Genome Biology and Evolution Advance Access published May 30, 2010 doi:10.1093/gbe/evq030

Genome Biology and Evolution

SMBE

Ecdysozoan mitogenomics: Evidence for a common origin of the legged invertebrates, the Panarthropoda.

Rota-Stabelli Omar[1,2*], Ehsan Kayal[3], Dianne Gleeson[4], Jennifer Daub[5], Jeffrey Boore[6], Max Telford[1], Davide Pisani[2], Mark Blaxter[5] and Dennis Lavrov[3*]

1 Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London WC1E 6BT, UK

2 Department of Biology, The National University of Ireland, Maynooth, Maynooth, Co. Kildare, Ireland

3 Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA

4 EcoGene, Landcare Research New Zealand Ltd., 231 Morrin Road, St Johns, Auckland NZ

5 Institute of Evolutionary Biology, The University of Edinburgh, Ashworth Laboratories, Edinburgh EH9 3JT, UK

6 Genome Project Solutions, 1024 Promenade Street, Hercules, CA 94547, USA

* Corresponding authors

Contacts: Omar Rota Stabelli: omar.rota-stabelli@nuim.ie, omar42@gmail.com. Dennis Lavrov: dlavrov@iastate.edu

Abstract

Ecdysozoa is the recently recognized clade of molting animals that comprises the vast majority of extant animal species and the most important invertebrate model organisms – the fruitfly and the nematode worm. Evolutionary relationships within the ecdysozoans remain, however, unresolved, impairing the correct interpretation of comparative genomic studies. In particular, the affinities of the three Panarthropoda phyla (Arthropoda, Onychophora, and Tardigrada) and the position of Myriapoda within Arthropoda (Mandibulata vs Myriochelata hypothesis) are among the most contentious issues in animal phylogenetics.

To elucidate these relationships, we have determined and analyzed complete mitochondrial genome sequences of two Tardigrada, *Hypsibius dujardini* and *Thulinia* sp. (the first genomes to date for this phylum), one Priapulida, *Halicryptus spigulosus*, and two Onychophora, *Peripatoides* sp. and *Epiperipatus biolleyi*, and a partial mitochondrial genome sequence of the Onychophora *Euperipatoides kanagrensis*. Tardigrada mitochondrial genomes resemble those of the arthropods in term of the gene order and strand asymmetry, while Onychophora genomes are characterized by numerous gene order rearrangements and strand asymmetry variations. In addition, Onychophora genomes are extremely enriched in A and T nucleotides, while Priapulida and Tardigrada are more balanced.

Phylogenetic analyses based on concatenated amino-acid coding sequences support a monophyletic origin of the Ecdysozoa and the position of Priapulida as the sister group of a monophyletic Panarthropoda (Tardigrada plus Onychophora plus Arthropoda). The position of Tardigrada is more problematic, most likely because of long branch attraction (LBA). However, experiments designed to reduce LBA suggest that the most likely placement of Tardigrada is as a sister group of Onychophora. The same analyses also recover monophyly of traditionally recognized arthropod lineages such as Arachnida and of the highly debated clade Mandibulata.

Introduction

In spite of an ongoing debate concerning their utility in phylogenetics (Curole and Kocher 1999, Delsuc et al. 2003, Cameron et al. 2004), mitogenomic studies have proven to be informative and insightful for phylogenetic studies (e.g. Boore et al. 1998, Lavrov and Lang 2005). This can be explained by conceptual advantages such as the conserved gene set, (almost) unambiguous orthology of genes, and presence of rare genomic changes, including gene rearrangement and changes in the genetic code, as well as some historical and methodological advantages such as the availability of primers for amplifying specific genes from many lineages and the relative ease of generating new data (Fendt et al. 2009, Boore, Macey and Medina 2005). On the other hand, phylogenies based on mitochondrial sequences are well known to be affected by a variety of reconstruction artifacts which may be responsible for dilution of the true phylogenetic signal and generation of homoplasies.

One of the main problems of mitogenomics is thought to be the lineage-specific compositional heterogeneity, which also influences the amino acid content of the encoded proteins (Foster, Jermiin and Hickey 1997; Gibson et al 2005). For example, some ecdysozoan lineages, such as some Arthropoda and Nematoda, have mitochondrial genomes enriched for A+T nucleotides and, in the absence of strong purifying selection, the corresponding proteins are enriched for amino acids encoded by A+T rich codons (Foster et al. 1997; Saccone et al. 2002). Another type of compositional heterogeneity, measured by GC and AT skews between the two strands of DNA (Perna and Kocher 1995), reflects a directional mutational bias driven by the asymmetric nature of replication of the mitochondrial genome and results in opposite compositional biases in genes with opposite transcriptional polarities (Saccone et al. 1999; Lavrov et al. 2000, Hassanin et al. 2005; Hassanin 2006, Jones et al. 2007). Gene inversions that cause a gene to change its orientation relative to the replication origin will result in a rapid compositional change as the sequence evolves from its ancestral nucleotide skews towards the ones driven by its new position (Helfenbein, Brown and Boore 2001). It has been shown that heterogeneities in both A+T proportion and AT, GC skew can cause erroneous results in phylogenetic inference (Gibson et al. 2005, Jones et al. 2007, Masta, Longhorne and Boore 2009).

Compositional heterogeneity is only one of the factors affecting mitochondrial phylogenies. Accelerated substitution rates may also play a role in masking and eroding phylogenetic signal through unrecognized homoplasy and lead to increased susceptibility to systematic biases such as long branch attraction (LBA, Felsenstein 1978, Brinkmann et al. 2006). Because of the variety of possibly confounding biases that could affect mitochondrial genomes concurrently, strong outgroup-effects should be expected and have been observed (Cameron et al 2004, Rota Stabelli and Telford 2008), with different outgroups suggesting alternative, equally well-supported, rooting positions for ingroup taxa. One approach to deal with these problems is to improve the standard, general time reversible (GTR) models of mitochondrial sequence evolution both at the nucleotide (Hassanin et al. 2005) and amino-acid level (Abascal et al. 2007, Rota Stabelli, Yang and Telford 2009). More sophisticated evolutionary models such as the heterogeneous CAT model, which accounts for among site heterogeneity (Lartillot and Philippe 2004) and the derived CAT-BP model, which also accounts for lineage-specific compositional heterogeneities (Blanquart and Lartillot 2008), can lessen the effects of compositional biases. Another approach to reduce bias is to increase taxonomic sampling. Sampling more taxa, particularly close to weakly supported nodes, can break long branches, allowing for a better elucidation of homoplastic similarities resulting from multiple substitutions and thus reducing the likelihood of the LBA artifact. Finally, site-stripping approaches (Brinkmann and Philippe 1999; Pisani 2004; Sperling et al. 2009) can be used to eliminate rapidly evolving sites and limit LBA artifacts, since the most rapidly evolving sites are expected to be the most heterogeneous in composition.

Panarthropoda

Our knowledge of metazoan evolution has changed dramatically since the seminal work of Aguinaldo et al. (1997), which first formally proposed the Ecdysozoa, a group of molting organisms that includes Arthropoda (eg: insects), Tardigrada (water bears), Onychophora (velvet worms), Nematoda (round worms), Nematomorpha (horsehair worms), Priapulida (penis worms), Kinorhyncha (mud dragons) and Loricifera. While a monophyletic origin of the Ecdysozoa is now widely accepted (reviewed in Telford et al. 2008), the relationships among the eight extant ecdysozoan phyla, in particular the position of Tardigrada, are still vigorously

debated. Although there is a strong support from morphological and developmental gene expression data for the monophyly of Panarthropoda, a group characterised by segmental, paired, locomotory appendages, comprising Arthropoda, Onychophora and Tardigrada (Nielsen 2001, Telford et al. 2008), these data are ambiguous about the placement of the Tardigrada and the Onychophora within Panarthropoda (Peterson and Eernisse 2001, Nielsen 2001, Mayer and Whitington 2009b). Furthermore, monophyly of Panarthropoda has found little molecular support. While arthropod affinity of the Onychophora is strongly supported by expressed sequence tag (EST)-derived phylogenomic data sets (Dunn et al. 2008, Roeding et al. 2009; Hejnol et al. 2009), mitochondrial data from the complete mitochondrial genomes of one Onychophoran placed it as the sister group of Arthropoda plus Priapulida (Podsiadlowski et al. 2008). The position of the Tardigrada is equally unclear. Ribosomal RNA sequences support a group of Tardigrada plus Onychophora as a sister lineage to the Arthropoda (Mallat and Giribet 2006), while EST data challenged the Panarthropoda hypothesis, grouping Tardigrada with Nematoda (Lartillot and Philippe 2008, Roeding et al. 2009, Hejnol et al. 2009). In these analyses, Tardigrada and Nematoda are characterized by long branches, suggesting that this clade could represent a phylogenetic artifact. This possibility is reinforced by the analyses of Dunn et al. (2008), also based on EST data, which recover a monophyletic Panarthropoda, suggesting that the placement of Tardigrada may be model-dependent.

Arthropoda

The monophyly of Crustacea plus Hexapoda (Tetraconata or Pancrustacea) within Arthropoda is now well accepted (Friedrich and Tautz 1995; Boore et al. 1998, Dohle 2001, Pisani 2009). However, the phylogenetic relationships among Pancrustacea, Myriapoda (millipedes, centipedes, symphylans) and Chelicerata (arachnids, ticks and their allies) are hotly debated. Many morphological and palaeontological studies group Myriapoda with Hexapoda and Crustacea in a clade called Mandibulata (Scholz, Mittmann, and Gerberding 1998, Harzsch, Muller and Wolf 2005, but see Mayer and Whitington 2009a). By contrast, different types of molecular data (mitochondrial, ribosomal, nuclear protein coding genes, and EST data sets) have supported a sister group relationship between Myriapoda and Chelicerata, which were placed together in Myriochelata or Paradoxopoda (Friedrich and Tautz 1995; Pisani et al. 2004,

Podsiadlowski et al. 2008, Mallatt et al 2006, Dunn et al. 2008, Hejnol et al. 2009, Roeding et al. 2009). Conversely, two recent analyses based on mixed markers and 62 nuclear genes found, convincing support for Mandibulata (Bourlat et al. 2008, Regier et al. 2010). Support for either Mandibulata or Myriochelata may depend on the outgroup used (Rota-Stabelli and Telford 2008), exclusion of sites (Pisani 2004) and/or method of phylogenetic inference (Regier et al. 2008), suggesting that signal is weak and that some phylogenetic conclusions may be prone to systematic errors.

Synopsis

To further clarify relationships within Ecdysozoa, shed light on the evolution of their mitochondrial genomes, and fill the gap in taxonomic representation that currently exists, we have sequenced and analyzed the complete mitochondrial genomes of five species: two tardigrades, *Hypsibius dujardini* and *Thulinia* sp. (the first tardigrade mitochondrial genomes to be sequenced); one priapulid, *Halicryptus spigulosus;* and two onychophorans, *Peripatoides* sp. and *Epiperipatus biolleyi*. We also determined a partial mitochondrial genome sequence from a third onychophoran *Euperipatoides kanagrensis*. Here, we briefly describe compositional and genomic characteristics of the mitochondrial genomes of Ecdysozoa, particularly focusing on these newly studied species. We show that Tardigrada species have an "arthropod like" mitochondrial genome and that the two priapulids share the same inverted fragment with an unexpected difference in GC skew. Finally, Onychophora mitochondrial genomes are highly divergent and contain several gene rearrangements.

