomia (Marletaz et al. 2006)	0.004*
10. *Bryozoa+Entoprocta*, sistergroup to *Rotifera+Acanthocephala*[#]	0.001*
11. *Bryozoa+Entoprocta*, sistergroup to spiralian clade (Hausdorf et al. 2007)	0.109

*Hypothesis rejected when P values < 0.05 for the approximately unbiased test. The studies in which the hypothesis is found or referenced are indicated. # Variations based on our trees. See text for discussion.

show it as sistergroup to Ambulacraria (first shown in Bourlat et al. 2003; Bourlat et al. 2006). The best BI tree (not shown), position it splitting after *Nemertodermatida,* as sistergroup to the rest of bilaterians. Moreover, this later position is not rejected by the AU test (Table 1, hypothesis 4), but further data is needed to corroborate their position. If *Xenoturbella* is removed from the dataset, the deuterostomes support increases to 83% BS (results not shown). Regarding chordates, although ML shows *Cephalochordata + Urochordata* as sister group to *Vertebrata,* but with low BS, (also shown in Zrzavý et al. 1998; Giribet et al. 2000; Peterson and Eernisse 2001), we credit the more robust BI phylogeny showing a clade with urochordata as a sister group to vertebrates (in agreement with Bourlat et al. 2006; Delsuc et al. 2006; Philippe et al. 2007; Dunn et al. 2008). It is noteworthy that, although we have not representatives of the Class Ophiuroidea, the other echinoderm classes robustly conform to previous studies (Littlewood et al. 1997).

Ecdysozoa and *Chaetognatha* The Ecdysozoa phylogeny shows two main clades: 1) Scalidophora (*Priapulida + Kinorhyncha*) plus *Nematomorpha* with maximum BPP and 2) a clade including *Nematoda, Onychophora*, *Chaetognatha* and *Arthropoda* with moderate BPP and low BS. Scalidophora monophyly is widely accepted by morphology (Ehlers et al. 1996; Schmidt-Rhaesa 1996) and molecules (Aguinaldo et al. 1997; Aleshin et al. 1998; Giribet et al. 2000; Peterson and

Eernisse 2001). Their relationship to nematomorpha was originally proposed by Malakov (Malakhov 1980; Adrianov and Malakhov 1995) and also hinted in Zrzavý (2001). But nematomorphs are usually linked to nematodes forming the Nematoida (Zrzavý et al. 1998; Giribet et al. 2000; Peterson and Eernisse 2001), and the AU test can not reject the Nematoida monophyly with our data (Table 1, hypotheses 5 and 6).

The second ecdysozoan group uniting nematodes, onychophorans, chaetognaths and arthropods is a candidate to be the consequence of a LBA artefact; to test this hypothesis we ran two ML analyses (not shown).The first, excluding chaetognaths, showed onychophorans within arthropods with 94% support, while the second, excluding onychophorans, places chaetognaths within nematodes (7%). These results point to an internal LBA effect between nematodes, onycophorans and the chaetognath which can also explain the lack of resolution for the relationships among the main ecdysozoan clades. Despite these problems, arthropods show a reliable internal phylogeny grouping Myriapoda with Chelicerata and Hexapoda with Crustacea, both groups also recovered in SSU (Peterson and Eernisse 2001), mitochondrial DNA (Hwang et al. 2001) and recent phylogenomic studies (Dunn et al. 2008).

The controversial phylogenetic position of *Chaetognatha* is one of the biggest conundrums in animal phylogeny. Molecular phylogenies drawn from SSU (Telford and Holland 1993; Zrzavý et al. 1998; Giribet et al. 2000) place them together with other fast-clock groups, most likely to be a product of LBA artifacts. Recent data on Hox cluster genes suggest a new position close to the base of the Bilateria (Papillon et al. 2003), while phylogeny based on ribosomal genes (Mallatt and Winchell 2002), mtDNA (Helfenbein et al. 2004; Papillon et al. 2004) and multigenic approaches (Marletaz et al. 2006; Matus et al. 2006; Philippe et al. 2007) place them as sister group to the protostomates or within that group.

Chaetognaths are placed within ecdysozoans (0.91 BPP) in our analyses and the comparison of topologies clearly rejects chaetognaths as sister group to all the protostomates, to Lophotrochozoa or to Ecdysozoa. Placement of chaetognaths inside the ecdysozoans has recently been suggested (Helmkampf 2007), albeit with much lower BPP values. Our conclusion is that LBA is partially affecting our results regarding the Ecdysozoa internal phylogeny and hence the solution to the chaetognaths riddle, although the BI support and the comparison of topologies point to an affiliation among them.

Lophotrochozoa Lophotrochozoa were first defined as the last common ancestor of annelids, molluscs, the lophophorate phyla (*Brachiopoda, Phoronida*, and *Bryozoa*), and all the descendants of that ancestor (Halanych et al. 1995). Hox gene residues (Balavoine, de Rosa, and Adoutte 2002) and other molecular markers (Ruiz-Trillo et al. 2002; Anderson, Cordoba, and Thollesson 2004; Peterson, McPeek, and Evans 2005) support this superclade. Their internal relationships, however, are far from settled, as highlighted by the wide variety of different proposals suggested (Zrzavý et al. 1998; Giribet et al. 2000; Peterson and Eernisse 2001; Mallatt and Winchell 2002; Passamaneck and Halanych 2006). Interestingly, the internal phylogeny obtained here has never previously been recovered, probably because earlier studies included fewer genes and taxon sampling.