We carried out phylogenetic analyses to understand the relationships within Panarthropoda. Results strongly support a sister relationship between the Onychophora and the Arthropoda, while the position of Tardigrada is more problematic. However, analyses performed using the CAT model (which is more robust to LBA) support the grouping of Onychophora and Tardigrada within a monophyletic Panarthropoda. This hypothesis is reinforced by sequential removal of rapidly evolving lineages and by the exploration of phylogenetic signal using partitions of sites with different evolutionary rates. We also revisit the relationships within Arthropoda, providing insight into the relationships among the major arthropod subphyla

(Chelicerata, Myriapoda and Crustacea and Hexapoda), as well as the relationships within Chelicerata.

Materials and Methods

Genome sequencing and annotation

The complete mitochondrial genomes of the onychophoran *Epiperipatus biolleyi* and *Peripatoides* sp., the tardigrades *Thulinia* sp and *Hypsibius dujardini* and the priapulid *Halicryptus spinulosus* were amplified and sequenced as described in Lavrov et al. (2000). Partial sequences encompassing five coding genes were identified in *Euperipatoides kanagrensis* using EST data. Open reading frames in the newly sequenced genomes were annotated based on comparisons with protein sequences from closely related species. In addition, the mitochondrial genome from *Metaperipatus inae* (Podsiadlowski et al. unpublished GenBank accession EF624055) was re-annotated based on the two other onychophoran mitochondrial genomes. tRNA genes were inferred using the tRNAscan-SE and ARWEN programs (Lowe and Eddy, 1997; Laslett and Canbäck, 2008) and checked manually. tRNA genes not found by the computer programs were identified based on expected anticodon sequences, conserved positions, potential secondary structures, and similarities with sequences from closely related species. If several potential tRNA gene sequences were found, we preferred the one with a more conserved gene order position. The *Epiperipatus biolleyi* mitochondrial genome has been previously independently analysed by Podsiadlowski et al. (2008). Comparison with the genome sequenced by us revealed a 97.9% identity at nucleotide level and 98.8% identity at amino acid level. However, our analysis resulted in a very different annotation of this genome, especially in respect to tRNA genes.

Compositional analyses

For each species of our dataset, we concatenated the 13 mt protein-coding genes to calculate (1) the overall nucleotide G+C % using all three codon positions and (2) the frequency of amino acids encoded by GC rich codons (G+A+R+P%). In addition, we calculated the amino acid composition (again as G+A+R+P%) expected from the nucleotide composition if the nucleotide frequences at all three codon positions are the same (F1x4 codon model). We have generated 10.000 random nucleotides using the same nucleotide composition of the real data and translated

them into amino acids using the appropriate genetic code. We plotted the corresponding values in figure 1 (white squares and dotted line).

We also calculated the GC skew (Perna and Kocher 1995) and the Skew index (Rota Stabelli and Telford 2008) on the concatenated alignment, and for each gene independently, using all three codon positions. To test if the strand asymmetry of genes was at equilibrium, we also calculated GC skew for the 1st+2nd and 3rd codon position separately. GC skew values were plotted for species of interest, using the inferred arthropod ancestral gene order (AAGO, which is the same as *Limulus polyphemus* and the ancestral ecdysozoan; Lavrov et al. 2000) as a reference to order genes on the abscissa of the plots in Figures 2 and 3. The skew index was calculated as an absolute value, and not referenced to a focal species, as in Rota Stabelli and Telford (2008). We summarize some of these statistics in Supplementary Table 1.

Mitochondrial gene and protein alignments

We downloaded nucleotide sequences of the 13 mitochondrial protein-coding genes for 240 metazoan species from the NCBI (http://www.ncbi.nlm.nih.gov) and the OGRe database (http://drake.physics.mcmaster.ca/ogre/compare.shtml). To this initial data set we added complete sequences for the 6 species determined for this study, thus generating a dataset of 245 taxa (246 minus the onychophoran *Epiperipatus biolleyi* previously published by Podsiadlowski et al., 2008). We translated nucleotide sequences into amino acids, using appropriate genetic codes, and aligned the 13 protein sets individually with ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). We back-aligned nucleotide sequences to the amino acid alignment using TranslatorX (downloadable from http://web.mac.com/maxtelford) and assembled a concatenated alignment of the nucleotide sequences of the 13 genes.

To avoid artifacts due to inadequate outgroup and ingroup selection, we followed the decision making strategy of Rota Stabelli and Telford (2008), generating a table (not shown, but available upon request) containing statistics for each of the 245 ingroup and outgroup species sampled. We thus sampled a set of Ecdysozoa and outgroups aiming to minimize root-to-tips distances and compositional heterogeneity across the data set. The key estimates used to identify optimal outgroups included, for each possible outgroup: (1) the maximum likelihood (ML) distance to the Ecdysozoa, (2) the G+C content, (3) the content of G+C-rich codon encoded

amino acids and (4) the two indicators of GC strand asymmetry: GC skew and the skew index, an indicator of gene overall strand asymmetry. For each possible outgroup an ML distance metric was calculated as the average ML distance to *Priapulus caudatus, Limulus polyphemus* and *Tribolium castaneum*, and the skew index was calculated with reference to the average across all Arthropoda with similar skew profile.

From the 245 taxa alignment, we selected a balanced sample of 66 species (listed in Supplementary Table 1) of which 10 were outgroup taxa. The nucleotide and the amino acid alignments were processed independently with Gblocks at default settings, followed by realignment using Muscle (Edgar, 2004; http://www.drive5.com/muscle) and a manual removal of poorly aligned regions, resulting in an amino acid alignment of 2016 residues and a nucleotide alignment of 7482 residues. Because Nematoda are rapidly evolving, their mitochondrial genomes are particularly A+T rich and could generate LBA artifacts, we did not include them in our initial 66 species dataset. However, Nematoda are a key taxon for resolution of the phylogenetic position of Tardigrada. Accordingly, we generated two additional data sets, partially based on our initial 66-taxon data set, with 11 nematodes, including the slowly-evolving enoplean nematodes (Rota Stabelli and Telford 2008). The two data sets that contained Nematoda had a total of 88 taxa and 2016 amino acid residues and 59 taxa and 2946 amino acid residues.

Phylogenetic analyses

We analyzed the 66-taxon nucleotide and amino acid datasets using a variety of evolutionary models and phylogenetic tools. The nucleotide alignment was analyzed under both Bayesian and ML frameworks using MrBayes3.1.2 and RAxML7.0.3 respectively (Huelsenbeck and Ronquist 2001, Stamatakis 2004). In both cases we excluded third codon positions and modeled 1st and 2nd codon partitions separately using two GTR models and a gamma distribution with five categories (Lanave et al 1984). For the RAxML analyses we used the fast ML method and performed a bootstrap analysis (100 replicates). For the nucleotide data we also used the NTE model of Hassanin (2006) with codon positions recoded using the program Recoder (Masta et al. 2009, http://web.pdx.edu/~stul/Software.html) and modeling the first and second codon positions with two distinct GTRs and the 3rd position with a 2-state character model. Two independent runs of one million generations did not converge and supported different tree

topologies. We therefore calculated the consensus tree by sampling trees only from the run associated with the higher mean log-likelihood.

The amino acid dataset was analyzed more extensively, using homogeneous and heterogeneous models of sequence evolution under both Bayesian and ML frameworks. Cross-validation analyses to test the fit of different models to our dataset were carried out with PhyloBayes 2.3 following the protocol described in the manual (Lartillot et al. 2009). We used the MtREV mitochondrial model (Adachi and Hasegawa 1996) as a reference to test the fit of other models: the CAT model (Lartillot and Philippe 2004), the mechanistic GTR model (Lanave et al 1984, Yang, Nielsen and Hasegawa 1998), MtZoa (Rota-Stabelli et al. 2009) and MtArt (Abascal et al. 2007). Using the MtREV model as a reference, results of the crossvalidation were as follow: mtREV versus: mtART = -80.1 (+/- 25.8); GTR = -85.4925 (+/- 25.3); mtZOA = -91.46 (+/- 21.2); CAT = -169.242 (+/- 18.8). Bearing in mind that negative values correspond to a better fit, we chose CAT and mtZoa for further analyses.

Bootstrap ML analyses with 100 replicates were carried out with the fast ML method implemented in RAxML using custom implementations of the MtZoa models and a 4-category gamma distribution. Bayesian analyses were carried out using both MrBayes and PhyloBayes. In both cases we described the among site rate variation with a gamma distribution using 4 categories. We ran two independent tree searches and stopped them after the likelihood of the sampled trees had stabilized and the two runs had satisfactorily converged (standard deviation of split frequencies lower than 0.02 in MrBayes and maxdiff less than 0.2, but in most of the cases less than 0.01, in PhyloBayes). Bayesian analyses under CAT, GTR and MtZoa models were performed with PhyloBayes.

Bayesian analyses were also performed using the CAT-BP model implemented in NH-PhyloBayes (Blanquart and Lartillot 2008) to account for among site-heterogeneity of the replacement process and among-branch heterogeneity of the stationary frequencies (Lartillot and Philippe 2004; Blanquart and Lartillot 2007). For the NH-Phylobayes analyses we set the number of categories to a value ranging from 120 and 140, as learned by using standard CAT analyses. We ran a minimum of two independent runs for each analysis in NH-Phylobayes, but it was impossible to obtain a meaningful convergence even after millions of generations and multiple runs. This can be explained by the large number of taxa in the datasets and the many free

parameters of the model. We therefore sampled trees from each run independently and compared the results of independent runs.

Sequential taxon and site removal.

In order to explore the signal in our data set, and clarify the placement of the Tardigrada, we sequentially removed rapidly evolving species, which show dubious relationship with Tardigrada in the 66-taxon alignments. We sequentially removed the two Pycnogonida (64 taxa data set), the two Symphyla (62 taxa data set) and then the outgroup taxa plus the rapidly evolving Arachnida (46 taxa data set). We inferred phylogenies from these datasets using PhyloBayes and RAxML, modeling the evolutionary process with the CAT and the MtZoa models respectively.

We further explored the signal in sequences by removing classes of rapidly and slowly evolving sites, using the slow-fast approach of Brinkmann and Philippe (1999) as modified in Sperling et al. (2009). It is well known (*e.g.,* Brinkmann and Philippe 2007, Pisani 2004, Sperling et al. 2009) that rapidly evolving sites present a problem for phylogenetic inference as they often contain no genuine phylogenetic signal but contribute to various phylogenetic artifacts. Castoe et al. (2009) recently pointed out that sites that evolve too slowly in mitochondrial coded proteins can also be misleading for phylogenetic analyses, because they are more likely to undergo adaptive convergent evolution in unrelated lineages. We thus decided to use the approach of Sperling et al. (2009) in which sites are partitioned into quartiles and only those from the two internal ones (i.e. those with the most homogeneous rate of substitution) are used for phylogenetic analyses, with the very slow and very fast sites analysed for comparison. The taxa were partitioned into seven monophyletic groups (Echinodermata, Lophotrochozoa, Aranea, Acari, Myriapoda, Hexapoda and Crustacea), and PAUP4b10 (http://paup.csit.fsu.edu/) was used to calculate site-specific parsimony scores. Phylogenetic analyses using CAT and MtZoa were then performed on two data sets: one including the sites from the 2nd and 3rd quartile of the distribution of parsimony scores and, as a matter of comparison, one including the sites from the 1st and 4th quartile. The sites from the internal quartiles are expected to be homogeneous among them, while sites from external quartiles are intrinsically heterogeneous.