A well-supported clade constituted by *Gnathostomulida* and *Gastrotricha* is shown as the first splitting lophotrochozoan lineage. A close relationship among gnathostomulans and gastrotrichans has been suggested on basis of the protonephridial ultrastructure and the monociliated epidermal cells (Rieger 1976; Sterrer, Mainitz, and Rieger 1985; Zrzavý et al. 1998). Also SSU and total evidence approaches have placed them together but within Ecdysozoa (Neotrichozoa; Zrzavý et al. 1998), or separated in the Lophotrochozoa (Giribet et al. 2000;

Todaro et al. 2006) or even as basal bilaterians (Peterson and Eernisse, 2001). Unfortunately these two phyla were not included in any of the recent SSU + LSU analyses (Mallat and Whinchell, 2002; Halanych and Passamaneck, 2005) and a recent ESTs study was not able to resolve their relationships (Dunn et al. 2008). Our trees propose a new position for them as the most basal lophotrochozoans and rejects other proposals such as Gnathifera (Rieger and Tyler 1995; Ahlrichs 1997) or Cycloneuralia (*sensu lato, Gastrotricha* + Nematoida + Scalidophora; Nielsen 2001).

Our results relate *Rotifera* and *Acanthocephala*, although BI tree shows a polyphyletic Rotifera. The relationship among rotiferans and acanthocephalans has been suggested by morphology (see review in Garey and Schmidt-Rhaesa 1998) and SSU (Winnepenninckx et al. 1995; named Syndermata in Zrzavy et al. 1998). The main difference between ML and BI inferences can be found in this node: while ML shows bryozoans and the robust syndermatans clustering together (Fig. 3), BI tree shows acanthocephalan and a polyphyletic rotifer splitting before the bryozoans that include the rogue rotiferan *Philodina* within them (Fig. 2). The unexpected syndermatan polyphyly is likely a consequence of the high proportion of missing data in *Philodina* (56%) and the acanthocephalan (63%) or maybe BI inference being trapped in local minima.

This discrepancy also affects the reliable positioning of *Bryozoa*. Former SSU studies have already shown bryozoans not to be lophophorates (Cohen 2000), although their position remained unsettled based on ribosomal data (Zrzavý et al. 1998; Giribet et al. 2000; Peterson and Eernisse 2001; Passamaneck and Halanych 2006). Recent ESTs analyses placed bryozoans close to spiralians (Hausdorf et al. 2007; Dunn et al. 2008) in one case together with entoproctans in the other not. Both of our trees agree with phylogenomic studies in the position of the entoproctan, but are not able to settle the bryozoans placement. Our AU test rejects the monophyly of *Bryozoa+Entoprocta* if the entoproctan is forced inside the Syndermata+*Bryozoa*, but does not reject it if the bryozoans are removed from syndermatans side and placed with entoproctans as sistergroup to spiralians. Therefore the position of entoprocta as spiralian sistergroup is highly supported by our analyses, but not much can be said for bryozoans.

Monophyly of the Spiralia (animals bearing a spiral-quartet cleavage) has been recovered in many molecular studies. However, once again, their internal phylogeny has never been solved with high support. In our trees, either *Nemertea* or a clade of platyhelminths and nemertines are the most basal spiralians. Molecular studies placed the *Platyhelminthes* as basal lophotrochozoans (Ruiz-Trillo et al. 1999)

or within the Platyzoa (Giribet et al. 2000). In our results platyhelminths either constitute a clade with nemertines as sistergroup to the rest of Spiralia+Brachiozoa (BI, Fig. 2) or, alone, are the sister group to the Spiralia+Brachiozoa clade (including nemertines, ML, Fig. 3). Nemertines were traditionally allied with platyhelminths because of the lack of coelom (Parenchymia; Nielsen 1995) or by the larva features (Nielsen 2001). Nevertheless, further works have shown them to be coelomates (Turbeville, Field, and Raff 1992) (de Rosa et al. 1999; Balavoine, de Rosa, and Adoutte 2002).

Although we cannot distinguish among the two alternatives found for Platyhelminthes and Nemertea, the BI tree strongly supports the platyhelminths relationship with the spiralians+Brachiozoa, and both trees show a basal situation of nemertines with respect to the rest of spiralian phyla (Zrzavý, 1998; Giribet et al., 2000; Peterson & Eernisse, 2001). Moreover, the internal phylogeny of the platyhelminths is robust and highly congruent with the modern systematics of the group (see a review in Baguñà and Riutort 2004)

The Spiralia appear paraphyletic due to the inclusion of *Phoronida + Brachiopoda* within this group. The *Phoronida + Brachiopoda* clade, often named Brachiozoa or Phoronozoa, is

increasingly recovered in recent molecular phylogenies (Zrzavý et al. 1998; Cohen 2000; Peterson and Eernisse 2001). The relationship between Brachiozoa and Spiralia has been hinted by recent ESTs studies (albeit relating them with nemertines rather than molluscs, Dunn et al. 2008) and recent paleontological studies have also pointed out a likely affiliation to molluscs (Vinther and Nielsen 2005), the later in concordance with our results. The other spiralian group relates *Echiura* and *Pogonophora* with *Annelida*. However, they are not nested within the annelids, as has been found in other studies (Hessling 2002; Bleidorn, Vogt, and Bartolomaeus 2003; Struck et al. 2007). The relationships among the annelids (paraphyletic basal polychaeta and the clade clitellata formed by oligochaeta and the erpobdellid hirudinea) are consistent with their contemporary phylogeny (Struck et al. 2007).

Concluding remarks

A fair balance between number of taxa and number of characters sampled seems a necessary prerequisite to obtain a robust phylogeny of the bilaterians. Furthermore, a deep taxon sampling, the use of probabilistic methods and adequate models has helped to recover a novel bilaterian phylogeny. Our results recover the monophyly of Deuterostomia, Protostomia, Lophotrochozoa and Ecdysozoa already seen in previous studies,

and shed new light in some dark, conflicting areas of the bilaterian tree. These regions appear better resolved here than with current ESTs analyses, namely the basal relationships of the Lophotrochozoa, the status of *Chordata*, the position of a paraphyletic *Acoelomorpha* as the earliest branching extant bilaterians, and the ecdysozoan nature of *Chaetognatha.* Our phylogeny also holds congruent internal relatitonships for many phyla. Unfortunately, our trees don't show high supports for some regions of the tree, as the status of *Xenoturbella,* the internal relationships of Ecdysozoa and the intermediate branches of Lophotrochozoa. A similar balanced approach, incorporating more genes or better EST collections, and better taxon sampling, will substantially improve our understanding of bilaterian evolution.

Author contributions JP obtained part of the biological material and generated the new molecular sequences. MR and JP designed the study and performed the analyses. MR, JP and JB prepared the manuscript. All authors read and approved the final manuscript.

Bibliography

Abouheif, E., R. Zardoya, and A. Meyer. 1998. Limitations of metazoan 18S rRNA sequence data: implications for reconstructing a phylogeny of the animal kingdom and inferring the reality of the Cambrian explosion. J Mol Evol **47**:394-405.

Adrianov, A. V., and V. V. Malakhov. 1995. The phylogeny and classification of the phylum Cephalorhyncha. Zoosyst. Ross. **3**:181-201.

Aguinaldo, A. M., J. M. Turbeville, L. S. Linford, M. C. Rivera, J. R. Garey, R. A. Raff, and J. A. Lake. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature **387**:489-493.

Ahlrichs, W. H. 1997. Epidermal ultrastructure of Seison nebaliae and Seison annulatus, and a comparison of epidermal structures within the Gnathifera. Zoomorphology **117**:41-48.

Aleshin, V. V., I. A. Milyutina, O. S. Kedrova, N. S. Vladychenskaya, and N. B. Petrov. 1998. Phylogeny of Nematoda and Cephalorhyncha Derived from 18S rDNA. Journal of Molecular Evolution **47**:597-605.

Anderson, F. E., A. J. Cordoba, and M. Thollesson. 2004. Bilaterian phylogeny based on analyses of a region of the sodium-potassium ATPase alpha-subunit gene. J Mol Evol **58**:252-268.

Baguñà, J., and M. Riutort. 2004. Molecular phylogeny of the Platyhelminthes. Can J Zool:168-193.

Balavoine, G., R. de Rosa, and A. Adoutte. 2002. Hox clusters and bilaterian phylogeny. Molecular Phylogenetics and Evolution **24**:366-373.

Bleidorn, C., L. Vogt, and T. Bartolomaeus. 2003. New insights into polychaete phylogeny (Annelida) inferred from 18S rDNA sequences. Molecular Phylogenetics and Evolution **29**:279-288.

Boore, J. L., J. R. Macey, and M. Medina. 2005. Sequencing and comparing whole mitochondrial genomes of animals. Methods Enzymol **395**:311-348.

Bourlat, S. J., T. Juliusdottir, C. J. Lowe, R. Freeman, J. Aronowicz, M. Kirschner, E. S. Lander, M. Thorndyke, H. Nakano, A. B. Kohn, A. Heyland, L. L. Moroz, R. R. Copley, and M. J. Telford. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature.

Bourlat, S. J., C. Nielsen, A. E. Lockyer, D. T. J. Littlewood, and M. J. Telford. 2003. Xenoturbella is a deuterostome that eats molluscs. Nature **424**:925-928.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution **17**:540-552.

Cohen, B. L. 2000. Monophyly of brachiopods and phoronids: reconciliation of molecular evidence with Linnaean classification (the subphylum Phoroniformea nov.). Proceedings of the Royal Society B: Biological Sciences **267**:225-231.

de Rosa, R., J. K. Grenier, T. Andreeva, C. E. Cook, and A. Adoutte. 1999. HOX genes in brachiopods and priapulids and protostome evolution. Nature **399**:772.