Results and Discussion

High degree of compositional heterogeneity

We explored the nucleotide compositional diversity of Ecdysozoa mitochondrial protein coding genes (Figure 1). Compared to outgroups, all the Ecdysozoas are characterised by mitochondrial coding sequences impoverished in G and C. However, the degree of heterogeneity among the main Ecdysozoa groups is remarkable. Sequences in Onychophora are extremely A+T rich, to a degree that is comparable only to those of the well-known compositionally problematic ticks and nematodes (Supplementary Table 1). Priapulida are characterised by a more balanced nucleotide composition, as, to a lesser extent, are tardigrades. Notably, the sequences of the four arthopods "subphyla" are also heterogeneous, with hexapods and chelicerates being A+T rich and myriapods and crustaceans less so. Interestingly, the Chromadorea nematodes are extremely A+T rich, but the relatively slower evolving Enoplea (Rota Stabelli and Telford 2008) have a less extreme nucleotide (and amino acid) composition. Such variability in nucleotide composition is known to result in erroneous phylogenetic reconstructions (see Mooers and Holmes, 2000).

As expected, the overall nucleotide composition and the proportion of the "GARP" amino acids are highly correlated (R^2=0.76). However, for some lineages such as Onychophora, Tardigrada and Hexapoda, the amino acid composition, and its standard deviation, is evidently less biased than the nucleotide one. Furthermore, comparison with the expected amino acid composition of randomized nucleotide sequences (white squares and dotted regression line) suggest that real amino acid sequences are less GARP biased then expected by chance alone. Also, the slope of the regression line of expected amino acid values is steeper than that of the real data. This observation suggests that some constrain is working at the amino acid level, and that the inference of phylogeny based on amino acids may be less prone to compositionally driven systematic errors (but see Delsuc et al., 2003).

Similar gene order and strand asymmetry in Tardigrada and Arthropoda

We compared the gene order (Figure 2) and strand asymmetry properties (Figure 3) of mitochondrial genomes sequenced for this study with those of other Ecdysozoa (Webster et al.

2006, Podsiadlowski et al. 2008) and the putative Arthropod Ancestral Gene Order (AAGO; identical to that of *Limulus polyphemus*) (Lavrov et al., 2000). Since protein coding genes in AAGO (represented by *L. polyphemus*) have two possible transcriptional orientations, they experience different strand asymmetry pressures. In addition, 1st and 2nd positions in the conserved genes of complex IV (*cox1, cox2* and *cox3*; Nardi et al. 2003) are slightly affected by strand bias while the more rapidly evolving genes of complex I (the NADH subunits) are clearly positively or negatively skewed. The 3rd codon position (right part of Figure 3) that is less constrained than the 1st plus 2nd positions (left part of Figure 3), also accumulates nucleotide skews more quickly, and is more likely to be at equilibrium comparing to them.

The mitochondrial gene order of the tardigrade *Thulinia* sp. differs from the AAGO only in the position of *trnI* which is located between *trnL1* and *trnL2* (as also observed in *Hypsibius dujardini)* and has an opposite transcriptional polarity (Figure 2). The *H. dujardini* mitochondrial genome displays several additional rearrangements not present in *Thulinia*. These autapomorphies include the inversion of *trnR*, an interchange of the positions of the *trnT-nad6-cob-trnS2* and the *nad1-trnL2* regions, and transpositions of *nad2* and two clusters of tRNAs (*trnW-trnC-trnY* and *trnK-trnD*). None of the protein coding genes change their transcriptional polarity in these rearrangements, and so their GC skew values resemble the AAGO strand profile (red and orange bars in Figure 3a).

Genome rearrangement and strand asymmetry reversal in Onychophora and Priapulida

The gene order in the onychophoran *Peripatoides* sp. is identical to the AAGO with the exception of an inversion of *trnQ* (Figure 2). Conversely, the two other representatives of Onychophora (Podsiadlowski et al. 2008, Podsiadlowski et al. unpublished GenBank: EF624055) display multiple gene rearrangements, autapomorphic for each species. As a result, the three onychophoran mitochondrial genomes share very few gene boundaries (only *atp6-atp8*, *nad1-trnL2* and *cob-trnS2*). The strand profile in Onychophora differs significantly from *L. polyphemus*, showing a reversal of strand asymmetry in some regions (Figure 3b). The three onychophorans share positive GC-skew signatures for *cox1, cox2, cox3, atp6, nad2* and *nad3* (genes that have negative skew in *L. polyphemus*) suggesting a shared ancestral global reversal of the skew, possibly due to an inversion of the control region. Some genes such as *nad5*, *nad4* and

nad4L and *cob* do not have a conserved strand profile across the three onychophorans, possibly because of a recent shift in the strand environment of these genes. The strong correlation of strand asymmetry at 1st + 2nd positions and at 3rd codon position for all genes in *E. biolleyi* illustrates that this genome is at equilibrium and supports the hypothesis of an ancestral inversion of the control region in the group. By contrast, the skew pattern in *Metaperipatus inae* appears to be out of equilibrium and displays a more complex series of rearrangement events. Additional mitochondrial genomes from related taxa will be very valuable in unraveling this conundrum.

The mitochondrial gene arrangement of the priapulid *Halicryptus spinulosus* is exactly the same as that of the previously published *Priapulus caudatus* (Webster et al. 2006). Both arrangements differ from AAGO by a single inversion of the *trnS1-rns* region (Figure 2). Because the GC skew is considered to result from the replication process, we expected the inverted genes in the priapulids to have an opposite skew to that of *L. polyphemus*. Indeed, the skew values at 1st + 2nd and 3rd positions for these genes in *H. spinulosus* are as expected (Figure 3c). However, in *P. caudatus* the skews at 1st and 2nd position are much reduced in magnitude, while at the 3rd position the skew is opposite to that predicted. One explanation for this discrepancy is that there has been a recent inversion of the control region in *P. caudatus*, and that skew values have not yet reached equilibrium in their new mutational pressure regime.

The lack of unambiguous shared derived (synapomorphic) gene rearrangements among Arthropoda, Tardigrada, Onychophora, and Priapulida means that no resolution can be achieved for their interrelationships using mitochondrial gene order data. This conclusion rejects some previous claims based on mitochondrial gene order data of close relationships between Arthropoda and Tardigrada (Ryu et al. 2007). Since the AAGO has also been inferred as the putative Protostome Ancestral Gene Order (Lavrov and Lang, 2005), no resolution for the relationships between Ecdysozoa and other protostome groups (such as Lophotrochozoa) can be achieved based on this character set. However, the presence of synapomorphies for Tardigrada and Priapulida (see above), as well as additional rearrangements in Tardigrada and Onychophora suggest that mitochondrial gene order data will be informative for phylogenetic studies within these groups.

The problem of Nematoda

The 66 taxon dataset analysed (see below) does not contain any Nematoda, although they are of key importance for resolving the affinities and internal relationships of the Ecdysozoas, in particular of the Tardigrada, which have been linked to Nematoda in phylogenomic studies (Lartillot and Philippe 2008, Dunn et al 2008). However, nematode mitochondrial genomes have high evolutionary rates and previous attempts to use them in phylogenetic reconstruction led to dubious assemblages of Nematoda with other rapidly evolving lineages (Mwinyi et al 2009, Podsiadlowski, Braband and Mayer 2008). We investigated the effect of Nematoda on phylogenetic inference by assembling preliminary datasets that included representatives from this group, in particular sampling slowly evolving enoplean nematodes. In these analyses, Nematoda were associated with rapidly evolving lophotrochozoan outgroups, implying a polyphyletic Ecdysozoa (Supplementary Figures 1a and 1b). These issues appear insurmountable with our current models of sequence evolution and we have therefore excluded Nematoda from our main study dataset.

Phylogenetic analyses suggest an unlikely sea spider affinity of the Tardigrada

Bayesian and ML analyses of 1st and 2nd codon positions performed under the GTR model (Figure 4) supported monophyly of Ecdysozoa, with Priapulida placed as the sister group of Onychophora plus Arthropoda, although with weak support. Arthropoda is paraphyletic in this tree, as Tardigrada are grouped with the rapidly evolving sea spider (Pycnogonida) and Symphyla. Bootstrap support values (BS; in bold in Figure 4) from the ML analyses are low, suggesting that either the phylogenetic signal in this dataset is weak or competing non-phylogenetic signals are present. Furthermore, inspection of branch lengths shows that Tardigrada, Pycnogonida and Symphyla are rapidly evolving lineages, suggesting that their grouping may be the result of the LBA. Mitochondrial genomes of Ecdysozoa are characterised by different patterns of strand asymmetry (Figure 3 and Supplementary Table 1). The NTE recoding strategy has been shown to reduce strand bias artifacts (Hassanin et al 2005, Jones et al. 2007), but analysis of the alignment under NTE recoding yields trees that are largely similar to those recovered using standard GTR models, grouping Tardigrada and Pycnogonida with the Symphyla (see PPs underlined in Figure 4). We conclude that artifacts due to strand asymmetry

are not driving the resolution of the relationships of Tardigrada as their mitochondrial genomes have a pattern of strand asymmetry typical for the majority of Arthropoda sampled (Figure 2a).

The amino acid content of ecdysozoan mitochondrially-encoded protein genes is markedly more homogeneous among different lineages than the nucleotide content (Figure 1), suggesting that structural constraints acting at the protein level may reduce the effects of mutational pressure acting at the nucleotide level. As the homogeneity of the stationary frequencies across the tree is an assumption of the majority of evolutionary models, the amino acid alignment appears to be a better substrate for inference of phylogeny (but see Delsuc et al., 2003). Consequently, we carried out more detailed analyses on a protein dataset of 66 taxa and 2307 amino acid residues. Initially, we performed a cross-validation analysis to test the fit of different models (MtREV, mtArt, mtZoa, GTR, CAT) to the dataset. Results clearly show that the heterogeneous CAT model best fits the 66 taxon dataset. Interestingly, the second best model is MtZoa, which fits the dataset better than the mechanistic GTR, probably as a result of the dataset not containing enough replacement information to satisfactorily estimate all the parameters of the GTR matrix (Rota Stabelli, Yang and Telford 2009). We thus chose the CAT and MtZoa models for further analyses and used the other models for comparative purposes only.

The consensus tree from the Bayesian and ML analyses using the MtZoa model (Figure 5a) resembles the nucleotide tree of Figure 4, with the exception that Tardigrada plus Pycnogonida was nested within paraphyletic Arachnida, and was not in a monophyletic clade with the Symphyla. Analyses performed using MtArt and MtREV resulted in topologies that were very similar to that derived using MtZoa (data not shown).

Tardigrada plus Chelicerata is due to LBA: Support for Panarthropoda using taxon removal and the CAT model

The grouping of Tardigrada, Pycnogonida and Symphyla has no support from morphological data and challenges two commonly accepted notions, monophyly of Chelicerata (supported for example by the presence of chelicerae) and monophyly of Arthropoda (which possesses articulated appendages). A possible LBA artifact is suggested by the extremely accelerated rate of evolution of the mitochondrial genomes of the sampled tardigrades, pycnogonids and Symphyla. Analyses under the NTE model and the exclusion of strand-biased

amino acids recovered the same topology as analysis of the original data set under GTR, suggesting that a Pycnogonida/Symphyla affinity of Tardigrada is not simply due to strand asymmetry-driven bias, but rather to a more general LBA artifact.