Delsuc, F., H. Brinkmann, D. Chourrout, and H. Philippe. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature **439**:965-968.

Dopazo, H., J. Santoyo, and J. Dopazo. 2004. Phylogenomics and the number of characters required for obtaining an accurate phylogeny of eukaryote model species. Bioinformatics **20**:i116-121.

Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, G. W. Rouse, M. Obst, G. D. Edgecombe, M. V. Sorensen, S. H. D. Haddock, A. Schmidt-Rhaesa, A. Okusu, R. M. Kristensen, W. C. Wheeler, M. Q. Martindale, and G. Giribet. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature **452**:745-749.

Ehlers, U., W. Ahlrichs, C. Lemburg, and A. Schmidt-Rhaesa. 1996. Phylogenetic systematization of the

Nemathelminthes (Aschelminthes). Verh Dtsch Zool Ges **89**:8.

Felsenstein, J. 1978. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading Syst Zool **27**:401-410.

Garey, J. R., and A. Schmidt-Rhaesa. 1998. The Essential Role of "Minor" Phyla in Molecular Studies of Animal Evolution. Amer. Zool. **38**:907-917.

Giribet, G., D. L. Distel, M. Polz, W. Sterrer, and W. C. Wheeler. 2000. Triploblastic relationships with emphasis on the acoelomates and the position of Gnathostomulida, Cycliophora, Plathelminthes, and Chaetognatha: a combined approach of 18S rDNA sequences and morphology. Syst Biol **49**:539-562.

Halanych, K. M. 2004. The new view of animal phylogeny. Annual Review of Ecology, Evolution, and Systematics **35**:229-256.

Halanych, K. M., J. D. Bacheller, A. M. Aguinaldo, S. M. Liva, D. M. Hillis, and J. A. Lake. 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science **267**:1641-1643.

Hausdorf, B., M. Helmkampf, A. Meyer, A. Witek, H. Herlyn, I. Bruchhaus, T. Hankeln, T. H. Struck, and B. Lieb. 2007. Spiralian Phylogenomics Supports the Resurrection of Bryozoa Comprising Ectoprocta and Entoprocta. Mol Biol Evol **24**:2723-2729.

Helfenbein, K. G., H. M. Fourcade, R. G. Vanjani, and J. L. Boore. 2004. The mitochondrial genome of Paraspadella gotoi is highly reduced and reveals that chaetognaths are a sister group to protostomes. PNAS **101**:10639-10643.

Helmkampf, M., Bruchhaus, I. & Hausdorf, B. 2007. Multigene analysis of lophophorate and chaetognath phylogenetic relationships. Molecular Phylogenetics and Evolution **In Press, Corrected Proof**.

Hessling, R., Westheide, W. 2002. Are Echiura derived from a segmented ancestor? Immunohistochemical analysis of the nervous system in developmental stages of Bonellia viridis. Journal of Morphology **252**:100-113.

Hooge, M. D., and S. Tyler. 2006. Concordance of molecular and morphological data: The example of the Acoela. Integr. Comp. Biol. **46**:118-124.

Hwang, U. W., M. Friedrich, D. Tautz, C. J. Park, and W. Kim. 2001. Mitochondrial protein phylogeny joins myriapods with chelicerates. Nature **413**:154-157.

Littlewood, D. T. J., A. B. Smith, K. A. Clough, and R. H. Emson. 1997. The interrelationships of the echinoderm classes: morphological and molecular evidence. Biological Journal of the Linnean Society **61**:409-438.

Malakhov, V. 1980. Cephalorhyncha, a new type of animal kingdom uniting Priapulida, Kinorhyncha, Gordiacea, and a system of aschelminthes worms. . Zool. Zh. **59**:485–499.

Mallatt, J., and G. Giribet. 2006. Further use of nearly complete 28S and 18S rRNA genes to classify Ecdysozoa: 37 more arthropods and a kinorhynch. Mol Phylogenet Evol **40**:772-794.

Mallatt, J., and C. J. Winchell. 2002. Testing the new animal phylogeny: first use of combined large-subunit and small-subunit rRNA gene sequences to classify the protostomes. Mol Biol Evol **19**:289-301.

Marletaz, F., E. Martin, Y. Perez, D. Papillon, X. Caubit, C. J. Lowe, B. Freeman, L. Fasano, C. Dossat, P. Wincker, J. Weissenbach, and Y. Le Parco. 2006. Chaetognath phylogenomics: a protostome with deuterostome-like development. Current Biology **16**:R577-R578.

Matus, D. Q., R. R. Copley, C. W. Dunn, A. Hejnol, H. Eccleston, K. M. Halanych, M. Q. Martindale, and M. J. Telford. 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. Current Biology **16**:R575-R576.

Nielsen, C. 1995. Animal Evolution: Interrelationships of the Living Phyla. Oxford University Press, Oxford.

Nielsen, C. 2001. Animal Evolution: Interrelationships of the Living Phyla. Oxford University Press; 2Rev Ed edition (31 Mar 2001), Oxford.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences:357-358.

Papillon, D., Y. Perez, X. Caubit, and Y. Le Parco. 2004. Identification of Chaetognaths as Protostomes Is Supported by the Analysis of Their Mitochondrial Genome. Mol Biol Evol **21**:2122-2129.

Papillon, D., Y. Perez, L. Fasano, Y. Le Parco, and X. Caubit. 2003. Hox gene survey in the chaetognath Spadella cephaloptera: evolutionary implications. Development Genes and Evolution **213**:142-148.

Passamaneck, Y., and K. M. Halanych. 2006. Lophotrochozoan phylogeny assessed with LSU and SSU data: evidence of lophophorate polyphyly. Mol Phylogenet Evol **40**:20-28.

Peterson, K. J., and D. J. Eernisse. 2001. Animal phylogeny and the ancestry of bilaterians: inferences from morphology and 18S rDNA gene sequences. Evol Dev **3**:170-205.

Peterson, K. J., J. B. Lyons, K. S. Nowak, C. M. Takacs, M. J. Wargo, and M. A. McPeek. 2004. Estimating metazoan divergence times with a molecular clock. Proc Natl Acad Sci U S A **101**:6536-6541.

Peterson, K. J., M. A. McPeek, and D. A. D. Evans. 2005. Tempo and mode of early animal evolution: inferences from rocks, Hox, and molecular clocks. Paleobiology **31**:36-55.

Philippe, H., H. Brinkmann, P. Martinez, M. Riutort, and J. Baguñà. 2007. Acoel Flatworms Are Not Platyhelminthes: Evidence from Phylogenomics. PLoS ONE **8**:e717.

Philippe, H., N. Lartillot, and H. Brinkmann. 2005. Multigene Analyses of Bilaterian Animals Corroborate the Monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol **22**:1246-1253.

Philippe, H., and M. J. Telford. 2006. Large-scale sequencing and the new animal phylogeny. Trends Ecol Evol **21**:614-620.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817-818.

Rieger, R. M. 1976. Monociliated epidermal cells in Gastrotricha: significance for concepts of early metazoan evolution. Zool. Syst. Evol. Forsch:198-226.

Rieger, R. M., and S. Tyler. 1995. Sister-Group Relationship of Gnathostomulida and Rotifera-Acanthocephala Invertebrate Biology **114**:186-188.

Rieger, R. M., Tyler, S., Smith III, J. P. S., & Rieger, G. E. 1991. Platyhelminthes: Turbellaria. Wiley-Liss, New York.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572-1574.

Ruiz-Trillo, I., J. Paps, M. Loukota, C. Ribera, U. Jondelius, J. Baguna, and M. Riutort. 2002. A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians. Proc Natl Acad Sci U S A **99**:11246-11251.

Ruiz-Trillo, I., M. Riutort, H. M. Fourcade, J. Baguna, and J. L. Boore. 2004. Mitochondrial genome data support the basal position of Acoelomorpha and the polyphyly of the Platyhelminthes. Mol Phylogenet Evol **33**:321-332.

Ruiz-Trillo, I., M. Riutort, D. T. Littlewood, E. A. Herniou, and J. Baguna. 1999. Acoel flatworms: earliest extant bilaterian Metazoans, not members of Platyhelminthes. Science **283**:1919-1923.

Schmidt-Rhaesa, A. 1996. The nervous system of Nectonema munidae and Gordius aquaticus, with implications for the ground pattern of Nematomorpha. Zoomorphology **116**:133.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502-504.

Sempere, L. F., C. N. Cole, M. A. McPeek, and K. J. Peterson. 2006. The phylogenetic distribution of metazoan microRNAs: insights into evolutionary complexity and constraint. J Exp Zoolog B Mol Dev Evol **306**:575-588.

Sempere, L. F., P. Martinez, C. Cole, J. Baguna, and K. J. Peterson. 2007. Phylogenetic distribution of microRNAs supports the basal position of acoel flatworms and the polyphyly of Platyhelminthes. Evolution & Development **9**:409-415.

Shimodaira, H., and M. Hasegawa. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17**:1246-1247.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688-2690.

Sterrer, W., M. Mainitz, and R. Rieger. 1985. Gnathostomulida: enigmatic as ever. Clarendon Press.

Struck, T., N. Schult, T. Kusen, E. Hickman, C. Bleidorn, D. McHugh, and K. Halanych. 2007. Annelid phylogeny

and the status of Sipuncula and Echiura. BMC Evolutionary Biology **7**:57.

Telford, M. J., and P. W. Holland. 1993. The phylogenetic affinities of the chaetognaths: a molecular analysis. Mol Biol Evol **10**:660-676.

Todaro, M. A., M. J. Telford, A. E. Lockyer, and D. T. J. Littlewood. 2006. Interrelationships of the Gastrotricha and their place among the Metazoa inferred from 18S rRNA genes. Zoologica Scripta **35**:251-259.

Turbeville, J. M., K. G. Field, and R. A. Raff. 1992. Phylogenetic position of phylum Nemertini, inferred from 18S rRNA sequences: molecular data as a test of morphological character homology. Mol Biol Evol **9**:235-249.

Vinther, J., and C. Nielsen. 2005. The Early Cambrian Halkieria is a mollusc. Zoologica Scripta **34**:81-89.