To test the possible effect of systematic LBA errors, we sequentially removed taxa from the 66 taxa dataset: the rapidly evolving Pycnogonida, the rapidly evolving Symphyla, and all the outgroups and some Chelicerata with accelerated rates of evolution and/or inversions of strand asymmetry (see Supplementary Table 1). Under the homogeneous MtZoa model, the position of Tardigrada was very unstable through this data reduction scheme. Using the full 66 taxa dataset Tardigrada were sister to the Pycnogonida (Figure 5A,). Considering only the ecdysozoan taxa, this corresponds to a four-taxon tree ((paraphyletic Arthropoda + Tardigrada), (Onychophora, Priapulida)) or ((pA+T),(O,P)). When Pycnogonida were excluded, Tardigrada were sister to Symphyla (i.e. ((pA+T),(O,P)); Figure 5B). Exclusion of Symphyla yielded Tardigrada as sister to all remaining Ecdysozoa and Priapulida as most closely related to Arthropoda (i.e. ((A,P),(O,T)); Figure 5C). Eventually, when the fastest evolving lineages were removed and only slowly evolving Priapulida retained, Tardigrada are weakly recovered as sisters to Onychophora (i.e. (P,(A,(O,T))); Figure 5D). Analyses, using MtREV, MtArt and GTR gave similar results (data not shown).

A clade of Tardigrada and Onychophora is also consistently recovered using the CAT model, whether the rapidly evolving species are included or not (Figure 6). The CAT model has been shown to be quite effective at overcoming the effects of LBA (Bourlat et al. 2009, Lartillot et al. 2007, Lartillot and Philippe 2008) and is the model that best fits our data. Consequently, the CAT topology should be regarded as more likely than that obtained using homogenous models (MtZoa or GTR) and the full set of taxa (respectively Figure 5A and Figure 4). We have also analyzed the dataset using the CAT-BP model (Blanquart and Lartillot 2008), although problems with convergence (see material and methods for more details), prevented us from drawing definite conclusions using this model. The use of the CAT-BP model in the analysis of the full dataset resulted in a sister group relationship between Tardigrada and Pycnogonida, while the same analysis in the absence of the fast evolving Pycnogonida tepidly supported a sister relationship between Tardigrada and Onychophora in accordance with the CAT analyses (supplementary Figure 3).

In conclusion, while the position of Tardigrada is highly unstable, we tentatively support its affiliation with Onychophora, as suggested by the CAT model and the analyses using reduced datasets and MtZoa. The outgroup roots in the Priapulida branch in the four-taxon tree in most analyses, suggesting the tree (outgroup,(P,(A,(O,T)))).

More support for Panarthropoda from the rate-partitioned dataset

Not all sites in an alignment have the same rate of evolution. Rapidly evolving sites accumulate multiple mutations and tend to be saturated and contribute to LBA (see Brinkmann and Philippe 1999; Pisani 2004, Sperling et al. 2009). Recently, Castoe et al. (2009) have shown that very slowly evolving sites (under strong purifying selection) can also be phylogenetically misleading as they may experience parallel adaptive changes in unrelated lineages. Accordingly, following Sperling et al. (2009) we explored the distribution of signal in the alignment by separating the sites with moderate evolutionary rates (i.e. those most likely to represent the most reliable source of phylogenetic signal) from the slowly and rapidly evolving sites (those more likely to convey misleading signal). To identify sites with moderate rates (see methods) we ranked individual sites according to their rate (inferred using the Slow-Fast method; Brinkmann and Philippe 1999). We then used sites in the second and third rate quartiles (moderately evolving) for phylogenetic reconstructions, and sites in the first and fourth quartile (slow and fast evolving sites) for comparison.

Unsurprisingly, the CAT tree built using the collection of rate heterogeneous sites (i.e. the fast and slowly evolving; Figure 7a), supports a tardigrades affinity for the chelicerates. Because these sites are most likely to be misleading (Brinkmann and Philippe 1999; Pisani 2004; Sperling et al. 2009; Castoe et al. 2009), this result confirms that a Chelicerata affinity for the Tardigrada is unlikely to be correct. The same tree also supports a group of onychophorans plus spiders as well as paraphyletic Pancrustacea (insects plus crustaceans), all extremely dubious topologies. This suggests that the signal associated with the set of the heterogeneous sites carries a high amount of non-phylogenetic signal. Furthermore, fast and slow evolving sites may be driven by different pressures (GC% for fast sites, positive selection for the slow sites) making the dataset in conflict internally.

On the other hand, a CAT tree based on the sites with moderate evolutionary rates (Figure 7b) supports monophyly of commonly accepted Pancrustacea, Chelicerata, and Arachnida groups rendered poly- or paraphyletic in the analyses of heterogeneous sites (Figure 7a). These results suggest that signal in moderately evolving sites is reliable. Notably, this partition supports a monophyletic origin of Panarthropoda, the tree (outgroup,(P,(A,(O,T)))). The sister relationship between Tardigrada and Onychophora is weakly supported (at PP of 58), as is the basal position of Priapulida in Ecdysozoa (PP 53). This weaker support could have resulted from the reduced dimensions of the alignment (and the exclusion of some sites conveying genuine phylogenetic signal) and the inability of the reduced dataset to efficiently resolve alternatives in the parameter-rich CAT model.

Tardigrada plus Onychophora?

When data transformations that are known to reduce LBA are implemented - the use of optimal outgroups (Figure 5D), using effective substitution models (CAT in Figure 6) and the exclusion of unreliable sites (Figure 7B) - support for a monophyletic Panarthropoda in which Onychophora and Tardigrada are sister groups emerges. The recent phylogenomic study of Henjol and colleagues (2009) supported a closer relationship between Onychophora and Arthropoda but places Tardigrada as a sister group to Nematoda + Nematomorpha. Similar phylogenetic position of Tardigrada was found in the study by Roeding et al. (2009). In these studies, however, Tardigrada and Nematoda are fast evolving lineages and their clustering may be due to phylogenetic artifacts such as LBA. In our dataset, the grouping of Tardigrada and Onychophora is unlikely due to LBA: mitochondrial sequences in Tardigrada are rapidly evolving while those in Onychophora have a moderate rate of evolution

There are, however, no commonly accepted synapomorphies of a Tardigrada plus Onychophora clade, though morphologists are divided over whether one of the two is the sister group of the Euarthropoda. A tentative character uniting the tardigrades and the onychophorans is their shared possession of non-articulated clawed appendages, as in the Cambrian lobopodian *Aysheaia*, but in contrast with arthropods which have articulated ones (Nielsen 2001). A lack of information from panarthropod stem group (and/or the difficulty to assess their phylogenetic position) prevents from possible polarisation of this character. Tardigrada lacks an ostiate heart

which is shared by Onychophora and Arthropoda, the two latter also sharing segmental leg musculature (Edgcombe 2010). Conversely, evidences from cuticular and developmental structures suggests a sister relationship between Arthropoda and Tardigrada (Nielsen 2001, Mayer and Whitingotn 2009b). It is, however, clear that morphological comparisons are complicated by the extremely reduced and derived nature of Tardigrada and possibility of parallel evolution.

We were not able to elucidate the phylogenetic position of Nematoda, therefore it is possible that it still forms a sister group to Tardigrada within the Panarthropoda. Clearly, our favourite topology (Panarthropoda) is inconsistent with that of published phylogenomic analyses (Tardigrada plus Nematoda with the exclusion of Onychophora): Hejnol et al. 2009; Roeding et al. 2009). However, it is clear that the mutual relationship of Nematoda and Tardigrada remain an open question, exacerbated by the derived nature of nematodes mitochondrial sequences. Complete mitochondrial genome from the nematode-like Nematomorpha may help in shorten the extremely long stem Nematoda branch and polarise possible characters as apomorphies of the nematodes.

An arthropod affinity for the Onychophora

Ecdysozoa is strongly recovered in all our analyses, with the Priapulida as the sister group of remaining ecdysozoans, the Panarthropoda. Regardless of the position of Tardigrada, the majority of our trees support a sister relationship between Onychophora and Arthropoda. The only exception is the tree in Figure 5C, which supports Priapulida as the sister group of the Arthropoda, a topology which can be interpreted as an artifact due to the mutual attraction of Tardigrada and the outgroup, which may have also pulled Onychophora (assuming they are the sister group of Tardigrada, see Figure 5D and 6) toward the base of the tree. This view is reinforced by the analysis of the same dataset of Figure 5C from which the Tardigrada were excluded, which recovers Onychophora as sister group of the Arthropoda (data not shown).

In a previous analysis, Podsiadlowski and colleagues (2008) could not recover the sister relationship between Onychophora and Arthropoda; this is easily explained by a limited taxon sampling in their analyses (eg: only one onychophoran and priapulid, no tardigrades). The addition of new sequences from *Peripatoides* and *Euperipatoides* appear to increase the

informative phylogenetic signal and thus resolution of the Onychophora sister group position. Within Onychophora, the Peripatopsidae (austral Onychophora) are monophyletic, with the Australian species (*E. kanagrensis*) more closely related to the New Zealand species (*Peripatoides* sp.) than to the Chilean species (*M. inae*), likely reflecting ancient Gondwanan distributions.

Relationships within Arthropoda: Mandibulata and monophyletic Arachnida.

A monophyletic origin of Pancrustacea (the clade comprising Hexapoda and Crustacea) is strongly supported in all our analyses. Our reduced pancrustacean taxon sampling (only Malacostraca and Branchiopoda for the crustaceans and no Collembola) prevented us from testing monophyly of hexapods, which has been challenged by previous mitochondrial analyses (Carapelli et al. 2007).

As for the position of Myriapoda, most of our analyses (Figures 4, 5A, 5B and 6) tend to support Myriochelata. However, as rapidly evolving lineages are removed (toward Figure 5), support for Myriochelata decays and in the dataset with all putative rapidly evolving species excluded (characterised by greater homogeneity of the rate of evolution among lineages), Myriapoda are grouped with the Pancrustacea, supporting the Mandibulata hypothesis (PP 0.96 in Figure 5D). Decrease in support for Myriochelata is also observed using the CAT model (Figure 6). Furthermore, when only the sites with moderate rates of evolution are analyzed, the CAT model strongly supports Mandibulata (PP 98 in Figure 7B). These results suggest that signal supporting Myriochelata is found in rapidly evolving sites or is associated with datasets containing rapidly evolving species. In particular the Symphyla, which tends to group with Chelicerata in most of the analyses (for example in Figures 4 and 5), render Myriapoda paraphyletic. Conversely, when sources of systematic error are reduced (excluding rapidly evolving sites and/or rapidly evolving lineages) this dataset lends support to Mandibulata.

Finally, in some of our phylogenies (Figure 4 and 5A), Chelicerata are paraphyletic due to the inclusion of Tardigrada as a sister group to Pycnogonida, an affinity we have above interpreted as LBA. When Pycnogonids are excluded from the analysis (Figure 5B, C and D), Tardigrada are placed in other parts of the tree leaving Chelicerata monophyletic. On the other hand, using the CAT model, Chelicerata is recovered as a monophyletic group with the

Pycnogonida being the sister group to the remaining eu-chelicerates (Figure 6). Unexpectedly, but in accordance with a recent mitochondrial study of chelicerates (Masta et al. 2009), the horseshoe crab *Limulus polyphemus* (Merostomata) is grouped with the harvestman *P. opilio* (Opilionidae) and the camel spiders *Nothopuga* sp. and *E. palpisetulosus* (Solifugae) in most of our analyses, rendering the Arachnida polyphyletic. However, as in the case of the Myriochelata, support for this dubious grouping decays as rapidly evolving lineages are excluded from the alignment. When long branch arachnids are excluded, both CAT (Figure 6) and MtZoa (Figure 5C and D) recover Merostomata as basal Chelicerata, while Opiliones and Solifugae join Acari, in a monophyletic Arachnida.

Conclusions

Given the ancient origin of Ecdysozoa and the high rate of mtDNA evolution in bilaterian animals, one can expect the phylogenetic signal in ecdysozoan mitochondrial genomes to be low. Furthermore, the recovery of this signal is impeded by lineage-specific rate heterogeneities (Supplementary Table 1), nucleotide composition biases (Figure 1) and strand asymmetrical properties (Figure 3). It is clear that the amount of phylogenetic information available for resolution of some nodes is meager in mitochondrial sequences. However, heterogeneous sequence composition did not play a key role in misleading phylogenetic reconstruction in our data set, as analyses designed to reduce the effect of strand bias (NTE model) and amino acid composition (CAT-BP model) do not ameliorate the problems. One explanation for the difficulty observed in robustly placing some lineages of interest is LBA. In particular, our analyses suggest that LBA is most likely responsible for grouping Tardigrada with the rapidly evolving Arthropoda (Figures 4 and 5A).

Here we have shown that experiments designed to reduce LBA recover a group of Tardigrada plus Onychophora as sister to the Arthropoda in agreement with morphological predictions of a common origin of paired walking appendages (and possibly segmentation) in the Panarthropoda. We note that the same experiments also recover monophyly of usually accepted groups such as Arachnida and Mandibulata.

Thus, while the phylogenetic signal in our mitochondrial datasets is limited, preventing us from drawing firm conclusions, the congruence of analyses that are expected to provide more accurate results in the presence of LBA, suggest the following hypotheses for Ecdysozoa: (outgroups, (Priapulida, (Arthropoda, (Tardigrada, Onychophora)))) and Arthropoda: (outgroups, (Chelicerata, (Myriapoda, (Hexapoda, Crustacea)))). The addition of the remaining ecdysozoan phyla, in particular of Nematomorpha, to the mitogenomic dataset may elucidate these relationships with more confidence.

Acknowledgements and Funding

Thanks to Martin Jones for useful comments on the manuscript, Stuart Longhorn for NTE-recoding the dataset and Samuel Blanquart for assistance with NH-PhyloBayes. O R-S was supported by the Marie Curie RTN 'ZOONET'and by Science Foundation Ireland FRP grant (08/RFP/EOB1595).

Figure Legends

Figure 1. Compositional properties of ecdysozoan mitochondrial coding sequences.

The G+C content of 1st and 2nd codon positions in the concatenated alignment is plotted against the percentage of amino acids encoded by G- and C-rich codons (glycine, alanine, arginine and proline [G+A+R+P]). Values are averaged for some major groups, with standard deviations indicated by a color code. All Ecdysozoa are A+T rich as compared to outgroup sequences. Onychophora are extremely A+T rich, while Priapulida and Tardigrada have more balanced nucleotide compositions. Amino acid frequencies are more homogenous within groups than are the corresponding nucleotide frequencies. As a matter of comparison we have plotted the expected amino acid composition (white squares and dotted regression line) for randomized nucleotide sequences. Color code is the same as in other Figures.

Figure 2. Mitochondrial gene order in Arthropoda, Tardigrada, Onychophora and Priapulida.

Mitochondrial gene order comparisons are shown for sampled Onychophora, Priapulida and Tardigrada, and the Arthropod Ancestral Gene Order (AAGO, exemplified by *Limulus polyphemus*). tRNAs are labelled by the one-letter code for their corresponding amino acids. Genes are transcribed from left to right unless underlined. Black arrows indicate inferred genome rearrangements. Red arrows show inferred synapomorphies of each of the two phyla Priapulida and Tardigrada. Multiple tRNA gene rearrangements inferred between *Peripatoides* sp. and the two other onychophoran species have been omitted for clarity.

Figure 3. Strand compositional asymmetry in Priapulida, Tardigrada and Onychophora.

GC skew calculated for 1st plus 2nd (on the left) and 3rd (on the right) codon positions for the 13 protein-coding genes of Tardigrada (A), Onychophora (B) and Priapulida (C). Genes are ordered as in the Ancestral Arthropod Gene Order (AAGO), and for each plot the values for *Limulus polyphemus* are given. Genes are named as following: N2 (*nad2*), C1 (*cox1*), C2 (*cox2*) A6

(*atp6*), C3 (*cox3*), N3 (*nad3*), N5 (*nad5*), N4 (*nad4*), NL (*nad4L*), N6 (*nad6*), CB (*cytb*) and N1 (*nad*1).

Figure 4. Phylogenetic analyses of nucleotide dataset support an unlikely Pycnogonida affinity of the Tardigrada

Consensus tree from the Bayesian analysis of partitioned 1st and 2nd codon positions using two distinct GTR models is shown. Support at nodes (from left to right) are the posterior probability (PP) from the Bayesian analysis (plain text), the bootstrap supports (BS) from the ML analysis using the same model (bold) and the PP from the Bayesian analysis using the NTE recoding and model (underlined). Tardigrada is consistently recovered as closely related to rapidly evolving Pycnogonida and Symphyla. An alternative position for the Tardigrada plus Pycnogonida using ML is shown by the dotted arrow.

Figure 5. The position of Tardigrada is sensitive to the taxa used in analysis under homogenous model.

Consensus trees from the Bayesian analysis of the amino acid dataset using the MtZoa model, with values at nodes being the PP from the Bayesian analysis (plain text) and the BS from the ML analysis (in bold). The original dataset (tree A) was modified by sequential removal of rapidly evolving lineages: Pycnogonida (B), Symphyla (C) and outgroups plus some rapidly evolving chelicerates (D). The position of Tardigrada (in red) changes as the taxon sampling is reduced, suggesting a reiterated LBA artifact. When all rapidly evolving lineages are excluded and only slowly evolving Priapulida (in pink) are used as outgroups (tree D), support for a group of Tardigrada plus Onychophora is recovered. We show a schematic version of the Bayesian trees with some lineages collapsed for clarity. The original Bayesian tree using the full dataset can be inspected in Supplementary Figure 2. An alternative position for the Tardigrada using ML is shown by the dotted arrow. Color code is the same as in other Figures.

Figure 6. Consistent support for Tardigrada plus Onychophora following sequential taxon removal under the CAT model.

Consensus tree from the Bayesian analysis of the amino acid dataset using the heterogeneous CAT model is shown. Rapidly evolving lineages were sequentially removed from the original dataset as in Figure 5. The four analyses resulted in similar topologies and consistently supported a group of Tardigrada plus Onychophora. Values at nodes are PP using (from left to right) the original 66 taxa dataset, and the sequential removal of Pycnogonida (branch square labelled 1), Symphyla (labelled 2) and the outgroups plus rapidly evolving Chelicerata (labelled 3). Where not indicated PP are 1. Angled slashes on branches indicate that branch length has been halved.

Figure 7. Signal decomposition supports Mandibulata and Panarthropoda. Consensus tree from the CAT Bayesian analysis of (A) sites with slow and fast evolutionary rates (corresponding to 1st and 4th quartiles of a slow fast distribution) and (B) the sites with moderate evolutionary rates (corresponding to 2nd and 3rd quartiles). Supports at nodes are posterior probabilities.

References

Abascal, F., D. Posada, and R. Zardoya. 2007. MtArt: a new model of amino acid replacement for Arthropoda. Mol. Biol. Evol. **24**:1-5.

Adachi, J. and M. Hasegawa. 1996. Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. **42**:459-68.

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.

Blanquart, S., and N. Lartillot. 2008. A site- and time-heterogeneous model of amino acid replacement. Mol. Biol. Evol. **25**:842-58.

Boore, J. L., D. V. Lavrov, and W. M. Brown. 1998. Gene translocation links insects and crustaceans. Nature **392**: 667-668.

Boore, J. L., J. R. Macey, and M. Medina. 2005. Sequencing and comparing whole mitochondrial genomes of animals. In Molecular Evolution: Producing the Biochemical Data, Part B (E. A. Zimmer and E. Roalson, eds.). Volume 395 of the Methods in Enzymology series, Elsevier, Burlington, Massachusetts: 311-348

Bourlat S. J, O. Rota-Stabelli, R. Lanfear, and M. J. Telford. 2009. The mitochondrial genome of Xenoturbella is ancestral within the deuterostome. BMC Evol. Biol. **9**:107.

Bourlat, S. J., C. Nielsen, A. D. Economou, and M. J. Telford. 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. Mol. Phylogenet Evol. 49:23-31.

Brinkmann, H., and Phillipe, H. 1999. Archaea sister group of bacteria? Indications from tree reconstruction artifacts in ancient phylogenies. Mol. Biol. Evol. **16**:817-825.

Brinkmann H., and H. Philippe. 2007. The diversity of eukaryotes and the root of the eukaryotic tree. Adv Exp Med Biol. **607**:20-37.

Cameron, S.L., K. B. Miller, C. A. D'Haese, M. F. Whiting, and S. C. Barker. 2004. Mitochondrial genome data alone are not enough to unambiguously resolve the relationships of Entognatha, Insecta and Crustacea sensu lato (Arthropoda). Cladistics **20**:534-557.

Carapelli, A., P. Liò, , F. Nardi, E. Van der Wath, and F. Frati. 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol. Biol. **16**;7 Suppl 2:S8.

Castoe, T. A., A. P. Jason de Koninga, H. Kima, W. Gua, B. P. Noonanb, G. Naylorc, Z. J. Jiangd, C. L. Parkinsond, and D. D. Pollock. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. PNAS. **106**. 8986-8991.

Curole, J. P., and T. D. Kocher. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol Evol. **14**:394-398.

Delsuc, F., M. J. Phillips, and D. Penny. 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?" Science **301**:1482.

Dohle, W. 2001. Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name 'Pancrustacea' for the monophyletic unit Crustacea+Hexapoda. Ann. Soc. Entomol. France. **37**:85–103.

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. 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**:745-9.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research **32**, 1792-97.

Edgecombe, G. E. 2010. Arthropod phylogeny: An overview from the perspectives of morphology, molecular data and the fossil record. Arthropod Structure & Development. doi:10.1016/j.asd.2009.10.002

.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**:401–410.

Fendt L., B. Zimmermann, M. Daniaux, and W.Parson 2009. Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences. BMC Genomics **10**:139.

Foster, P. G., L. S. Jermiin, and D. A. Hickey. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. **44**:282-8.

Foster, P.G., and Hickey, D.A. 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. **48**:284-90.

Friedrich, M., and D. Tautz. 1995. rDNA phylogeny of the major extant arthropod classes and the evolution of myriapods. Nature **376**:165–167.