Wallberg, A., M. Curini-Galletti, A. Ahmadzadeh, and U. Jondelius. 2007. Dismissal of Acoelomorpha: Acoela and Nemertodermatida are separate early bilaterian clades. Zoologica Scripta **36**:509-523.

Winnepenninckx, B., T. Backeljau, L. Y. Mackey, J. M. Brooks, R. Dewachter, S. Kumar, and J. R. Garey. 1995. 18s Ribosomal-Rna Data Indicate That Aschelminthes Are Polyphyletic in Origin and Consist of at Least 3 Distinct Clades. Molecular Biology and Evolution **12**:1132-1137.

Xia, X., and Z. Xie. 2001. DAMBE: Data analysis in molecular biology and evolution. Journal of Heredity **92**:371-373.

Zrzavý, J. 2001. The interrelationships of metazoan parasites: a review of phylum-and higher-level hypotheses from recent morphological and molecular phylogenetic analyses. Folia Parasitolofica **48**:81-103.

Zrzavy, J., S. Mihulka, P. Kepka, A. Bezdek, and D. Tietz. 1998. Phylogeny of the Metazoa Based on Morphological and 18S Ribosomal DNA Evidence. Cladistics **14**:249-285.

Zrzavý, J., S. Mihulka, P. Kepka, A. Bezdek, and D. Tietz. 1998. Phylogeny of the Metazoa Based on Morphological and 18S Ribosomal DNA Evidence. Cladistics **14**:249-285.

Supplementary Information

Supplementary Table 1. Species collected and providers.

Group	Species	Provider/Collection	SMART
Diploblastica			
Porifera			
	Unclassified Sycon	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Chondrosia reniformis	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
Cnidaria			
	Bunodactis verrucosa	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Metridium senile	Collected at Kristineberg Marine Station with the help of Dr. Ul.Jondelius (Sweden, summer 2003)	
	Parazoanthus	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Alcionum acaule	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Eunicella	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Anemonia sulcata	Dr. Carles Ribera (Universitat de Barcelona, Spain)	Yes
	Podocoryne carnea	Dr. Volker Schmidt (University of Bielefeld, Germany)	
Myxozoa	Unclassified Myxozoa	Dr. Humbert Salvador (Universitat de Barcelona, Spain)	Yes
Placozoa	*Trichoplax adhaerens*	Dr. Bernd Schierwater (Institute of Animal Ecology & Cell Biology, Germany)	
Ctenophora			
	Beroe sp	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	
	Bolinopsis sp	Dr. Andreas Wallberg (Uppsala University, Sweden)	Yes
Acoelomorpha			
Acoela			
	Paratomela rubra	Collected by Jordi Paps at Sitges, Spain (autumn 2001)	Yes
	Convoluta roscofensis	Dr. Emili Saló (Universitat de Barcelona, Spain)	

Group	Species	Provider/Collection	SMART
	Convoluta convoluta	Dr. Adam McCoy (Harvard University, USA)	Yes
	Conviluta pulchra	Dr. Pere Martínez (Universitat de Barcelona, Spain)	
	Parafanostoma triangulifera	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	
Nemertodermatida	*Nemertodermatida westbladi*	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, autumn 2001)	Yes
Lophotrochozoa			
Platyhelminthes			
	Stenostomum bicaudatum	Dr. Cristina Damborenea (Museo de La Plata, Argentina)	Yes
	Stenostomum tenuicauda	Dr. Cristina Damborenea (Museo de La Plata, Argentina)	
	Stenostomum leucops aquariorum	Dr. Maria Reuter (Institute of Biology, Finland)	
	Thysanozoon sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Discocelis sp	Collected by Jordi Paps at Cubelles, Spain (spring 2002)	
	Leptoplana	Collected by Jordi Paps at Cubelles, Spain (spring 2002)	
	Procerodes plebeja	Collected by Jordi Paps at Cubelles, Spain (spring 2002)	
	Dendrocoelum lacteum	Dr. Emili Saló (Universitat de Barcelona, Spain)	
	Schmidtea mediterranea	Dr. Emili Saló (Universitat de Barcelona, Spain)	
	Dugesia sicula	Collected by Jordi Paps at Sot del Ferrer, Spain (autumn 2002)	
	Dugesia japónica	Dr. Emili Saló (Universitat de Barcelona, Spain)	
	Girardia tigrina	Dr. Emili Saló (Universitat de Barcelona, Spain)	
Nemertea			
	Lineus sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Tubulanus sp	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	Yes
Mollusca			
	Epimenia babai	Dr. Gonzalo Giribet and Dr. Akiko Okusu (Harvard University, USA)	Yes