Gibson, A., V. Gowri-Shankar, P. G. Higgs, and M. Rattray. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol. Biol. Evol. **22**: 251-64.

Harzsch, S., C. H. Muller, and H. Wolf. 2005. From variable to constant cell numbers: cellular characteristics of the arthropod nervous system argue against a sister-group relationship of Chelicerata and "Myriapoda" but favour the Mandibulata concept. Dev. Genes. Evol. **215**:53-68.

Hassanin, A., N. Leger, and J. Deutsch. 2005 Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Systematic Biology **54**:277-298.

Hassanin, A. 2006. Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. Mol. Phylogenet. Evol. **38**:100-16.

Hejnol, A., M. Obst, A. Stamatakis, M. Ott, G.W. Rouse, G.D. Edgecombe, P. Martinez, J. Baguñà, X. Bailly, U. Jondelius, M. Wiens, W.E. Müller, E. Seaver, W.C. Wheeler, M.Q. Martindale, G. Giribet, and C.W. Dunn. 2009. Assessing the root of bilaterian animals with scalable phylogenomic methods.Proc. Biol. Sci. **276**:4261-70.

Helfenbein, K. G., Brown W. M., and Boore J. L.. 2001 The complete mitochondrial genome of a lophophorate, the brachiopod Terebratalia transversa. Molecular Biology and Evolution **18**: 1734-1744.

Huelsenbeck, J. P., and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754-755.

Jones, M., B. Gantenbein, V.Fet, and M. Blaxter. 2007. The effect of model choice on phylogenetic inference using mitochondrial sequence data: Lessons from the scorpions. Mol Phylogenet Evol. **43**:583-95.

Lanave, C., G. Preparata, C.Saccone, and G. Serio 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. **20**:86-93.

Lartillot, N., H. Brinkmann and H. Philippe. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. **8**;7 Suppl 1:S4.

Lartillot, N., T. Lepage, S. Blanquart. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics. **25**:2286-8.

Lartillot, N., and H. Philippe. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria.. Philos Trans R Soc Lond B Biol Sci. **363**:1463-72.

Lartillot, N., and H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol **21**:1095-109.

Laslett, D., and B. Canbäck. 2008. ARWEN, a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics **24**:172-175.

Lavrov D. V., J. L. Boore, and W. M. Brown. 2000. The Complete Mitochondrial DNA Sequence of the Horseshoe Crab *Limulus polyphemus*. Mol. Biol. Evol. **17**:813-824.

Lavrov D. V., and B.F. Lang. 2005. Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements Systematic Biology **54**:651-659.

Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**:955-64.

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.

Masta, S. E., S. J. Longhorn, and J. L. Boore. 2009. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. Mol Phylogenet Evol. **50**:117-28

Mayer, G., and P. M. Whitington. 2009a. Neural development in Onychophora (velvet worms) suggests a step-wise evolution of segmentation in the nervous system of Panarthropoda. Dev Biol. **335**:263-75.

Mayer, G., and P. M. Whitington. 2009b. Velvet worm development links myriapods with chelicerates. Proc. Biol. Sci. **276**:3571-9.

Mooers, A. O. and E. C. Holmes. 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. **15**:365-369.

Mwinyi, A., A. Meyer, C. Bleidorn, B. Lieb, T. Bartolomaeus, and L. Podsiadlowski. 2009. Mitochondrial genome sequence and gene order of Sipunculus nudus give additional support for an inclusion of Sipuncula into Annelida. BMC Genomics. **16**;10:27.

Nielsen, C. 2001 Animal evolution. Interrelationships of the living phyla, 2nd ed. Oxford, UK: Oxford University Press.

Perna, N. T, Kocher, T. D. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol. **41**:353–35.

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.

Pisani, D. 2004. Removing Fast Evolving Sites Using Compatibility Methods: An Example from the Arthropoda. Systematic Biology **53**:983-994.

Pisani, D. 2009. Arthropoda. In: The Timetree of Life (S.B. Hedges and S. Kumar eds). Oxford University Press**.**

Pisani, D., L. Poling, and S. B. Hedges. 2004. A Molecular Clock Analysis of Arthropod Evolution and the Colonization of Land by Animals. BMC Biology **2**:1.

Podsiadlowski, L., A. Braband, and G. Mayer. 2008. The Complete Mitochondrial Genome of the Onychophoran Epiperipatus biolleyi Reveals a Unique Transfer RNA Set and Provides Further Support for the Ecdysozoa Hypothesis. Mol. Biol. Evol. **25**:42-51.

Regier, J. C. et al. 2008. Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Syst. Biol. **57**:920-938.

Regier, J. C., J.W. Shultz, A. Zwick, A. Hussey, B. Ball, R. Wetzer, J.W. Martin and C.W. Cunningham. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature **463**, 1079-1083.

Roeding, F., J. Borner, M. Kube, S. Klages, R. Reinhardt, and T. Burmester. 2009. A 454 sequencing approach for large scale phylogenomic analysis of the common emperor scorpion (Pandinus imperator). Mol.Phylogenet Evol. doi:10.1016/j.ympev.2009.08.014

Rota-Stabelli, O., and M. J. Telford. 2008. A multi criterion approach for the selection of optimal outgroups in phylogeny: Recovering some support for Mandibulata over Myriochelata using mitogenomics. Mol Phylogenet Evol **48**:103-11.

Rota-Stabelli, O., Yang, Z. and Telford, M. J. 2009. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. Mol Phylogenet Evol. **52**:268-272.

Ryu, S. H., J. M. Lee, K. H. Jang, E. H. Choi, S. J. Park, C. Y. Chang, W. Kim, and U.W. Hwang. 2007. Partial mitochondrial gene arrangements support a close relationship between Tardigrada and Arthropoda. Mol Cells. **24**:351-7.

Saccone, C., C. De Giorgi, C. Gissi, G. Pesole, and A. Reyes. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene **238**:195-209.

Saccone, C., C. Gissi, A. Reyes, A. Larizza, E. Sbisa, and G. Pesole. 2002. Mitochondrial DNA in metazoa: degree of freedom in a frozen event. Gene **286**: 312.

Scholz, G., B. Mittmann, and M. Gerberding. 1998. The pattern of Distal-less expression in the mouthparts of crustaceans, myriapods and insects: new evidence for a gnathobasic mandible and the common origin of the Mandibulata. Int. J. Dev. Biol. **42**:801-810.

Sperling, E. A., K. J. Peterson, and D. Pisani. 2009. Phylogenetic-signal dissection of nuclear housekeeping genes supports the paraphyly of sponges and the monophyly of Eumetazoa. Mol. Biol. Evol. **26**:2261:2274.

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

Telford, M. J., S. J. Bourlat, A. Economou, D. Papillon, and O. Rota-Stabelli. 2008. The evolution of the Ecdysozoa. Philos Trans R Soc Lond B Biol Sci **363**:1529-37.

Webster, B. L., R. R. Copley, R. A. Jenner, J. A. Mackenzie-Dodds, S. J.Bourlat, O. Rota-Stabelli, D. T. Littlewood, and M. J. Telford. 2006. Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral Ecdysozoan. Evol. Dev. **8**:502-10.

Yang, Z., R. Nielsen, and M. Hasegawa. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. Mol. Biol. Evol. **15**:1600-11.

Fig 1

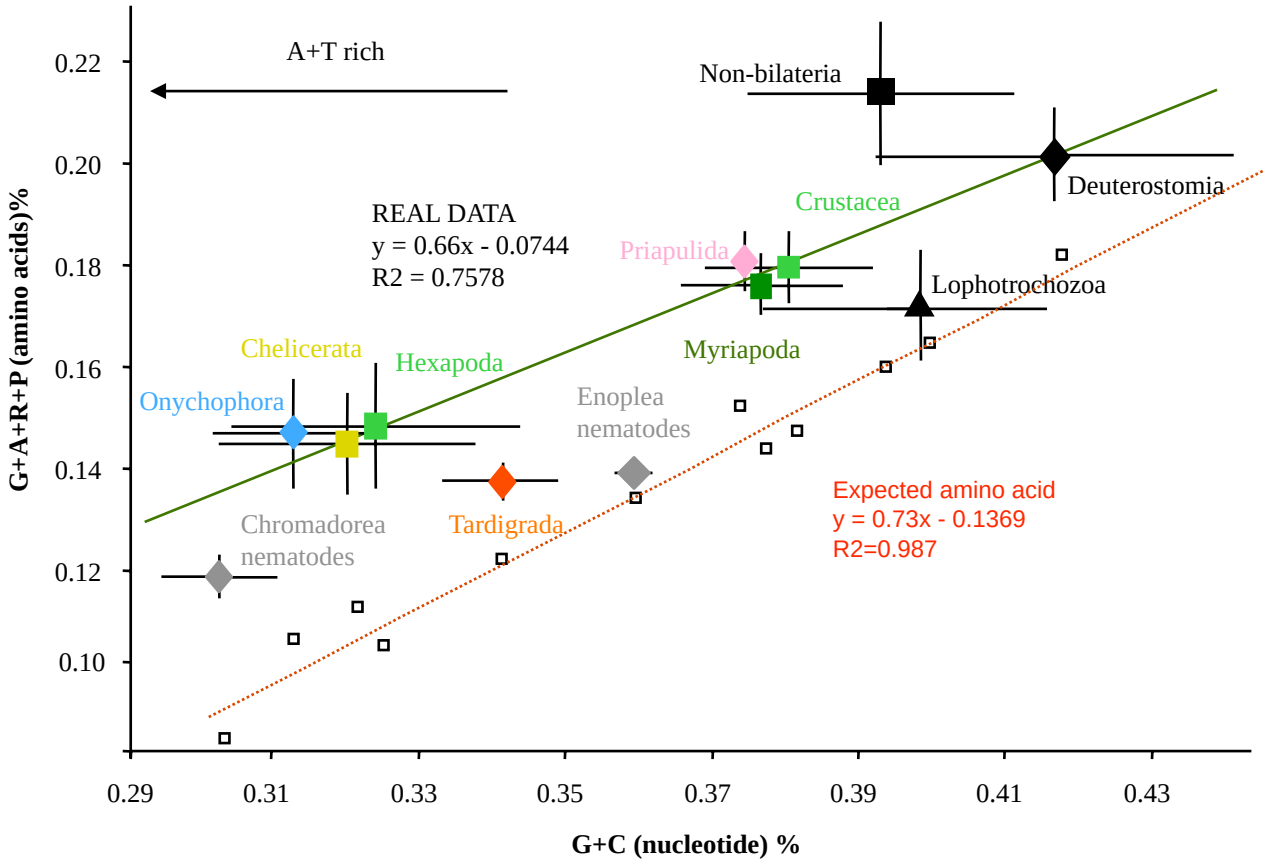


Fig 2

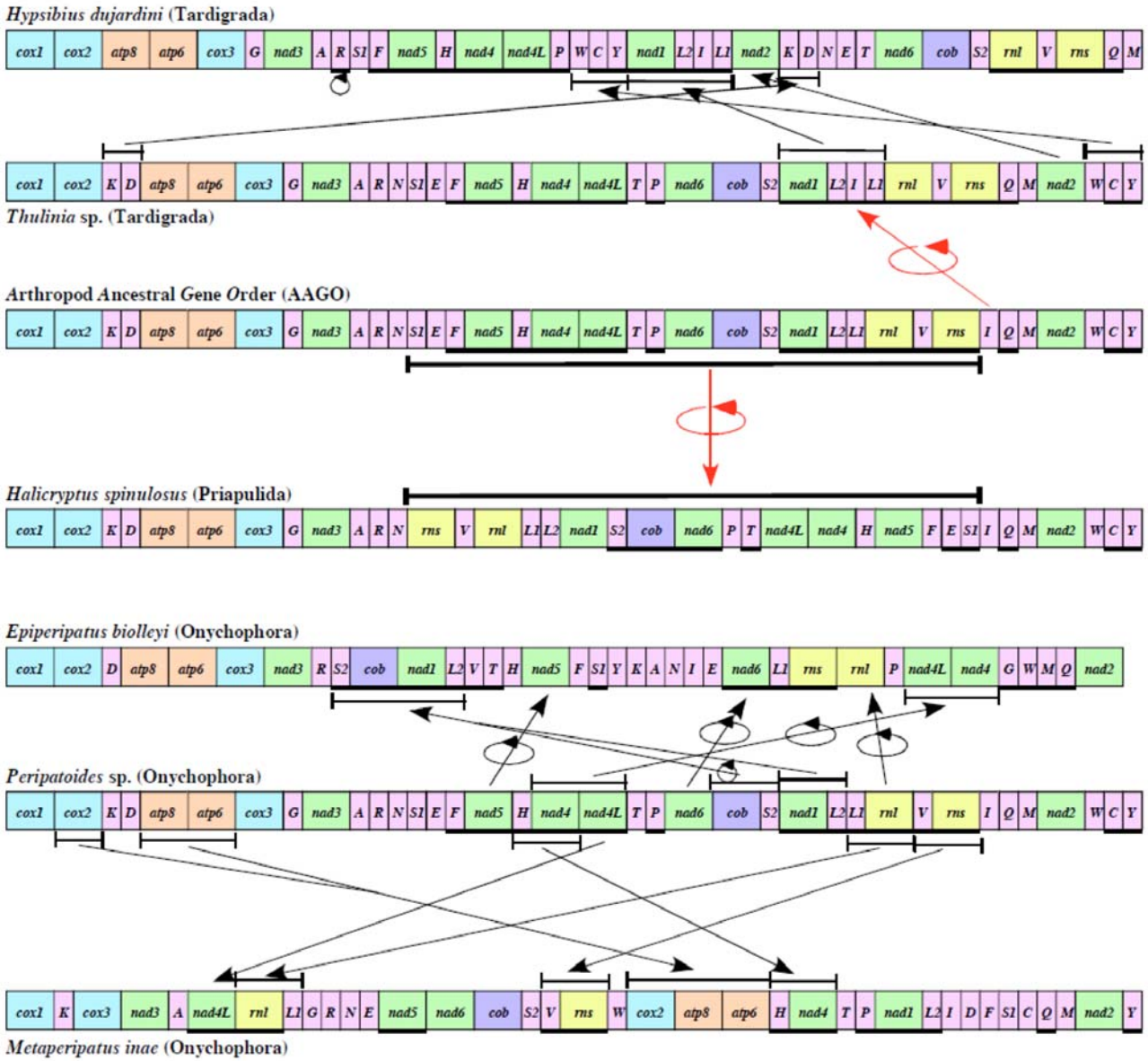


Fig 3

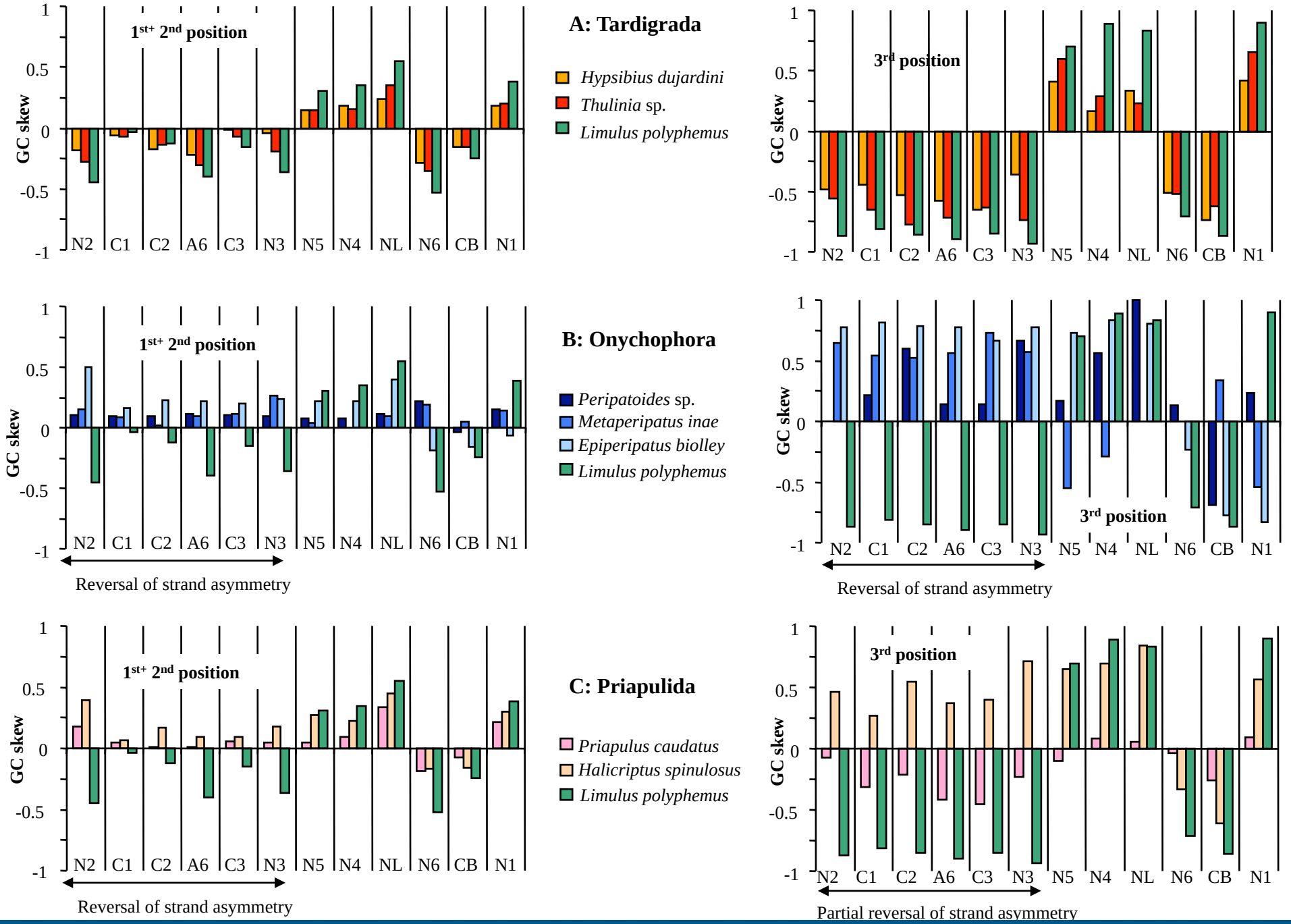


Fig 4

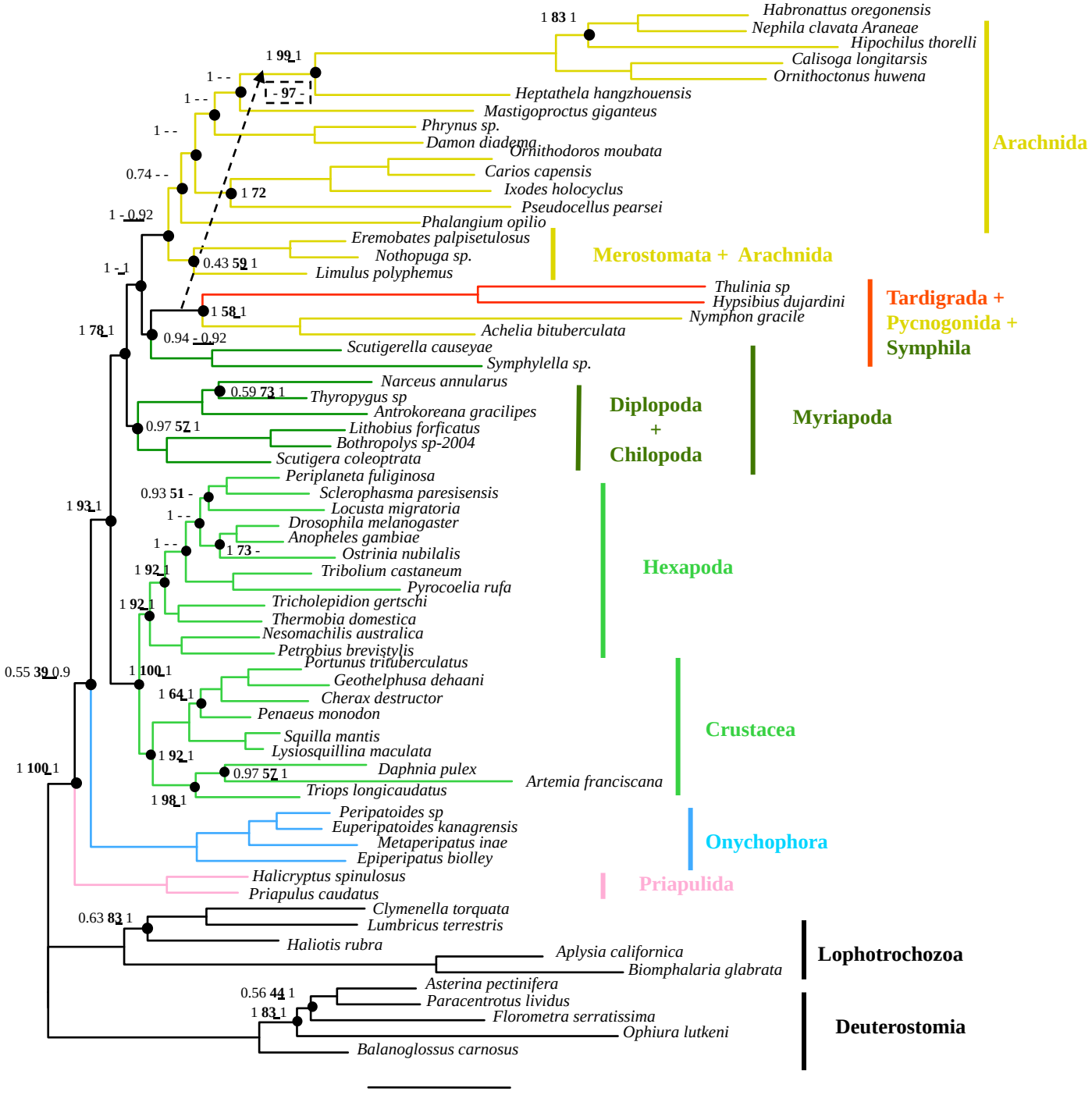


Fig 5

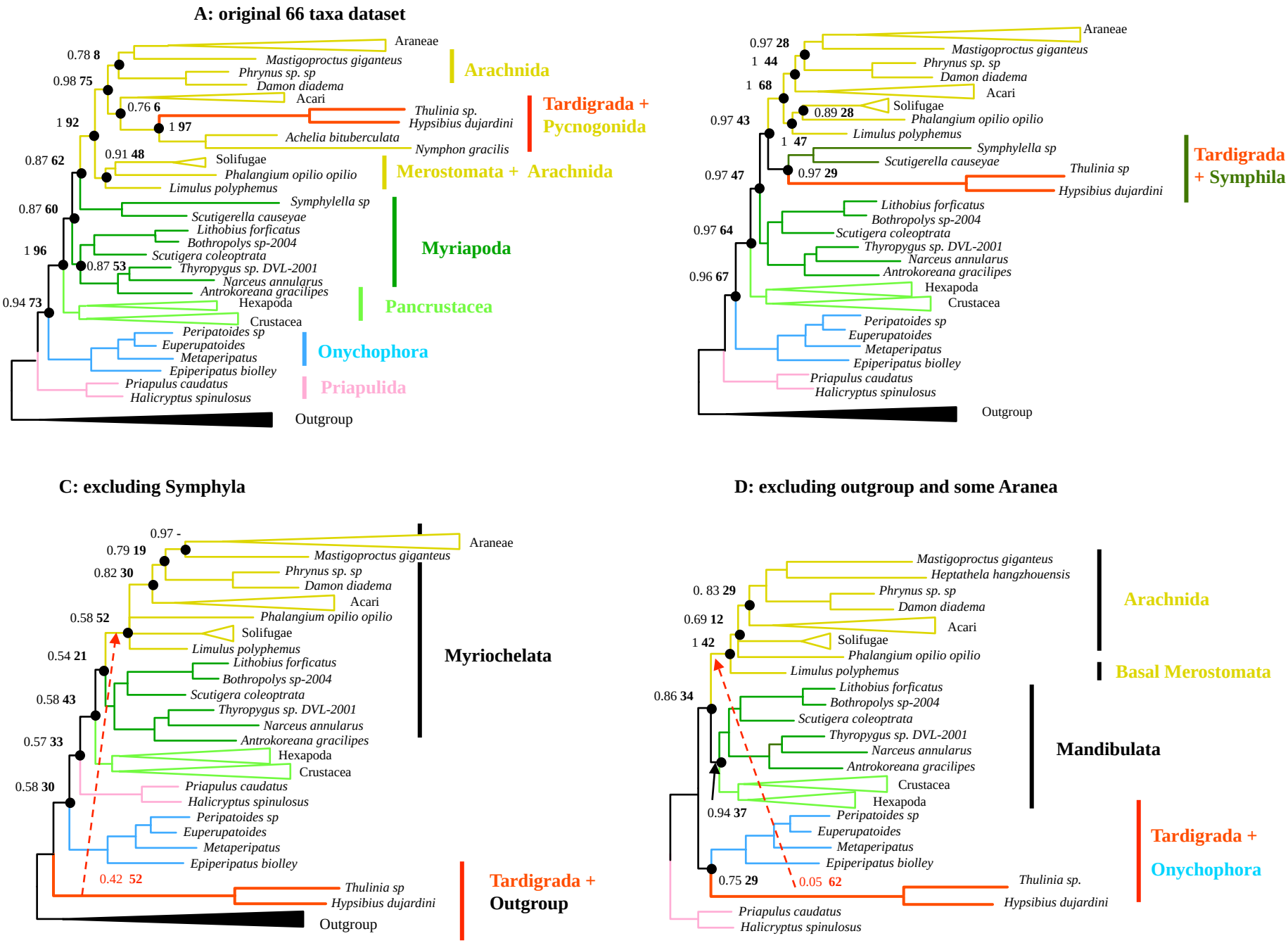