Group	Species	Provider/Collection	SMART
	Lepidochitona sp	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2002)	
	Donax trunculus (tallarina)	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Cerastoderma edule	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Haliotis sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Antalis entalis	MEDIT-ES Campaign (east coast of Spain, summer 2003)	
	Bolinus brandarus	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Deroceras sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Ceritium sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Eledone cirrhosa	MEDIT-ES Campaign (east coast of Spain, summer 2003)	
	Illex coindetii	MEDIT-ES Campaign (east coast of Spain, summer 2003)	
Sipuncula			
	Phascolosoma granulatum	Dr. Carles Ribera (Universitat de Barcelona, Spain)	Yes
Annelida			
	Protodrilus rubropharyngeus	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	
	Trilobodrilus sp	Collected at Sylt Marine Station with the help of Dr. Schmidt-Rhaesa and Dr. Liesenjohann (Germany, autumn 2003)	
	Euplinia sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Unclassified Polichaeta	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Unclassfied Oligochaeta	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Unclassified Erpobdellidae	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
Echiura			
	Echiurus echiurus	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	
	Bonellia viridis	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	Yes

Group	Species	Provider/Collection	SMART
Gnathostomulida			
	Gnathostomula peregrina	Dr. Martin Vinter Sorensen (Harvard University, USA)	
	Haplognathia ruber.?	Dr. Martin Vinter Sorensen (Harvard University, USA)	
Cycliophora			
	Symbion pandora	Dr. Matthias Obst (Harvard University, USA)	
Rotifera			
	Brachionus plicatilis	Dr. Maria Rosa Miracle (Universitat de Valencia, Spain)	
	Brachionus MVS	Dr. Martin Vinter Sorensen (Harvard University, USA)	
	Philodina 1	Carolina Biological Supplies	
	Philodina 2	Carolina Biological Supplies	
	Philodina MVS	Dr. Martin Vinter Sorensen (Harvard University, USA)	
	Macrotrachela quadricornifera	Dr. Claudia Ricci (State University of Milan, Italy)	
Bryozoa			
	Miriapora truncata	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Celaria sp 1 & 2	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	Yes
	Turbicellepora sp 1 & 2	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	Yes
	Alcyonidium sp	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	
	Securiflustra sp	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	
Brachiopoda			
	Unclassified Brachipoda	Dr. Jordi García Fernández (Universitat de Barcelona, Spain)	
	Unclassified Brachipoda	Dr. Ricard Albalat (Universitat de Barcelona, Spain)	

Group	Species	Provider/Collection	SMART
	Unclassified Brachipoda	Dr. Gonzalo Giribet (Harvard University, USA)	
Phoronida			
	Phoronis hipporepia	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
Entoprocta			
	Barentsia matsushimana	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2002)	Yes
Gastrotricha			
	Unclassified Gastrotricha	Dr. Humbert Salvador (University of Barcelona, Spain)	
	Neodays sp	Collected by Jordi Paps at Sitges, Spain (summer 2005)	
	Holichaetonotus aculifer	Dr. Antonio Todaro (Università di Modena & Reggio Emilia, Italy)	
	Tetranchyroderma papii	Dr. Antonio Todaro (Università di Modena & Reggio Emilia, Italy)	
	Mesodasys adenotubulatum	Dr. Antonio Todaro (Università di Modena & Reggio Emilia, Italy)	
Ecdysozoa			
Onychophora			
	Uncalssified Onychophora	Dr. Sean Carrol (R.M. Bock Laboratories, USA)	
Nematomorpha			
	Unclassified Nematomorpha	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Unclassified Nematomorpha	Miquel Vila (Universitat de Barcelona, Spain)	
Kinorhyncha			
	Pycnophyes kielensis	Collected at Sylt Marine Station with the help of Dr. Schmidt-Rhaesa and Dr. Liesenjohann (Germany, autumn 2003)	Yes
	Pycnophyes sp.	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2002)	

Group	Species	Provider/Collection	SMART
Priapulida			
	Priapulus caudatus	Dr. Sean Carrol (R.M. Bock Laboratories, USA)	
	Priapulus caudatus	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	
Deuterostomia			
Echinodermata			
	Unclassified Crinoidea (from	Dr. Creu Palacín (Universitat de Barcelona, Spain)	Yes
	Amphiura filiformis	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	
	Ofiuroderma sp	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	
	Ofiuroderma sp	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Ofiura	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Unclassified Holoturia	Collected at Roses with the help of Dr. Palacin, Dr. Turon and Dr Ballesteros (Spain, autumn 2003)	
	Unclassified Holoturia Apoda	Dr. Creu Palacín (Universitat de Barcelona, Spain)	
	Strongylocentrotus bachiensis	Dr. Gonzalo Giribet and Dr. Stephanie Huff (Harvard University, USA)	
	Paracentrotus lividus	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
Hemichordata			
	Unclassified Hemichordata	Dr. Jordi García Fernández (Universitat de Barcelona, Spain)	
	Unclassified Hemichordata	Dr. Ricard Albalat (Universitat de Barcelona, Spain)	
	Hemicordat	Dr. Ricard Albalat (Universitat de Barcelona, Spain)	
	Enteropneust	Dr. Gonzalo Giribet (Harvard University, USA)	
Urochordata			
	Halocinthia papillosa	Dr. Carles Ribera (Universitat de Barcelona, Spain)	
	Ciona intestinalis	Dr. Ricard Albalat (Universitat de Barcelona, Spain)	Yes

Group	Species	Provider/Collection	SMART
Cephalochordata	*Branchiostoma floridae*	Dr. Jordi García Fernández (Universitat de Barcelona, Spain)	
	Branchiostoma lanceolatum	Dr. Ricard Albalat (Universitat de Barcelona, Spain)	
Chaetognatha			
	Sagitta elegans	Collected at Kristineberg Marine Station with the help of Dr. Ulf Jondelius (Sweden, summer 2003)	Yes
	Flaccisagitta enflata	Dr. Mark Martindale (Kewalo University, USA)	

Supplementary Table 2. Primer sequences and their Tm.