Fig 6

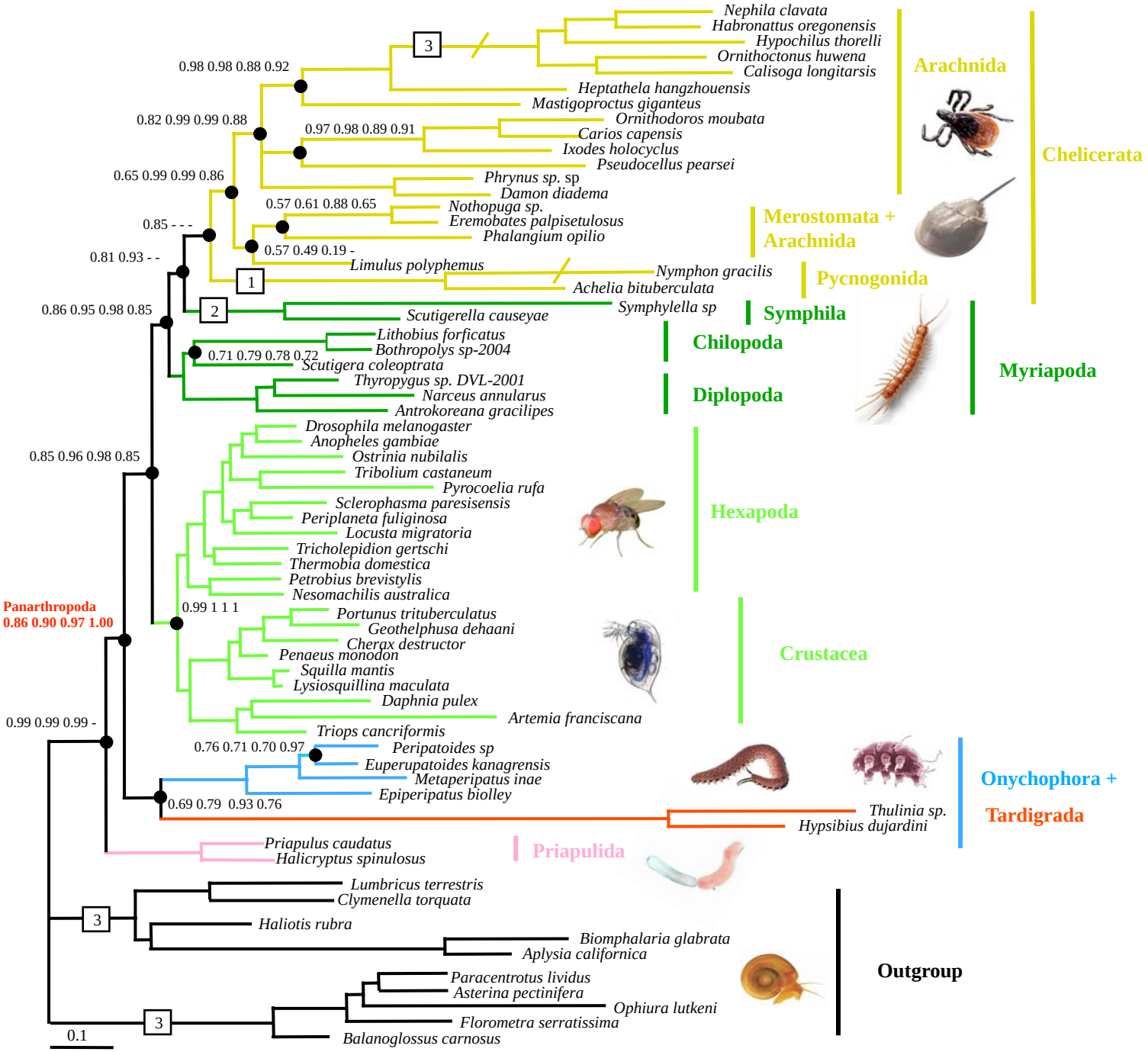
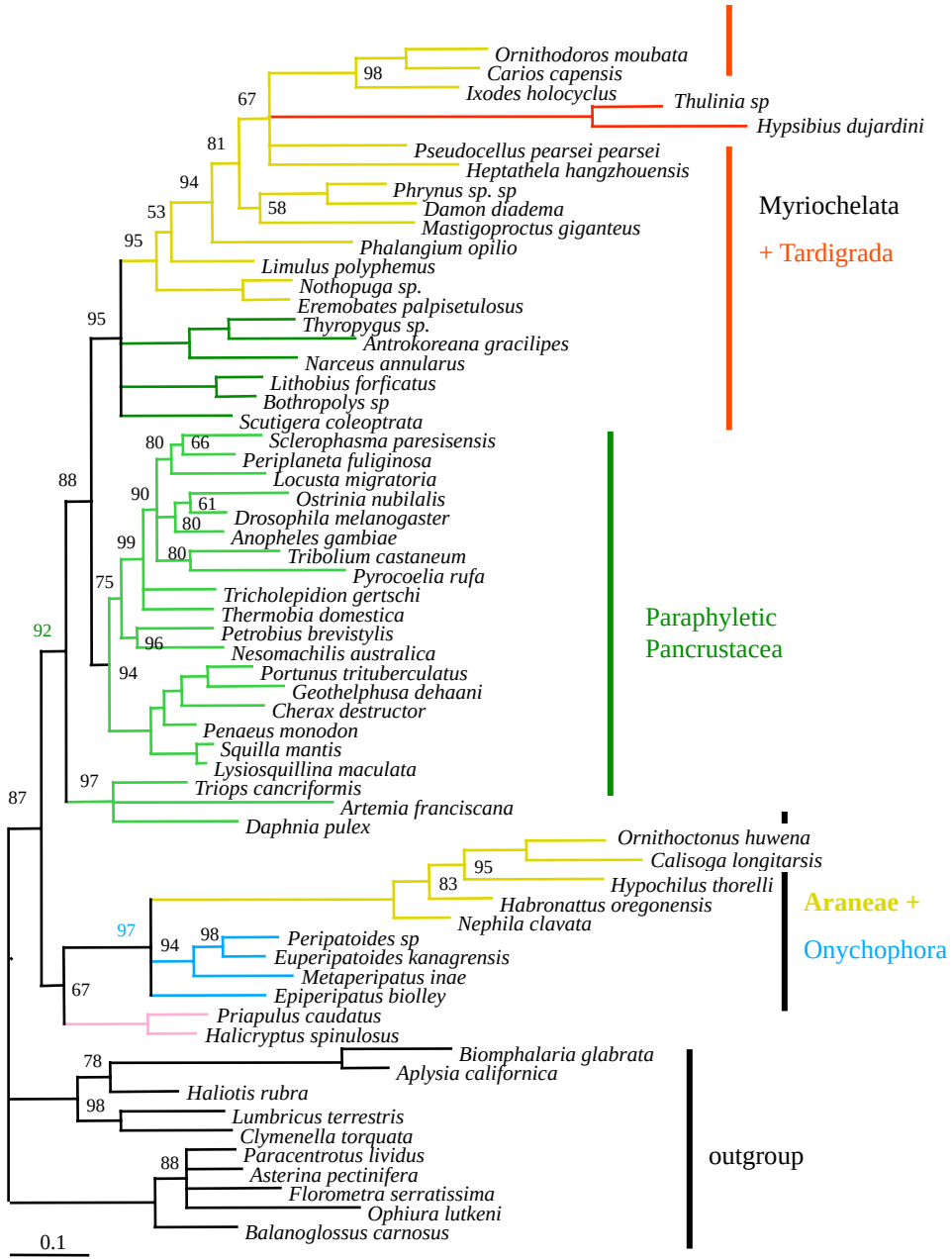


Fig 7

A: : fast and slow evolving sites (external quartiles)



B: : homogenously evolving sites (internal quartiles)