Histone H3 (Edgecombe et al, 2000)

H3aF	ATGGCTCGTACCAAGCAGACVGC	50ºC
H3aR	ATATCCTTRGGCATRATRGTGAC	50ºC

Glyceraldehyde-3-phosphate dehydrogenase (developed in this study)

GAPDH_F2	ATYAAYGGATTTGGYCGYATCGG	52ºC
GAPDH_R2	GTARCCRAACTCRTTRTCRTACC	52ºC
GAPDH_F1	ATTAACGGATTTGGYCGYATCGG	54ºC
GAPDH_R1	GTARCCRAACTCGTTGTCGTACC	54ºC

Myosin heavy chain type II (Ruiz-Trillo et al, 2002)

Mio6 (R)	CCYTCMARYACACCRTTRCA	53ºC
Mio7 (F)	TGYATCAAYTWYACYAAYGAG	53ºC

Tropomyosin (developed in this study)

TropoF	ATYRAGAAGAARATGNBKGCVATG	53ºC
TropoR	GTHYGRTCCARTTGNYCACT	53ºC

Intermediate filaments (developed in this study)

FilF1	TACATCGAGAAGGTGCGTTTCCTGG	48ºC
FilF3	TACATCGAGAAGGTGCGTTTCCTGG	50ºC
FilR1	CCTCACCCTCCAGCAGCTTTCTGTA	48ºC
FilR3	CYTCNCCYTCCAGCAGYTTYCTGTA	50ºC

Sodium-potassium ATPase alpha-subunit (Anderson et al, 2004)

fATPa	ATGACNGTNGCNCATATGTGGT	52ºC
rATPa	ATNGGGTGGTCNCCNGTNACCAT	52ºC
fATPb	GTNATGAAGGGNGCNCCNGA	52ºC
rATPb	CCCATNGCNACNCCNATGTCNGCTTT	52ºC
fATPc	ATGGTNACNGGNGATCATCCNAT	52ºC
rATPc	ATNGCNGGNACCATGTCNGTNCC	52ºC

Aldolase (Peterson et al, 2004)

AldolasaF	GGGAARGGNATHYTNGCNGC	50ºC
AldolasaR	GGGGTNACCATRTTNGGYTT	50ºC

ATP synthase beta-chain (Peterson et al, 2004)

ATPB-F	GTNGAYGTNCARTTYGAYGA	50ºC
ATPB-R	NCCNACCATRTARAANGC	50ºC

Catalase (Peterson et al, 2004)

CatalasaF	GAYGARATGDSNCAYTTYGAYMG	50ºC
CatalasaR	CCNARNCKRTGNMDRTGNGTRTC	50ºC

Methionine adenosyltransferase (Peterson et al., 2004)

MAT-F	GGNGARGGNCAYCCNGAYAA	50ºC
MAT-R	CCNGGNCKNARRTCRAARTT	50ºC

Phosphofructokinase (Peterson et al., 2004)

PFK-F	GGNGGNGAYGCNCARGGNATGAA	50ºC
PFK-R	GGNCKNARNARCCACCAYTG	50ºC

Supplementary Table 3. Summary of the sequence matrices used in the study.

For each phylum, the number of sequences present (first column), the average number of genes/representative (middle column) and the percentage of missing data (last column) are indicated.

Group (number of OTUs)	Present sequences	Genes/OTU	Average missing data
Cnidaria (6)	52	8,6	36%
"Acoelomorpha"(7)	38	5,4	52%
Platyhelminthes (7)	45	6,4	48%
Annelida (5)	43	8,6	30%
Pogonophora (1)	4	4	53%
Mollusca (6)	44	7,3	42%
Nemertea (3)	31	10,3	15%
Echiura (1)	6	6	43%
Brachiopoda (3)	24	8	30%
Phoronida (1)	11	11	9%
Rotifera (2)	13	6,5	41%
Acanthocephala (1)	3	3	63%
Entoprocta (1)	4	4	54%
Bryozoa (3)	22	7,3	25%
Gnathostomulida (2)	6	3	73%
Gastrotricha (2)	7	3,5	76%
Priapulida (1)	12	12	10%
Kinorhyncha (1)	4	4	49%
Nematoda (2)	18	9	24%
Nematomorpha (2)	7	3,5	61%
Onychophora (1)	3	3	57%
Arthropoda (13)	82	6,3	43%
Echinodermata (7)	52	7,4	37%
Hemichordata (2)	21	10,5	17%
Urochordata (1)	8	8	35%
Cephalochordata (1)	6	6	45%
Vertebrata (4)	48	12	10%
Chaetognatha(3)	19	6,3	51%
Xenoturbella (1)	10	10	49%
Totals	**643**	**5,9**	**40%**